ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20250508002655
Original scientific paper

HFTN-SR: High-Frequency Feature Transfer Network for MR Image Super-Resolution

Yan JIN

Abstract: Magnetic resonance imaging plays a crucial role in clinical diagnosis due to its ability to provide information on soft tissue structure. At present, multi-contrast super-resolution (SR) methods for magnetic resonance (MR) images have been widely studied and have achieved good results. However, most studies overlook the impact of modal differences between reference and low-resolution (LR) image features on feature reconstruction, which may result in inaccurate reconstruction of detail features due to improper alignment of structural features. To address this issue, we propose a high-frequency feature transfer network for MR image SR task (HFTN-SR), which consists of two feature extraction branches and one high-frequency feature transfer branch. Considering the modal differences between the target and reference images, a high-low frequency decomposition method is used to decompose the reference image into high-frequency and low-frequency components, where the low-frequency components are used in the subsequent network to match LR target image features. A feature extraction block (FEB) is constructed to extract and integrate high-frequency and low-frequency features of the reference image, as well as features of the LR target image. In response to the modal differences between two image features, a feature transfer block (FTB) is designed to establish the correlation between the low-frequency features of the reference image and the target image features, and use the correlation matrix to transfer the high-frequency features of the reference image to the target image features. To further reduce the loss of shallow features caused by the increase in network depth, a standardized combined residual feature module (SCRFM) is constructed to supplement the shallow features of the target image into the final reconstructed features. Experiments on the public dataset FastMRI and the self-built dataset AXA show that the performance of HFTN-SR is superior to some state-of-the-art (SOTA) methods. Notably, HFTN-SR achieves the highest PSNR and SSIM scores across all tested datasets, with significant improvements in visual quality and detail reconstruction.

Keywords: feature transfer; high-frequency; MRI; super-resolution

1 INTRODUCTION

Magnetic resonance imaging (MRI) stands as one of the most widely used medical imaging modalities. In comparison to computed tomography (CT) imaging and positron emission tomography (PET) imaging, MRI is widely regarded as the most reliable imaging modality for detecting minute pathological changes in the brain, spinal cord, and related systems. However, acquiring high-resolution (HR) MR images typically demands extended scanning durations. Due to constraints imposed by imaging hardware and limitations in post-processing capabilities, the resulting MR images often exhibit low resolution. Efforts to enhance spatial resolution can lead to a reduction in signal-to-noise ratio, increased strain on hardware resources, accelerated equipment wear, and elevated operational costs [1, 2]. As a solution, image super-resolution (SR) techniques [3, 4] have emerged as effective post-acquisition methods to enhance image clarity without modifying existing hardware configurations, and have thus become an essential tool in improving the quality of MR imaging.

SR approaches are generally classified into two main types: single-contrast and multi-contrast methods. Single-contrast SR techniques [5-8] rely exclusively on the data contained within a single LR image to reconstruct a corresponding HRimage output. Due to the limited feature information in LR images, this type of method is difficult to reconstruct rich texture information. However, the texture information of medical images is an important basis for doctors' diagnosis. Due to the lack of more structural information in LR medical images, single-contrast SR methods [9, 10] for LR medical images cannot achieve good performance.

Multi-contrast SR techniques use an HR MR image from another modality as a reference to reconstruct richer texture details in the low-resolution MR image of the same anatomical region. At present, most of the SR approaches for MR images adopt multi-contrast reconstruction methods, which use two different MR modalities to reconstruct one modality's image. Fig. 1 presents two pairs of MR images captured using different modalities but depicting identical anatomical regions. These include combinations such as fluid attenuation inversion recovery (FLAIR) with T2-Weighted imaging (T2WI), and proton density weighted imaging (PDWI) with fat-suppressed proton density Weighted Imaging (FS-PDWI). These image variations are produced by altering the acquisition parameters on the MRI scanner [11, 12]. As illustrated, although each image pair shares the same structural features, they highlight distinct tissue characteristics. This complementarity enables one modality to be effectively used as a reference to support the super-resolution reconstruction of the other.

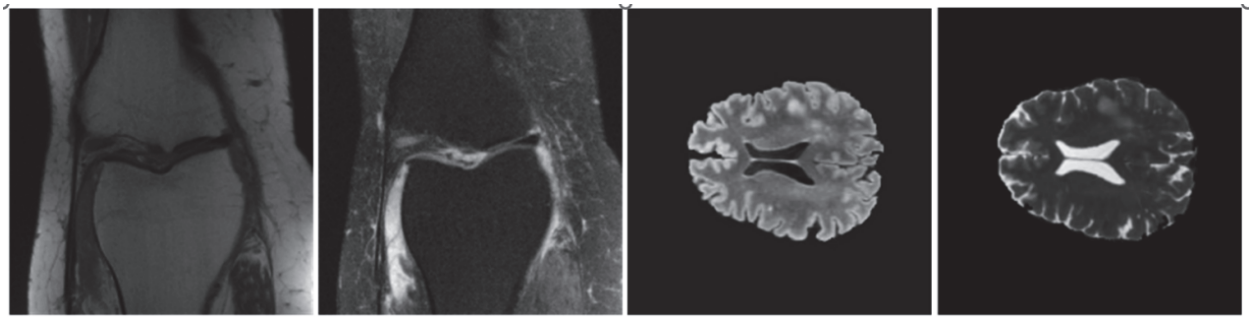

Figure 1 Two sets of different modal MR images

The powerful feature representation capabilities of convolutional neural networks (CNNs) have driven significant progress in multi-contrast super-resolution (SR) methods [13-16], which leverage deep learning techniques to enhance MR image quality. Zeng et al. [17] proposed a deep CNN-based model capable of simultaneously performing both single-contrast and multi-contrast SR reconstruction tasks. Lyu et al. [18] designed an SR network with two architectures, a one-level non-progressive network and a two-level progressive network, for multi-contrast SR reconstruction. Feng et al. [19] proposed a multi-level feature fusion network for multi-contrast SR reconstruction, which explicitly models the dependency relationships between multi-contrast images across different stages. Li et al. [20] proposed the McMRRS method that utilizes a multi-scale context matching and aggregation scheme to achieve stepwise interactive aggregation of multi-scale matching features. Liu et al. [21] introduced a Flexible Alignment Super-Resolution Network (FASRNet) tailored for multi-contrast MRI super-resolution tasks. This network incorporates two alignment modules, Single-Multi Pyramid Alignment (SA) and Multi-Multi Pyramid Alignment (M-A) to effectively align features across different contrasts. Zhang et al. [28] proposed a deep unfolding network with spatial alignment (DUN-SA), which incorporates spatial alignment into the reconstruction process. By jointly optimizing cross-modal alignment and reconstruction, their method effectively improves image reconstruction quality. Wang et al. [29] introduced a cross-modality MRI reconstruction framework that integrates deep learning with optimal transport theory, significantly mitigating inter-modality misalignment and achieving efficient and accurate image reconstruction. Although the aforementioned methods have achieved certain success in reconstruction performance, the target modality image often lacks abundant high-frequency texture details, while the reference modality preserves more structural boundaries and rich texture information. Therefore, effectively leveraging the high-frequency features from the reference modality to compensate for missing details in the target modality is essential for improving super-resolution reconstruction quality. However, current methods do not explicitly address the distributional differences between the two modalities in the feature space, and instead directly fuse their features. Such fusion strategies, which overlook modality-specific discrepancies, often result in inaccurate structural alignment and consequently lead to blurred edges and the loss of fine details in the reconstructed images.

To address this issue, we propose a high-frequency feature transfer network for MR image SR (HFTN-SR). This network explicitly decomposes the reference image into high-frequency and low-frequency components, and effectively transfers high-frequency features to target image features by establishing the correlation between these two component features and the target image features. In this way, HFTN-SR reconstructs high-resolution MR images with enhanced structural details.

Specifically, we employ a frequency decomposition technique to separate the reference image into high-frequency and low-frequency components. A feature extraction block (FEB) is constructed to extract features from both the target image and the high/low-frequency components of the reference image. Considering the spatial resolution similarity between the low-frequency components and the target image, we further design a feature transfer block (FTB) that computes a correlation matrix between the low-frequency features and the target features. This correlation is then used to selectively transfer high-frequency information from the reference image to the target image features, thereby significantly improving edge sharpness and detail representation in the super-resolved outputs. The contributions of this work are as follows:

1) A HFTN-SR for multi-contrast MR image SR is proposed, which transfers high-frequency features from the reference image to the target image features by leveraging the correlation between the low-frequency features of the reference image and the target image features, thereby obtaining SR images with enhanced details.

2) Two feature extraction branches and one feature fusion branch are constructed to gradually achieve the extraction and fusion of two types of image features. A FEB is constructed for feature extraction, and an FTB is designed to achieve the fusion of two image features by transferring high-frequency features from the reference image to the target image features.

3) To reduce the feature loss caused by the increase in network depth, a standardized combined residual feature module (SCRFM) is constructed to achieve the fusion of transfer fusion features and shallow features of the target image, supplementing shallow features for reconstructed features.

2 PROPOSED METHOD

To generate high-quality SR images from low-resolution MRI data, we propose a HFTN-SR for MRI image SR task. As shown in Fig. 2, this network is composed of three sub-branches: two feature extraction branches and one high-frequency feature transfer branch. In HFTN-SR, two feature extraction branches are constructed to extract features of the target image and high-frequency features of the reference image at different depths, respectively. To facilitate the transfer of high-frequency details from the reference image to the target image, a dedicated high-frequency feature transfer branch is designed. This branch establishes feature correlations between the two images and integrates fine-grained details from the reference image into the target image features to obtain fused features with rich details. Then, to reduce the loss of features in the fused result, SCRFM is designed to achieve the fusion of fused features with shallow features of the reference image. Finally, to enhance the detail information in the final fusion result, a residual structure is adopted to supplement the detail features by utilizing the high-frequency features of the reference image.

Below, we will provide a detailed introduction to the three branches and main modules in the network.

2.1 Two Feature Extraction Branches

Considering the differences in feature representations

between the two images, we construct two structurally distinct feature extraction branches, one for extracting target image features and the other for extracting high-frequency component features of the reference image. As shown in Fig. 2, these two branches are respectively referred to as the upper branch and the lower branch.

In the upper branch, first, the target image is upsampled to the same size as the reference image. Then, a 3 × 3 convolution is employed to obtain shallow features F_{Tar}^{0} of the target image. Finally, multiple feature extraction blocks (FEBs) are constructed to achieve feature integration and extraction from the fusion branch at different depths, and the output of each FEB is fed into the feature fusion branch for fusion with the reference image features.

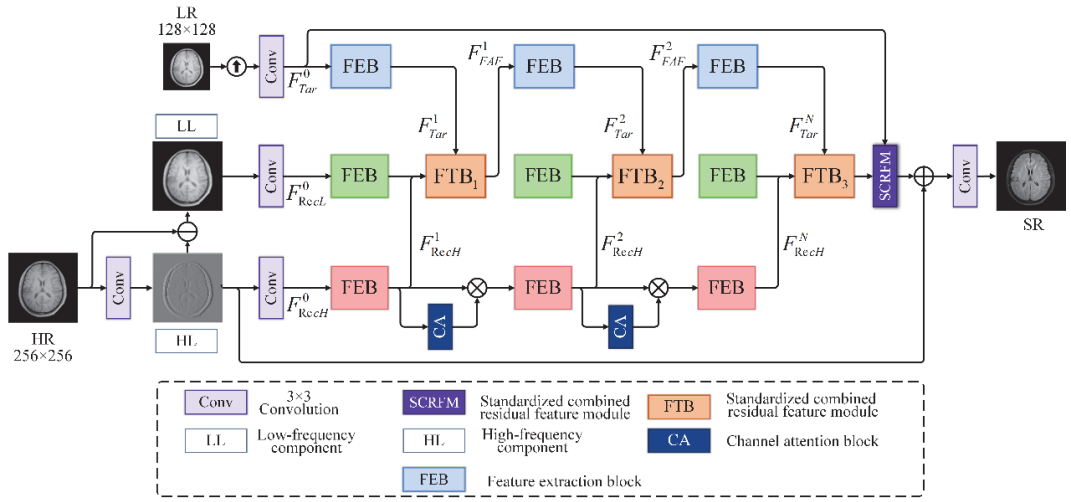

Figure 2 Architecture of HFTN-SR

The FEB is constructed as a residual block containing two 3 × 3 convolutions and a ReLU function. The above operations can be expressed as the following equation.

$$F_{\mathrm{Tar}}^{0}=\mathrm{Conv}_{3\times3}\left(\mathrm{Up}\left(I_{\mathrm{LR}}\right)\right) \tag{1}$$

$$\begin{aligned}F_{\mathrm{Tar}}^{i}&=FEB\left(F_{\mathrm{FAF}}^{i-1}\right)\\&=F_{\mathrm{FAF}}^{i-1}+\mathrm{Conv}_{3\times3}\left(\mathrm{Relu}\left(\mathrm{Conv}_{3\times3}\left(F_{\mathrm{FAF}}^{i-1}\right)\right)\right)\\F_{\mathrm{FAF}}^{0}&=F_{\mathrm{Tar}}^{0},\ i=1,\ldots,n\end{aligned} \tag{2}$$

where Up(·) and $\mathrm{Conv}_{3\times3}(\cdot)$ represent upsampling and 3 × 3 convolution operations, respectively. FEB(·) represents the operation of FEB, and its output is represented as F_{Tar}^{i} (i = 1, …, n). F_{FAF}^{i} represents the fusion features from feature fusion branch, and n is the number of FEBs.

In the lower branch, to effectively extract fine-grained details from the reference image I_{Re}, a high-low frequency decomposition operation is first performed. Specifically, a Gaussian blur is applied to smooth the image and extract its low-frequency component, which is then subtracted from the original image to obtain the high-frequency component. Compared to ideal low-pass filters or hard thresholding approaches, Gaussian filtering provides a smooth and continuous low-pass operation that effectively suppresses high-frequency noise while avoiding artifact introduction. Moreover, it preserves the overall structural information of the image, allowing for a more natural and reliable separation of high- and low-frequency features. This property facilitates more accurate feature extraction and contributes to improve reconstruction performance.

The high-frequency and low-frequency components are separately fed into a 3 × 3 convolutional layer to obtain shallow features F_{Rec}^{0} and F_{RecL}^{0}. Then, feature extraction blocks with the same structure as those in the upper branch are used to extract the features of high-frequency component. In addition, to obtain more informative representations, channel attention is employed to enhance high-frequency features at each depth. The above operations are as follows.

$$\left\{\mathrm{LF},\mathrm{HF}\right\}=\left\{\mathrm{GB}\left(I_{\mathrm{Re}}\right),I_{\mathrm{Re}}-\mathrm{GB}\left(I_{\mathrm{Re}}\right)\right\} \tag{3}$$

$$\begin{aligned}F_{\mathrm{Rec}}^{0}&=\mathrm{Conv}_{3\times3}\left(\mathrm{HF}\right),\\F_{\mathrm{RecH}}^{i}&=\mathrm{CA}\left(\mathrm{FEB}\left(F_{\mathrm{RecH}}^{\mathrm{i\text{-}1}}\right)\right)\otimes\left(\mathrm{FEB}\left(F_{\mathrm{RecH}}^{i-1}\right)\right),\\&i=1,\ldots,n\end{aligned} \tag{4}$$

where GB(·) and CA(·) represent Gaussian blur and channel attention operations, respectively, and ⊗ represents the Hadamard product. LF and HF represent low frequency and high frequency components, respectively, and F_{RecH}^{i} represents the high-frequency features at different depths, where $F_{\mathrm{RecH}}^{0}=F_{\mathrm{Rec}}^{0}$.

2.2 High-Frequency Feature Transfer Branch

High-frequency feature transfer branch is constructed to transfer high-frequency features from the reference image to the target image. At various depths, feature extraction blocks are employed to capture progressively deeper features. Simultaneously, a Feature Transfer Block (FTB) is introduced to transfer high-frequency features from the reference image to the target image features by

establishing feature correlations between the two modalities. The operations of this branch are represented as follows.

$$F_{\mathrm{RecL}}^{0} = \mathrm{Conv}_{3\times3}\left(\mathrm{LF}\right) \tag{5}$$

$$F_{\mathrm{RecL}}^{i} = \mathrm{FEB}\left(F_{\mathrm{RecL}}^{i-1}\right) \tag{6}$$

$$F_{\mathrm{FAF}}^{i} = \mathrm{FTB}\left(F_{\mathrm{RecL}}^{i}, F_{\mathrm{RecH}}^{i}, F_{\mathrm{Tar}}^{\mathrm{i}}\right) \tag{7}$$

where FTB(·) represents the operation of FTB. F_{RecL}^{i} and F_{FAF}^{i} represents the output of feature extraction block and fusion features at different depths. The specific structure of FTB is illustrated below. As shown in Fig. 3, FTB mainly consists of three steps: relevance embedding, transfer attention and feature synthesis.

2.2.1 Relevance Embedding

There is a significant difference in high-frequency information between the reference image and the target image, while the low-frequency information between the two has a high correlation. Therefore, to transfer high-frequency features from the reference image, we first need to calculate the correlation matrix R between the features from the upper branch and the low-frequency features of the reference image. Specifically, F_{Tar}^{i} and F_{RecL}^{i} are first divided into non-overlapping patches using the operation Div(·). Each patch is then reshaped into a 1D vector with Re(·), and projected into a lower-dimensional embedding space via a learnable linear projection:

$$\begin{aligned} q_p &= \mathrm{Linear}\left(\mathrm{Re}\left(\mathrm{Div}\left(F_{\mathrm{Tar}}^{i}\right)\right)\right), \\ k_j &= \mathrm{Linear}\left(\mathrm{Re}\left(\mathrm{Div}\left(F_{\mathrm{RecL}}^{i}\right)\right)\right) \end{aligned} \tag{8}$$

where $q_p(\in[1,\ h\times w])$ and $k_j(\in[1,\ h\times w])$ represent the query and key vectors corresponding to a patch with a size of $h \times w$, respectively.

To measure the similarity between the target and reference patches, we compute the normalized inner product (i.e., cosine similarity) between each pair of query and key vectors:

$$r_{p,j} = \frac{q_t}{|q_t|}, \frac{k_j}{|k_j|} \tag{9}$$

where $\langle\cdot\rangle$ represents the inner product, and $|\cdot|$ represents the operation of vector modulus.

Based on this correlation matrix, we then construct a relevance embedding matrix T to guide the high-frequency feature transfer. Specifically, to transfer high-frequency features from the reference image, the embedding position t_p corresponding to each q_p can be obtained by the following formula:

$$t_p = \mathrm{Pos}\left(\mathrm{argmax}_j\left(r_{p,j}\right)\right) \tag{10}$$

where argmax_j represents taking the maximum value in the j-th row of R, and Pos(·) represents the operation of obtaining the positions corresponding to the correlation coefficients. Here, $t_p\left(p\in\left[1,\ \frac{h}{2}\times\frac{w}{2}\right]\right)$ is the p-th element in T.

2.2.2 Transfer Attention

This step aims to select the high-frequency features to be migrated from the feature vectors V. The traditional attention mechanism utilizes an attention matrix to perform weighted summation operations on feature maps. Due to the differences between the features of two different modalities, the direct integration of the two features can lead to blurred edges in the image. Therefore, unlike traditional attention mechanisms, we directly select the high-frequency features from V based on the positional information in the relevance embedding matrix T to obtain the transferred high-frequency features C. Here, except for the positions recorded by T, the values of other positions in C are all 0.

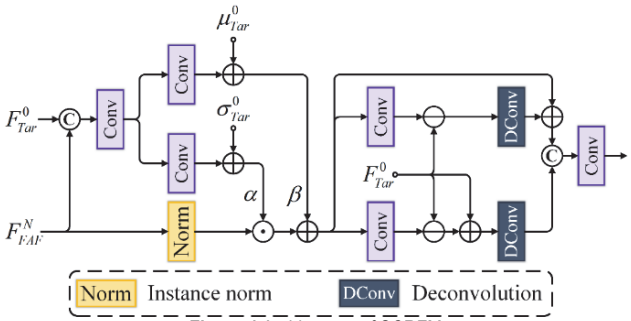

Figure 4 Architecture of SCRFM

2.2.3 Feature Synthesis

The transferred high-frequency features are integrated with the features F_{Tar}^{i} from the upper branch through a concatenation operation and a 3 × 3 convolutional layer to obtain the supplementary features Z. To further enhance the structural information of the supplementary features, we utilize the correlation matrix R to weight the supplementary features. Enhanced features are added to the features from the upper branch to obtain the output of FTB, thereby achieving the transfer of high-frequency features from the reference image to the target image features. The above operations can be expressed as:

$$F_{\mathrm{FAF}}^{i} = F_{\mathrm{Tar}}^{i} + \mathrm{Re}\left(\mathrm{Conv}_{3\times3}\left(\mathrm{Concat}\left(C, Q\right)\right)\otimes R\right) \tag{11}$$

where Concat(·) represents the concatenation operation.

2.3 Standardized Combined Residual Feature Module (SCRFM)

As shown in Fig. 4, to reduce the loss of the features, a SCRFM is designed to fuse the shallow features from the

target image with the fused features F_{FAF}^{N} from the fusion branch. This module consists of two stages: standardized feature fusion and refined feature fusion.

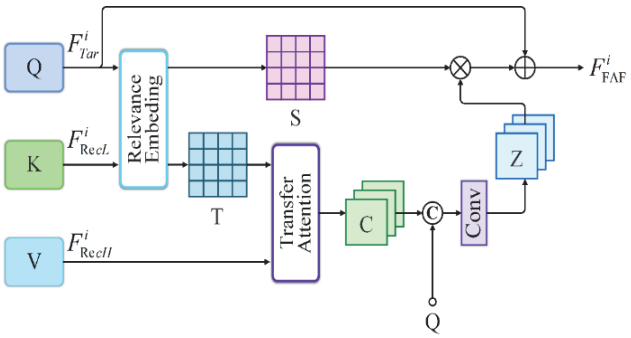

Figure 3 Architecture of FTB

In the first stage, considering the differences between deep and shallow features, standardized feature fusion is performed by remapping the distribution of fused features to that of low-resolution target image features. Specifically, each channel of the fused features is first normalized through instance normalization operation. Then, the shallow features and fused features are fed together into three 3 × 3 convolutional layers to obtain two integrated features, which are added to the mean and variance of the shallow features to obtain two parameters α and β. These two parameters are used to perform instance normalization operations on the fused features. The operations of the first stage are as follows:

$$\begin{aligned} IF_{\mathrm{FAF}}^{N} &= IN\left(F_{\mathrm{FAF}}^{N}\right), \\ F_{\mathrm{FT}} &= \mathrm{Conv}_{3\times3}\left(\mathrm{Concat}\left(F_{\mathrm{FAF}}^{N}, F_{\mathrm{Tar}}^{0}\right)\right) \end{aligned} \tag{12}$$

$$\alpha = \mathrm{Conv}_{3\times3}\left(F_{\mathrm{FT}}\right) + \mathrm{Var}\left(F_{\mathrm{Tar}}^{0}\right) \tag{13}$$

$$\beta = \mathrm{Conv}_{3\times3}\left(F_{\mathrm{FT}}\right) + \mathrm{Mean}\left(F_{\mathrm{Tar}}^{0}\right) \tag{14}$$

$$F_{sff} = \alpha \cdot IF_{\mathrm{FAF}}^{N} + \beta \tag{15}$$

where $IN(\cdot)$ represents the operation of instance normalization, and its output is IF_{FAF}^{N}. Var(·) and Mean(·) represent the calculation operations of matrix mean and variance. F_{sff} is the output of the first stage.

In the second stage, two 3 × 3 convolutional layers are first used to integrate the fused features from the first stage. Then, these fused features are subtracted from shallow features to obtain differential features, which are used for further feature supplementation. Next, two residual operations are employed to supplement the differential features into the shallow and fused features, and two deconvolution operations are used to restore the size of the feature maps. Finally, the two supplemented features SF_{sff} and SF_{Tar}^{0} are integrated through an addition operation and a 3 × 3 convolutional layer to obtain the fusion features F_f. The operations of the second stage are as follows:

$$SF_{sff} = D\mathrm{Conv}_{3\times3}\left(F_{\mathrm{Tar}}^{0} - \mathrm{Conv}_{3\times3}\left(F_{sff}\right)\right) + F_{sff} \tag{16}$$

$$SF_{\mathrm{Tar}}^{0} = D\mathrm{Conv}_{3\times3}\left(\left(F_{\mathrm{Tar}}^{0} - \mathrm{Conv}_{3\times3}\left(F_{sff}\right)\right) + F_{\mathrm{Tar}}^{0}\right) \tag{17}$$

$$F_f = \mathrm{Conv}_{3\times3}\left(\mathrm{Concat}\left(SF_{sff}, SF_{\mathrm{Tar}}^{0}\right)\right) \tag{18}$$

where $D\mathrm{Conv}_{3\times3}(\cdot)$ represents the deconvolution operation.

2.4 Loss Function

For network training, L_1 loss is employed to define the reconstruction loss L_{rec} used for network training, which can evaluate the dissimilarity between the super-resolution result I_{SR} and corresponding ground-truth (GT) image I_{GT}. The loss is defined as follows:

$$L_{\mathrm{rec}} = \frac{1}{LWH}\left\| I_{\mathrm{GT}} - I_{\mathrm{SR}} \right\|_1 \tag{19}$$

where $\|\cdot\|_1$ represents the L_1 loss. W and H is the width and height of the images, and L is the number of channels.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Datasets

To assess the performance of the proposed technique, we performed comprehensive evaluations on both a widely recognized public dataset and a custom-constructed dataset. The public dataset, FastMRI [22], is currently the largest open-access MRI collection. In parallel, we developed the AXA dataset using clinical data acquired in-house. For Fast MRI, we adopted the data selection strategy used in LMKSP [23], extracting 227 image pairs for training and 24 for validation, covering both PD and PDFS modalities. The AXA dataset was compiled from scans of 50 patients, obtained with a 3T Siemens MagnetomSkyra scanner. Each scan includes fully sampled k-space data for T2 and Flair sequences, with scanning parameters set to TRFlair = 9000 ms, TEFlair = 120 ms, TRT2 = 5725 ms, and TET2 = 100 ms. Ethical approval for the clinical data collection was granted by the Institutional Review Board. The AXA dataset was partitioned into training, validation, and test subsets using a 70:10:20 split. In experiments involving FastMRI, PD serves as the input modality while PDFS is treated as the reconstruction target. In contrast, for AXA experiments, T2 images function as inputs and Flair images as outputs.

3.2 Experimental Setup

The model was implemented on a single NVIDIA RTX A6000 GPU equipped with 48 GB of memory. Similar to MINet [19] and SANet [24], our training parameters were set as follows: training was conducted for 50 epochs using the Adam optimizer, with a fixed learning rate of 5×10^{-5}. The network parameter N was configured to 8. For the Fast MRI dataset, performance metrics for all baseline methods were sourced directly from SANet [24]. In contrast, for the AXA dataset, each comparative method

was independently trained and evaluated to generate the corresponding performance metrics.

3.3 Experimental Results

We evaluate the relative performance of different methods using two widely accepted quantitative measures: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). These indicators are widely used in image reconstruction tasks, with higher scores reflecting improved reconstruction accuracy. In order to validate the robustness of our proposed framework, we conduct a systematic comparison against several state-of-the-art (SOTA) models, namely EDSR [26], MCSR [17], NEU [27], MINet [19], MSDT [25], and SANet [24]. The reference results for these methods are obtained from their officially reported benchmarks. Tab. 1 and Tab. 2 present the average evaluation outcomes for upscaling factors of ×2 and ×4, evaluated on the FastMRI and AXA datasets, respectively. It is evident from the tabulated data that our approach achieves superior results across all metrics. These findings highlight the model's capacity to reconstruct high-resolution images with greater fidelity to the GT images.

Table 1 Average quantitative results of all comparison methods at upscaling factors of ×2 and ×4 on the FastMRI dataset

Dataset / Scale factor	FastMRI			
	×2		×4	
Method	Metric			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
EDSR [26]	26.70	0.512	18.40	0.208
MCSR [17]	29.50	0.670	23.30	0.476
NEU [27]	30.30	0.682	28.20	0.634
MINet [19]	31.80	0.709	29.80	0.641
MSDT [25]	31.98	0.713	30.38	0.615
SANet [24]	32.00	0.716	30.40	0.662
Ours	32.67	0.731	30.85	0.671

Table 2 Average quantitative results of all comparison methods at upscaling factors of ×2 and ×4 on the AXA dataset

Dataset / Scale factor	AXA			
	×2		×4	
Method	Metric			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
EDSR [26]	29.71	0.803	24.32	0.628
MCSR [17]	33.64	0.879	29.98	0.738
NEU [27]	35.22	0.902	31.27	0.791
MINet [19]	36.41	0.916	33.38	0.831
MSDT [25]	36.17	0.914	33.41	0.834
SANet [24]	36.54	0.925	33.66	0.844
Ours	37.02	0.933	33.81	0.857

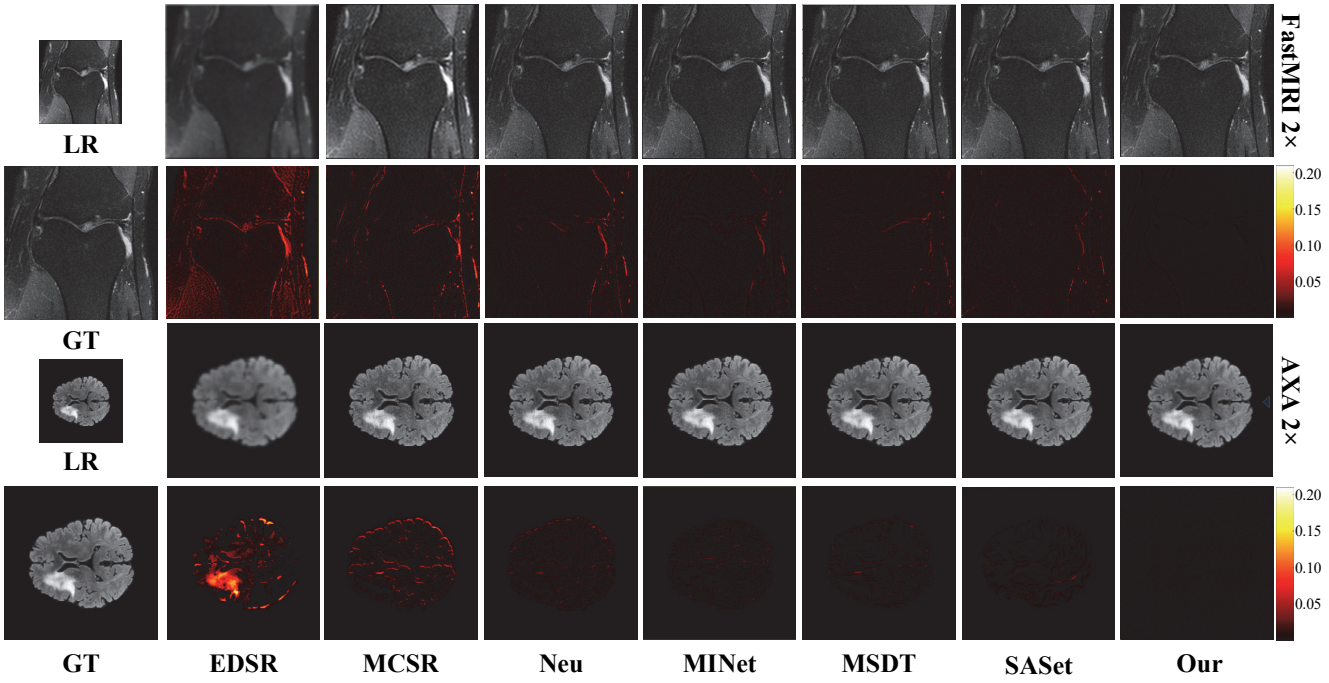

Figure 5 The visualization results of SR × 2 of all comparison methods on the FastMRI and AXA datasets. The first and second lines present the SR results and corresponding residual images of an image from the FastMRI dataset. The third and fourth lines show the SR results and corresponding residual images of an image from the AXA dataset

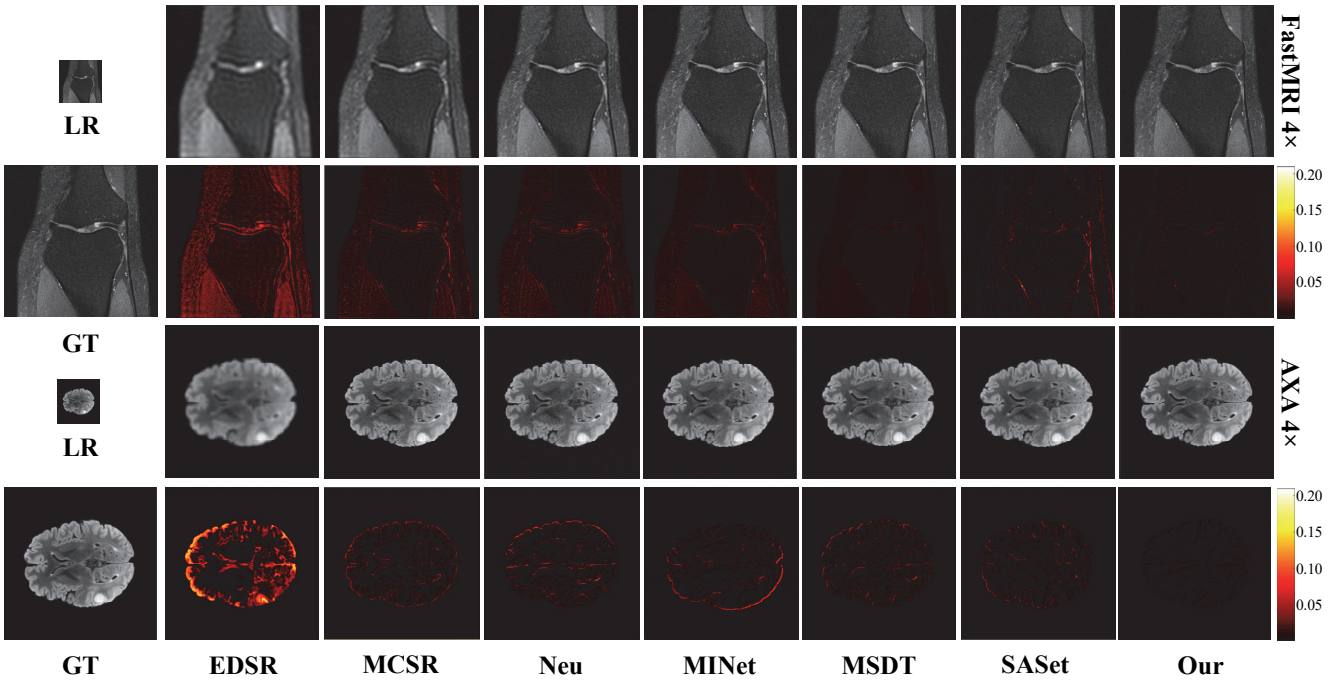

Figure 6 The visualization results of SR × 4 of all comparison methods on the FastMRI and AXA datasets. The first and second lines present the SR results and corresponding residual images of an image from the FastMRI dataset. The third and fourth lines show the SR results and corresponding residual images of an image from the AXA dataset

Fig. 5 and Fig.6 illustrate the visual comparisons for super-resolution factors of ×2 and ×4, using sample images drawn from each of the two datasets. To better highlight the distinctions between the reconstructed outputs and their corresponding GT images, we compute and display the residual maps. As observed in the visualizations, the EDSR model, which relies solely on a single low-resolution input, tends to produce results lacking fine details and exhibiting noticeable edge blurring. Consequently, its residual maps show a higher concentration of reconstruction errors. In contrast, the other methods - including our proposed approach leverage additional reference images during reconstruction, enabling them to recover richer textures and structural content. Compared to other methods, our approach achieves refined reconstruction results with more details. These improvements are particularly evident in the highlighted regions of the reconstructed images, and our method preserves structural integrity with minimal distortion, resulting in outputs that are closer to the ground truth and exhibit higher fidelity. However, as the scaling factor increases, noticeable ringing artifacts emerge in the results of EDSR, MCSR, and NEU. Analysis of the residual maps reveals that our method achieves superior detail preservation and artifact suppression, with substantially lower residuals than baseline methods, reflecting higher fidelity to the original GT images.

3.4 Ablation Studies

To validate the effectiveness of the proposed components, we conducted several ablation experiments on high-low frequency decomposition (HLD), FTB, and SCRFM, and the results are shown in Tab. 3. Model 1 represents a model without HLD, FTB, and SCRFM, where the input high-frequency and low-frequency components are directly replaced by the input source images. Model 2 represents a model without HLD and with FTB and SCRFM, where the features from the three branches are directly concatenated and fed into the subsequent network. Model 3 represents a model without SCRFM and with HLD and FTB, where the fused features from the fusion branch are directly integrated with shallow features through a 3 × 3 convolutional layer. From the table, it can be seen that after removing the three core modules, the performance of the model showed a significant decline. The addition of each module has improved the performance of the model, and our model achieved the best results with three modules included. This indicates that the proposed modules are all effective.

Fig. 7 shows the visual results of SR × 4 of several ablation experiments on the AXA dataset. From the figure, it can be seen that that the addition of different modules can enable the model to reconstruct more structural information, and the model with three modules reconstructed the richest detailed features. This also demonstrates the effectiveness of the three proposed modules.

Table 3 Average quantitative results of SR × 2 of ablation experiments on the FastMRI dataset for HLD, FTB and SCRFM

models	HLD	FTB	SCRFM	PSNR↑	SSIM↑
1	×	×	×	27.62	0.573
2	√	×	×	29.74	0.601
3	√	√	×	32.16	0.718
our	√	√	√	32.67	0.742

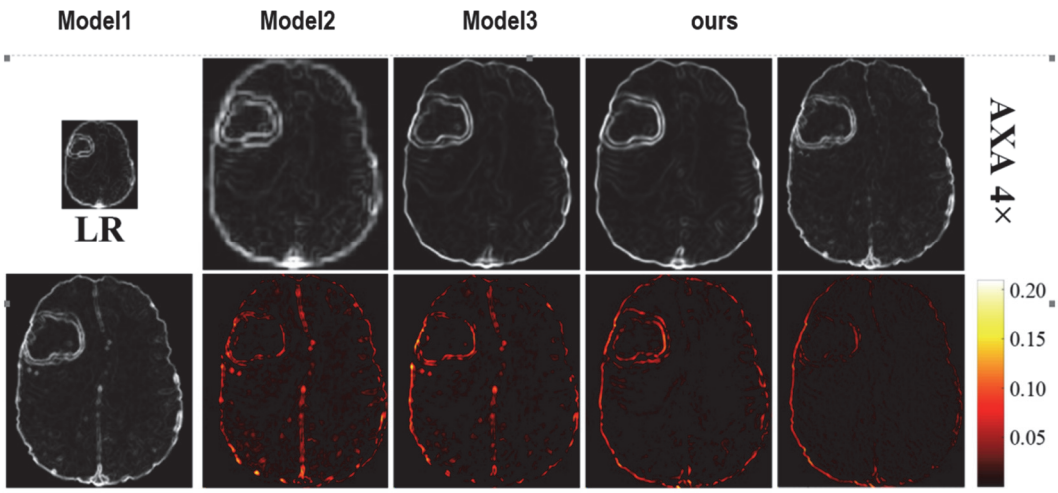

Figure 7 The visualization results of SR × 4 of ablation experiments on the AXA datasets

4 DISCUSSION

This paper proposes a high-frequency feature transfer network called HFTN-SR for MR image super-resolution tasks. By considering the feature differences between the reference and target modalities, the network utilizes the low-frequency components of the reference image as structural guidance and effectively transfers high-frequency information to the target image, demonstrating significant advantages in detail reconstruction. Beyond improvements in quantitative metrics, the method also holds substantial clinical relevance. Accurate depiction of anatomical structures is critical for early disease detection and therapeutic decision-making. Experimental results on the FastMRI and AXA datasets show that HFTN-SR enhances the representation of structural boundaries and texture details, which can assist radiologists in more accurately identifying potential lesions and thus improve diagnostic accuracy. However, the method has certain limitations. The current architecture incurs high computational overhead, posing challenges for real-time deployment in practical clinical settings. Future research will focus on developing lightweight network designs to enhance the model's clinical applicability and generalizability.

5 CONCLUSIONS

In this paper, we propose an HFTN-SR that achieves SR reconstruction of low-resolution MRI by using features from different modality images as guidance. This network achieves feature extraction of the reference image and the target image by constructing two feature extraction branches, and another fusion branch is designed to achieve image detail reconstruction by establishing the correlation between the low-frequency features of the reference image and the features of the target image. All three branches use a single FEB with the same structure to extract and integrate features, and an FTB is designed to transfer high-frequency features from reference images at different depths to target image features. In addition, a SCRFM is constructed to fuse features from the fusion branch with shallow features of the target image to ensure that shallow features are not lost. Numerous experiments have shown that our network substantially improves the structural clarity and detail restoration of MR images, which can assist clinicians in more accurately identifying subtle lesions and holds promise for enhancing diagnostic accuracy in clinical practice. However, the current model imposes considerable computational overhead, making it challenging to meet the requirements for real-time clinical deployment. Future work will focus on designing lightweight architectures to promote practical and efficient application in real-world scenarios.

6 REFERENCES

[1] Van Reeth, E., Tham, I. W., Tan, C. H., & Poh, C. L. (2012). Super-resolution in magnetic resonance imaging: a review. *Concepts in Magnetic Resonance Part A*, *40*(6), 306-325. https://doi.org/10.1002/cmr.a.21249

[2] Feng, C. M., Wang, K., Lu, S., Xu, Y., & Li, X. (2021). Brain MRI super-resolution using coupled-projection residual network. *Neurocomputing*, *456*, 190-199. https://doi.org/10.1016/j.neucom.2021.01.130

[3] Zhang, R., Mo, H., Wang, J., Jie, B., He, Y., Jin, N., & Zhu, L. (2024). UTSRMorph: A Unified Transformer and Superresolution Network for Unsupervised Medical Image Registration. *IEEE Transactions on Medical Imaging*, *44*(2), 891-902. https://doi.org/10.1109/TMI.2024.3467919

[4] Song, T., Wen, R., & Zhang, L. (2024). RoughSet-DDPM: An Image Super-Resolution Method Based on Rough set Denoising Diffusion Probability Model. *Tehnički vjesnik*, *31*(1), 162-170. https://doi.org/10.17559/TV-20230717000808

[5] Wan, W., Wang, Z., Wang, Z., Gu, L., Sun, J., & Wang, Q. (2024). Arbitrary-Scale Image Super-Resolution via Degradation Perception. *IEEE Transactions on Computational Imaging*, 10, 666-676. https://doi.org/10.1109/TCI.2024.3393712

[6] Jia, X., Li, X., Wang, Z., Hao, Z., Ren, D., Liu, H., Du, Y., & Ling, F. (2024). Enhancing Cropland Mapping with Spatial Super-Resolution Reconstruction by Optimizing Training Samples for Image Super-Resolution Models. *Remote Sensing*, *16*(24), 4678. https://doi.org/10.3390/rs16244678

[7] Liu, X. (2020). Real-time Defogging of Single Image of IoTs-based Surveillance Video Based on MAP. *Tehnički vjesnik*, *27*(4), 1262-1269. https://doi.org/10.17559/TV-20200527085338

[8] Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., & Zeng, T. (2022). Transformer for single image super-resolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 457-466.

https://doi.org/10.1109/CVPRW56347.2022.00061
[9] Qiu, D., Zhang, S., Liu, Y., Zhu, J., & Zheng, L. (2020). Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Computer methods and programs in biomedicine*, *187*, 105059. https://doi.org/10.1016/j.cmpb.2019.105059
[10] Zhang, Y., Li, K., Li, K., & Fu, Y. (2021). MR image super-resolution with squeeze and excitation reasoning attention network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13425-13434. https://doi.org/10.1109/CVPR46437.2021.01322
[11] Lyu, Q., Shan, H., & Wang, G. (2020). MRI super-resolution with ensemble learning and complementary priors. *IEEE Transactions on Computational Imaging*, *6*, 615-624. https://doi.org/10.1109/TCI.2020.2964201
[12] Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., & Shen, D. (2018). Deep-learning-based multi-modal fusion for fast MR reconstruction. *IEEE Transactions on Biomedical Engineering*, *66*(7), 2105-2114. https://doi.org/10.1109/TBME.2018.2883958
[13] Zhang, Z., Wang, Z., Lin, Z., & Qi, H. (2019). Image super-resolution by neural texture transfer. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7982-7991. https://doi.org/10.1109/CVPR.2019.00817
[14] Yang, F., Yang, H., Fu, J., Lu, H., & Guo, B. (2020). Learning texture transformer network for image super-resolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5791-5800. https://doi.org/10.1109/CVPR42600.2020.00583
[15] Lu, L., Li, W., Tao, X., Lu, J., & Jia, J. (2021). Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6368-6377. https://doi.org/10.1109/CVPR46437.2021.00630
[16] Zhang, Y., Yang, Q., Chandler, D. M., & Mou, X. (2024). Reference-Based Multi-Stage Progressive Restoration for Multi-Degraded Images. *IEEE Transactions on Image Processing*, 4982-4997. https://doi.org/10.1109/TIP.2024.3451939
[17] Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., & Chen, Z. (2018). Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Computers in biology and medicine*, *99*, 133-141. https://doi.org/10.1016/j.compbiomed.2018.06.010
[18] Lyu, Q., Shan, H., Steber, C., Helis, C., Whitlow, C., Chan, M., & Wang, G. (2020). Multi-contrast super-resolution MRI through a progressive network. *IEEE transactions on medical imaging*, *39*(9), 2738-2749. https://doi.org/10.1109/TMI.2020.2974858
[19] Feng, C. M., Fu, H., Yuan, S., & Xu, Y. (2021). Multi-contrast MRI super-resolution via a multi-stage integration network. *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, 140-149. https://doi.org/10.1007/978-3-030-87231-1_14
[20] Li, G., Lv, J., Tian, Y., Dou, Q., Wang, C., Xu, C., & Qin, J. (2022). Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20636-20645. https://doi.org/10.1109/CVPR52688.2022.01998
[21] Liu, Y., Zhang, M., Jiang, B., Hou, B., Liu, D., Chen, J., & Lian, H. (2023). Flexible alignment super-resolution network for multi-contrast magnetic resonance imaging. *IEEE Transactions on Multimedia*, *26*, 5159-5169. https://doi.org/10.1109/TMM.2023.3330085
[22] Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M. J. et al. (2018). fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.
[23] Xuan, K., Sun, S., Xue, Z., Wang, Q., & Liao, S. (2020). Learning MRI *k*-space subsampling pattern using progressive weight pruning. *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, 178-187. https://doi.org/10.1007/978-3-030-59713-9_18
[24] Feng, C. M., Yan, Y., Yu, K., Xu, Y., Fu, H., Yang, J., & Shao, L. (2024). Exploring separable attention for multi-contrast MR image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(9), 12251-12262. https://doi.org/10.1109/TNNLS.2023.3253557
[25] Zou, B., Ji, Z., Zhu, C., Dai, Y., Zhang, W., & Kui, X. (2023). Multi-scale deformable transformer for multi-contrast knee MRI super-resolution. *Biomedical Signal Processing and Control*, *79*, 104154. https://doi.org/10.1016/j.bspc.2022.104154
[26] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 136-144. https://doi.org/10.1109/CVPRW.2017.151
[27] Neubert, A., Bourgeat, P., Wood, J., Engstrom, C., Chandra, S. S., Crozier, S., & Fripp, J. (2020). Simultaneous super-resolution and contrast synthesis of routine clinical magnetic resonance images of the knee for improving automatic segmentation of joint cartilage: data from the osteoarthritis initiative. *Medical Physics*, *47*(10), 4939-4948. https://doi.org/10.1002/mp.14421
[28] Zhang, H., Wang, Q., Shi, J., Ying, S., & Wen, Z. (2025). Deep unfolding network with spatial alignment for multi-modal mri reconstruction. *Medical Image Analysis*, *99*, 103331. https://doi.org/10.1016/j.media.2024.103331
[29] Wang, Q., Wen, Z., Shi, J., Wang, Q., Shen, D., & Ying, S. (2024). Spatial and Modal Optimal Transport for Fast Cross-Modal MRI Reconstruction. *IEEE Transactions on Medical Imaging*, *43*(11),3924-3935 https://doi.org/10.1109/TMI.2024.3406559

Contact information:

Yan JIN
Tiangong University, School of Computer Science and Technology,
Tianjin 300387, China
E-mail: jinyan@tiangong.edu.cn