Super-Resolution Network Based On Spatial and Frequency Domains for Multi-Contrast Magnetic Resonance Imaging

Ying Li

***Abstract*—Magnetic Resonance Imaging(MRI) has a wide range of applications in the medical field, such as diagnosis, treatment, pathological research,etc.However,due to hardware limitations, obtaining high-quality MR images is often challeng ing in clinical practice. Therefore, reconstructing high-quality MRimages through partial images has important significance in medical research. Existing super-resolution methods usually use multi-contrast MRI to reconstruct MR images. However, existing methods usually use single-scale MR images for reconstruction and do not combine with the specificity of MR images. To address this issue, we propose a multi-scale feature transfer network(SFSR) based on spatial and frequency domains, which comprises four components, including the shallow feature extrac tor, and the Multi-Scale Frequency Attention Block(MFAB), and the Multi-Scale Spatial Attention Block(MSAB), and the Multi Scale Fusion Block(MSFB).Firstly, we utilize the shallow feature extractor to extract features at three scales from both the target and reference images. These features are then separately fed into the Multi-Scale Frequency Attention Block and the Multi-Scale Spatial Attention Block to align the features. Finally, the Multi Scale Fusion Block are employed to fuse the aligned features across different scales.Extensive experiments on IXI and FastMRI datasets show that SFSR achieves the most competitive results over state-of-the-art approaches.**

***Index Terms*—About four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

Magnetic resonance imaging (MRI) is one of the most widely used medical imaging modalities. High-resolution med ical images(HR) can provide detailed texture and structural information for doctors to accurately diagnose and quanti tatively analyze patient conditions. However, due to various factors such as hardware conditions, reconstruction algorithms, acquisition time, physiological motion, and patient-acceptable radiation dose, it is difficult to obtain high-quality MR images while taking into account the patient's physical health.Super resolution reconstruction technology can improve image qual ity without changing external hardware and is often used for post-processing of medical images [1] [2]. In clinical practice, Multi-contrast images with the same anatomical structure can be obtained through different settings, including T1 weighted images (T1), T2 weighted images, proton density weighted im ages (PD) and fat-suppressed proton density weighted images (FS-PD), which can provide complementary information to each other [3].Specifically,

Manuscript received May 10, 2025

Ying Li, School of Software, Tiangong University, Tianjin, China

T1 is used to obtain morphological and structural features with shorter repetition and echo times than T2, T2 is used to illustrate edema and inflammation, PD is used to provide articular cartilage structure with high signal to-noise ratio, FS-PD suppresses fat signals and enhances the visibility of tissue structure [4].Therefore, T1 and PD images with low cost but rich texture can be used to provide supplementary texture information for T2 or FS-PD images for super-resolution.At present, there are many methods for multi contrast reconstruction, Zeng et al [5]. Convolutional neural networks are used to simultaneously perform single-contrast and multi-contrast super-resolution. Lyu et al. introduced a progressive network based on GAN to reconstruct multi contrast MR images [6]. Feng et al. used a multi-level feature fusion mechanism for multi-contrast SR [7].Li et al. employed multi-scale context matching and aggregation schemes, as well as gradually interactive aggregation of multi-scale matching features [8]. Liu et al. used a dual-branch transformer network to reconstruct multi-contrast MR images [9]. Although these methods have high effectiveness, we still face the following challenges: (1) How to effectively align the features of the reference image and the target image (2) How to better fuse the aligned features.Many existing methods [6] [7]directly upsample low resolution images with reference images(Ref) for feature extraction and fusion, ignoring the presence of different features between images of different scales, and do not process MR images in the frequency domain. To address these issues, we propose a multi-scale and multi-contrast MRI super-resolution framework that combines spatial and frequency domains, called SFSR. Our contributions can be summarized as follows:

- We propose a Multi-scale Spatial Attention Block based on Tranformer(MSAB) and Multi-scale Frequency Attention Block based on Tranformer(MFAB). Specifically, we align the features of multi-contrast MR images of different scales in the spatial and frequency domains. Due to the characteristics of MR images, dividing MR images into real and imaginary parts in the frequency domain can more accurately align the features.
- In order to effectively fuse aligned multi-scale features, we propose a multi-scale feature fusion block that com bines channel selection block.
- Wedesign a dual-branch multi-scale transformer network that combines spatial and frequency domains, which can effectively align and fuse features of multi- contrast MR images for super-resolution reconstruction of MRI im ages. After extensive experiments, our method has been proven to be significantly superior to other MRI super resolution methods on the IXI and FastMRI datasets.

II. RELATED WORK

A. Single Image Super-Resolution

Bicubic interpolation and B-spline algorithm are the most commonly used upsampling interpolation methods.Recently, the single image super-resolution (SISR) method based on deep learning has shown excellent results in MRI image super resolution. Qui et al [10]used convolutional neural networks (CNN) for knee joint MRI SR. Christian et al. [11] adopt Generative Adversarial Network (GAN) in super-resolution tasks. Li et al. [12]used the attention mechanism and cyclic loss in Generative Adversarial Network for pelvic image SR reconstruction.Wang et al. [16] further proposed an enhanced generator and discriminator, achieving more perceptually com petitive results. Liang et al. [14] used a closed form Laplace pyramid to accelerate MRI super-resolution tasks.Recently, methods of knowledge distillation [15] [16] and diffusion [17] [18] have been frequently applied in MRI super-resolution tasks. Although these effects have shown excellent results in super-resolution reconstruction, like interpolation methods, SISR methods often introduce artifacts into the reconstructed high-resolution images, which can interfere with doctor di agnosis and lead to misdiagnosis. Therefore, single image super-resolution methods are not suitable for MR images reconstruction.

B. Single Image Super-Resolution

Compared with the single image super-resolution method (SISR), the reference image based super-resolution method(Ref-SR) uses additional high-resolution reference images as references to perform super-resolution on low resolution images(LR). This method is easier to obtain accurate texture information and reduces the probability of artifacts. For MR images, high-resolution contrast images are often used as reference images.CrossNet [19]estimates the flow between Ref and LR images at multi-scales and distorts Ref features based on the flow. However, traffic is obtained through pre trained networks, resulting in high computational complexity and inaccurate estimation.SSEN [20]utilizes deformable convolution to align and extract features of Ref, expanding the receptive field.SRNTT [21] extracted features between Ref and LR for matching in pretrained VGG, and transferred texture information from the reference image to assign low-resolution details based on similarity scores. C2matching [22] introduces contrastive correspondence networks and teacher-student correlation distillation to align images at the pixel level.TTSR [23] retains the idea of cross-attention and adds a soft-attention module. MASA [24] considered potential differences and reduced computational costs. FASR [9] uses a single pyramid alignment module and a multi-pyramid alignment module to align features, solving the problem of scale matching between LR and Ref.However, these methods only align features in the spatial domain. In this paper, we use a combination of spatial and frequency domains to align features.

III. METHOD

A. Overview

During the acquisition process of MR images, a series of multi-contrast images are generated simultaneously. Therefore, high-resolution PD images can be directly used as reference images(Ref) to provide details for the super-resolution of low-resolution and high cost T2 images.And the raw data of MR images is divided into two parts: the real part and the imaginary part. Usually, in the frequency domain, the real part contains the edge information between the tissues in the MR image, while the imaginary part contains the structural information inside the tissues, as shown in the figure. Therefore, combining the characteristics of MR images in the frequency domain mentioned above, we propose a new fea ture alignment super-resolution network based on spatial and frequency domain rates. PD images and T2 images are used as reference images and low resolution images, and feature alignment is performed in spatial and frequency domain rates, respectively.

The architecture of SFSR is shown in the fig. 2. Specifically, REF and REF ↓↑ represent PD images and PD images that have been downsampled and upsampled with the same factor, respectively. LR ↑ represents T2 images that have been upsam pled to the same scale as REF through bicubic interpolation. After upsampling, the consistency of LR and REF sizes can be ensured.

Overall, SFSR can be divided into three parts: the shallow feature extractor, dual-domain feature alignment block, and multi-scale fusion block. The shallow feature extractor aims to extract features of different scales from LR and Ref, facilitating subsequent modules to align features at different scales. The dual-domain feature alignment block consists of the multi-scale frequency attention block, and the multi-scale spatial attention block, which align feature images of different scales in the spatial and frequency domains, respectively. Finally, a multi-scale fusion module is used to fuse the aligned features with the LR.

B. shallow feature extractor

We choose the CNN [25] network based on the VGG architecture as the shallow feature extractor, responsible for extracting three-level features from MR images, as shown in Fig. 2.

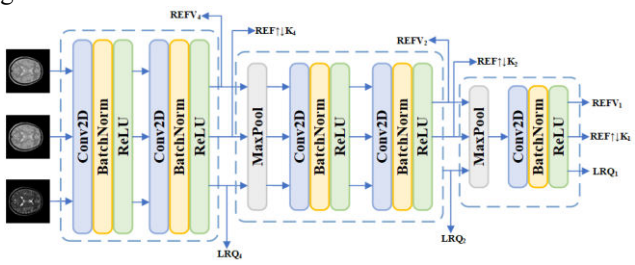

Fig. 2 Architecture diagram of shallow feature extractor (SFE)

The shallow feature extractor consists of three con volutional modules, where the first two convolutional blocks contain two convolutional layers composed of convolution operations, BN layers, and ReLU functions, and the last convolutional block only contains one convolutional layer.Due to the higher resolution of the input Ref compared to the LR, in order to establish the feature correspondence between the Ref and the LR, we need to upsample the LR, and the reference image needs to be degraded through downsampling and upsampling to obtain a feature distribution relationship similar to the LR. Therefore, the images input to the multi feature extractor are the upsampled

LR ↑ , the downsampled and upsampled Ref ↑ ↓ , and the original Ref.And use the extracted shallow features of the same level for q, k, and v in subsequent Transformer block.

C. *Multi-Scale Spatial Attention Block(MSAB)*

Drawing inspiration from TTSR [23], we developed a Multi Scale Spatial Attention Block(MSAB).As shown in Fig. 3 To fully utilize the interrelationships between different levels of features, we use the deepest level feature of ILR↑as the query (q) in the MSAB, which contains a large number of

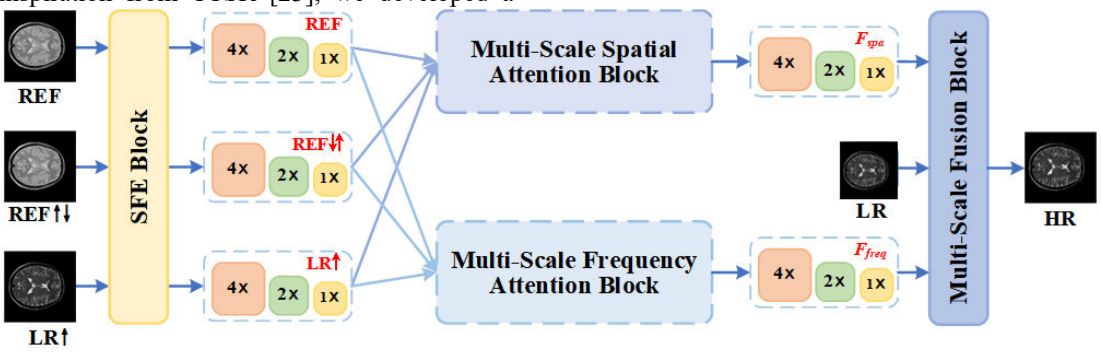

Fig. 1 Overview of SFSR. SFSR consists of three parts, the shallow feature extractor, the dual-domain feature alignment module, and the multi-scale fusion block. The input REF, REF ↑↓, and LR are first processed through the shallow feature extractor block to obtain multi-scale feature maps. Then, the dual-domain feature alignment module aligns features of different scales in the spatial and frequency domains respectively, and outputs the aligned features Fspa and Ffraq

semantic features. Then, we unfold the Ref↑↓ and Ref of the three levels separately as keys (k), k={$k^{4\times}$, $k^{2\times}$, $k^{1\times}$} and values (v), v={$v^{4\times}$, $v^{2\times}$, $v^{1\times}$}, which respectively contain the structural features, edge features, and semantic features of Ref. Calculate the correlation between q and k and activate it with Softmax.

$$RM_{spa}^{n\times} = Softmax\left(\left\langle q, k^{n\times}\right\rangle\right) \quad n=\{1,2,4\} \tag{1}$$

$$H_{spa}^{n\times}(i) = argmax RM_{spa}^{n\times}(i,j) \quad n=\{1,2,4\} \tag{2}$$

$$T_{spa}^{n\times} = v^{n\times}\left(H_{spa}^{n\times}(i)\right) \quad n=\{1,2,4\} \tag{3}$$

Since v is obtained by recombining the Ref image, we also add the aligned features of q={$q^{4\times}$, $q^{2\times}$, $q^{1\times}$} at three scales and pass through multiple residual blocks to supplement the structural, edge, and semantic features of LR ↑ that are missing in the Ref into the feature map. The final feature map obtained can be obtained by the following formula:

$$F_{spa}^{n\times} = T_{spa}^{n\times} + q^{n\times} \quad n=\{1,2,4\} \tag{4}$$

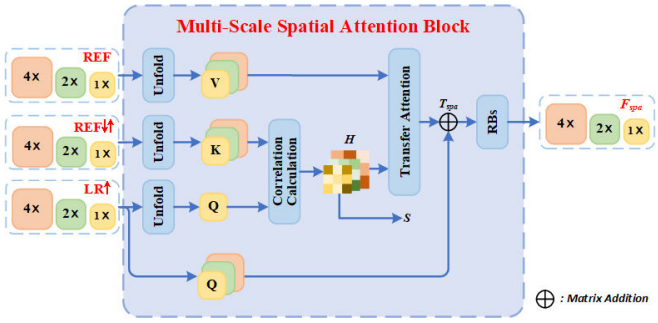

Fig. 3 The Multi-Scale Spatial Attention Block(MSAB) is shown in the above figure. After calculating the correlation matrix between q and k, the attention matrix is used to guide feature alignment, and the aligned features are added to q for feature supplementation. After passing through the residual block, the final feature is obtained.

D. *Multi-Scale Spatial Attention Block(MSAB)*

Feature alignment on a single scale only may lead to misalignment of LR images and Ref images, and feature alignment in the frequency domain can better preserve the edge information in the image, which makes the edge boundaries of different tissues in MRI images clearer and more conducive to the diagnosis of lesions. Inspired by [26], we built a Multi-Scale Frequency Attention Block(MFAB) based on frequency domain rate, as shown in Fig. 4. In order to convert the image into real and imaginary parts, we need the magnitude and phase information of the image, and in general, the magnitude can be obtained by the following equation,

$$\left|Z_{j,t}\right|e^{i\theta_{j,t}} = Z_{j,t}e^{i\theta_{j,t}} \quad Z_{j,t} > 0 \tag{5}$$

$$\left|Z_{j,t}\right|e^{i\theta_{j,t}} = Z_{j,t}e^{i(\theta_{j,t}+\pi)} \tag{6}$$

where $Z_{j,t}$,$\Theta_{j,t}$ denotes the magnitude $Z_{j,t}$ and phase $\Theta_{j,t}$ of the t-th element, Z_j represents the input vector. In order to capture the specific attributes of each input separately, we use the estimation module Θ to generate the phase information based on the input feature Xj , i.e., θj = Θ(xj , Wθ), where W_θ denotes the learnable parameters. In summary, the component can be specifically obtained by the following equation:

$$Z_{Real} = \left|Z_j\right| \otimes \cos\theta_j \tag{7}$$

$$Z_{Imag} = \left|Z_j\right| \otimes sin\theta_j \tag{8}$$

We compute the real and imaginary parts of the inputs LR↑, Ref↑↓, and Ref, respectively, and expand them as q_{real}, q_{imag}, k_{real}, k_{imag}, v_{real}, v_{imag}.

The computation of the similarity matrix is similar to Eq. (1) and can be expressed as:

$$RM_{real}^{n\times} = Soft\max\left(\left\langle q_{real}^{n\times}, k_{real}^{n\times}\right\rangle\right) \quad n=\{1,2,4\} \tag{9}$$

$$RM_{imag}^{n\times} = Soft\max\left(\left\langle q_{imag}^{n\times}, k_{imag}^{n\times}\right\rangle\right) \quad n=\{1,2,4\} \tag{10}$$

The correlation matrices obtained for the x1 and x2 sizes in the real and imaginary parts, respectively, are subjected to the up-sampling operation and summed with the correlation matrix for the x4 size to obtain the final attention matrices in the real and imaginary parts, which can be expressed as follow:

$$RM_{real} = \cup\left(RM_{real}^{1\times}\right) + \cup\left(RM_{real}^{2\times}\right) + RM_{real}^{4\times} \tag{11}$$

$$RM_{imag} = \cup\left(RM_{imag}^{1\times}\right) + \cup\left(RM_{imag}^{2\times}\right) + RM_{imag}^{4\times} \tag{12}$$

$$H_{real}^{n\times} = argmax RM_{real}(i,j) \quad n=\{1,2,4\} \tag{13}$$

$$H_{imag}^{n\times} = argmaxRM_{imag}(i,j) \quad n=\{1,2,4\} \tag{14}$$

Similarly, we perform feature alignment in the real and imaginary parts of the image, respectively. The alignment features in the real and imaginary parts of images T1 and T2 can be formulated as follows:

$$T_{real}^{n\times} = v^{n\times}\left(H_{real}^{n\times}(i)\right) \quad n=\{1,2,4\} \tag{15}$$

$$T_{imag}^{n\times} = v^{n\times}\left(H_{imag}^{n\times}(i)\right) \quad n=\{1,2,4\} \tag{16}$$

Finally we perform a magnitude operation on the real and imaginary parts of the aligned features and the attention matrix to obtain the final attention matrix and aligned features, which can be formulated as:

$$H_{freq}^{n\times} = \sqrt{\left(H_{real}^{n\times}\right)^2 + \left(H_{imag}^{n\times}\right)^2} + F_{Ref}^{n\times} \quad n=\{1,2,4\} \tag{17}$$

$$F_{freq}^{n\times} = \sqrt{\left(T_{real}^{n\times}\right)^2 + \left(T_{imag}^{n\times}\right)^2} + F_{Ref}^{n\times} \quad n=\{1,2,4\} \tag{18}$$

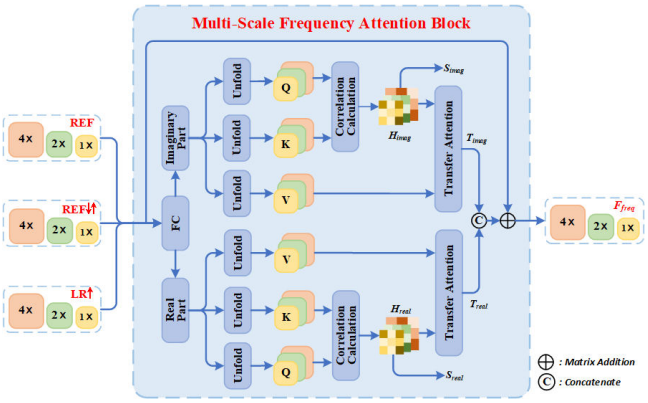

Fig. 4. Multi-Scale Frequency Attention Block is shown in the above figure. Mapping q and k to the frequency domain, the correlation between q and k is computed on the real and imaginary parts respectively to get the correlation matrix H_{real} and H_{imag} to guide v for feature alignment to obtain the aligned features T_{real} and T_{imag} . The to the final aligned feature after calculating through the magnitude .

E. Multi-scale Feature Fusion Block(MSFB)

Commonly used fusion methods such as concatenation (splicing) and averaging often lead to insufficient interaction of multi-scale image information. To address this problem, we propose a multi-scale fusion block fusing features aligned in the spatial and frequency domain rates respectively. As shown in Fig. 5, it's input is a low-resolution image(LR).

$$S_{fus}^{n\times} = \frac{\left(S_{spa}^{n\times} + \sqrt{\left(S_{real}^{n\times}\right)^2 + \left(S_{imag}^{n\times}\right)^2}\right)}{2} \quad n=\{1,2,4\} \tag{19}$$

In addition, for better feature interaction between different scales, we also designed the channel selection block as in Fig. 6. Specifically, in CSB, Our channel selection block can be specifically obtained from the following equation:

$$F_{CA}^{2\times} = CA\left(F_{fus}^{2\times}\right)\cdot CA\left(F_{fus}^{2\times} + F_{fus}^{1\times}\right)\cdot F_{fus}^{2\times} + F_{fus}^{2\times} \tag{20}$$

$$F_{CA}^{1\times} = CA\left(F_{fus}^{1\times}\right)\cdot CA\left(F_{fus}^{2\times} + F_{fus}^{1\times}\right)\cdot F_{fus}^{1\times} + F_{fus}^{1\times} \tag{21}$$

$$F_{out}^{2\times} = F_{CA}^{1\times} + F_{CA}^{2\times} \tag{22}$$

In channel selection block that performs 4x zoom, the second selection module we select 1x scale and 4x scale feature maps, 2x scale and 4x scale feature maps for feature selection.

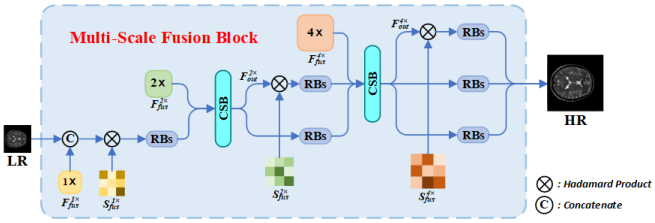

Fig. 5. The multi-scale feature fusion block is shown in the above figure. In order to fully fuse the features on multiple scales of the image, we fused in the multi-scale fusion module $F^{1\times}$, the $F^{2\times}$ and $F^{4\times}$ Features. In order to fully integrate the features between different scales, we introduce a channel selection module. as shown in Fig. 6.

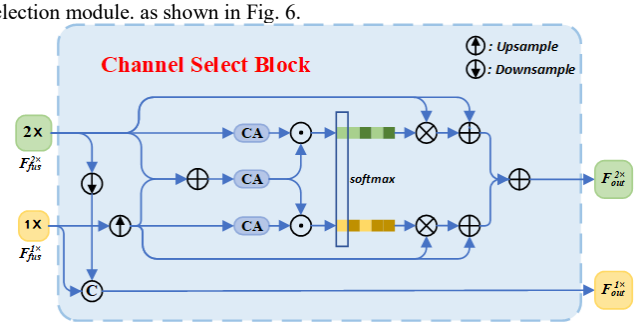

Fig. 6. the channel selection block is shown in the figure above. the CA is the channel attention block.

IV. Loss

L1 Loss: In order to compute the difference between the Isr and GT images, a method of computing the pixel-level difference between the two images can be used. This method directs our network to focus more on the detailed part of the image. Previous experiments have demonstrated his effectiveness. We choose the L1 loss as the most pixel-level loss for the network, which is given as

$$L_1 = \left|I_{SR} - I_{GT}\right| \tag{23}$$

V. EXPERIMENTS

A. Datebase

(1) IXI dataset: The IXI dataset[29] contains MRIs of 578 patients, which include T1, T2 and PD-weighted images, among others. Among them, T1, T2 and PD images are images of different modalities under the same anatomical structure. Our experiments use downsampled T2 images as input, original T2 images as GT, and PD images as reference images. Specifically, the specific sizes of both T2 and PD images in the IXI dataset are 256 × 256 × 3. In order to achieve 2x scale and 4x scale of super-resolution results, we downsampled T2 at the corresponding multiplicity. Before training, all images were normalized to the range [-1, 1]. We selected 8000 pairs and 800 pairs of T2 and PD images in the dataset as the training and validation sets.

(2) FastMRI Dataset: The FastMRI dataset [30] [31] con tains four types of data from knee MRI and brain MRI. We selected the knee part of the dataset, cropped it to 256×256 size in k-space and used the inverse Fourier transform to transform the cropped data to the image domain to produce the original image, and downsampled the PDFS image according to the corresponding magnification to produce the LR image, and other operations were the same as those of the IXI dataset.We selected 216 pairs and 20 pairs of PD and PDFS images in the dataset as the training and validation sets

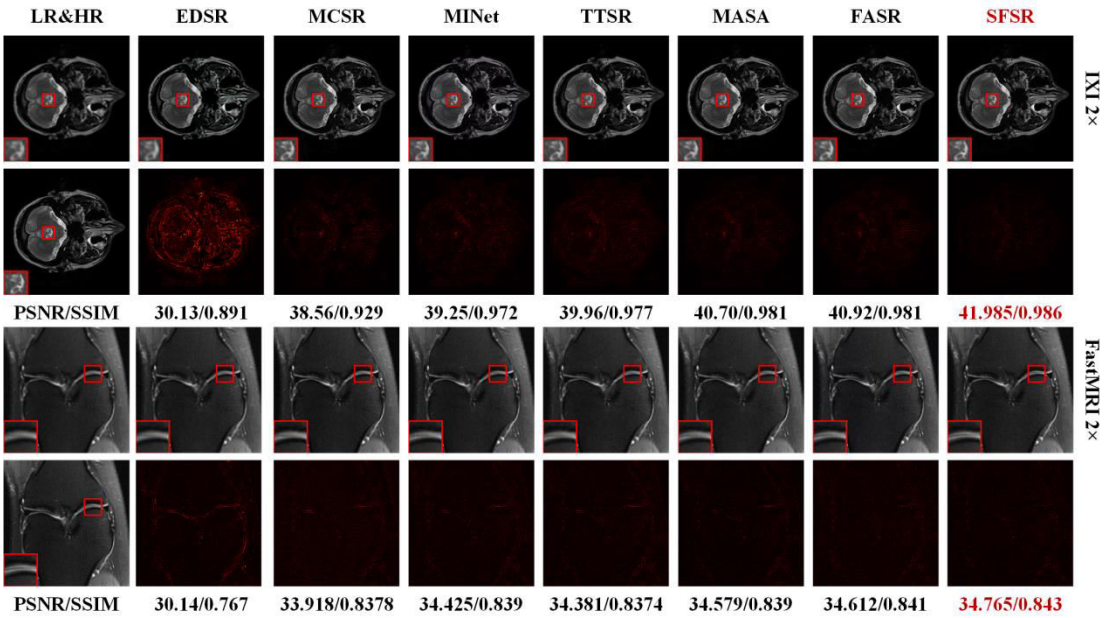

Fig. 7. The above figure presents a visual comparison of the outputs and residual maps from different models at 2 × magnification on the IXI and FastMRI datasets. The residual map quantifies the absolute pixel differences between the model's output and the corresponding ground truth (GT). LR stands for low resolution, created through downsampling. HR denotes the ground truth.

(3) IMPLEMENTATION DETAILS: We implement our model us ing one NVIDIA RTX A6000 GPU with single-card 48GB memory. Our model is trained using the Adam optimizer for 50 epochs, and a learning rate is set to 1e-5.

B. Objective and Subjective Comparison

To demonstrate that our proposed model achieves better super-resolution results, we performed experiments on the IXI dataset and FastMRI dataset, respectively.

(1) IXI Dataset: We compared our results with previous work on 2× scale and 4× scale on the IXI dataset, including EDSR, MCSR, MINet, TTSR, MASA, and FASR, and can conclude that our method achieves the best results compared to existing methods.Specifically, MCSR also utilizes multi modal images as a reference map, but lacks accurate fusion of features from the reference image, and falls short of our approach.MINet utilizes the reference map to learn hierarchi cal feature representations from multiple convolutional stages for each image of different contrasts, and achieves better results.TTSR and MASA align low resolution images of ref erence images at the semantic level, achieving more effective results.FASR used a multi-contrast flexible alignment network and achieved significant results. Due to our feature alignment in the frequency domain, our method outperforms FASR and MASA.In the fig. 7, we show the ISR of the 4× scale output and the corresponding residual plots to demonstrate the best visualization of our model. The residual plots exhibit the absolute pixel value difference between the resolution results and the GT, as shown in Table. I, where our method achieves the best results.

TABLE 1
QUANTITATIVE COMPARISON OF IXI DATASETS

DATASET	IXI			
SCALE	×2		×4	
METHODS	METRIC			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
BICUBIC	24.21	0.769	20.32	0.672
EDSR	30.11	0.891	29.52	0.85
MCSR	38.55	0.928	33.81	0.891
MINET	39.23	0.972	37.51	0.933
TTSR	39.95	0.977	37.89	0.961
MASA	40.69	0.981	38.26	0.971
FASR	40.91	0.981	38.52	0.974
OUR	41.984	0.986	39.03	0.9746

(2) FastMRI Dataset: We also compared our method on FastMRI dataset with other methods. The evaluation metrics, Table. II, show that our method outperforms the other methods. The corresponding residual plots are shown in Fig. 8.

TABLE 2
QUANTITATIVE COMPARISON OF FASTMRI DATASETS

DATASET	FASTMRI			
SCALE	×2		×4	
METHODS	METRIC			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
BICUBIC	26.9	0.613	18.52	0.605
EDSR	30.12	0.767	24.28	0.684
MCSR	33.920	0.837	29.74	0.731
MINET	34.427	0.839	31.60	0.7371
TTSR	34.380	0.837	31.86	0.743
MASA	34.582	0.840	31.93	0.745
FASR	34.615	0.841	32.24	0.752
OUR	34.768	0.844	32.50	0.759

C. Ablation Study

In order to verify the effectiveness of the proposed compo nents, we conducted several experiments on IXI 4× scale to verify the effectiveness of the components as shown in Table.It includes (1) Multi-Scale Spatial Attention Block(MSAB)(2) Multi-Scale Frequency Attention Block(MFAB) (3) Multi Scale Feature Block(MSFB).The quantitative results are shown in Table 3. And the visual results are shown in Fig. 9 In the MSAB-free experiment, we

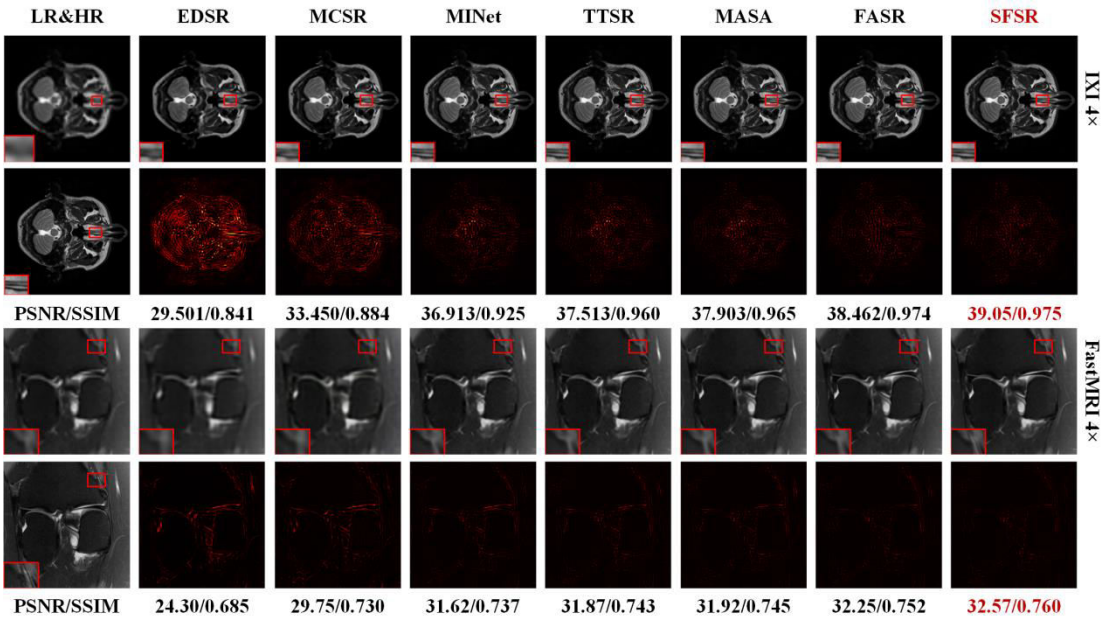

Fig. 8. The above figure presents a visual comparison of the outputs and residual maps from different models at 4× magnification on the IXI and FastMRI datasets. The residual map quantifies the absolute pixel differences between the model's output and the corresponding ground truth (GT). LR stands for low resolution, created through downsampling. HR denotes the ground truth.

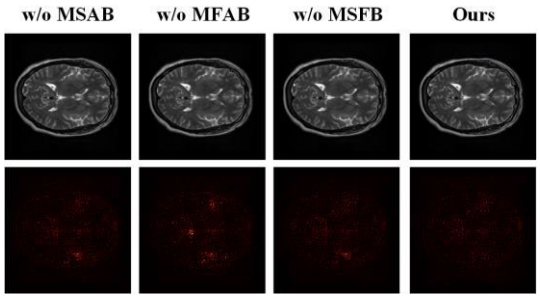

Fig. 9. The above figure presents a visual comparison of the output and residual images from the ablation experiment conducted at 4x magnification on the IXI dataset

TABLE 3
MODULE ABLATION EXPERIMENTS BASED ON THE IXI DATASET

SETTING	SPAATTEN	FREQATTEN	MSFB	PSNR	SSIM
MSAB		√	√	38.01	0.971
MFAB	√		√	37.08	0.966
MSFB	√	√		38.38	0.973
ALL	√	√	√	39.03	0.974

removed the null-domain-based alignment module and kept only the frequency-domain-based feature alignment module, which resulted in a decrease in both PSNR and SSIM metrics compared to the full structure. We also removed the Multi-Scale Frequency Attention Block, and it can be seen that the metrics by a significant decrease, indicating that the frequency-domain based feature alignment module has a more significant effect on the alignment of fea tures.In the experiments without MSFB module, we use con volutional layer and up-sampling operation instead of MSFB fusion module, and we can conclude that compared with direct fusion, our MSFB module can fuse features between different scales more effectively.

VI. CONCLUSION

In this paper, we propose a multi-scale MRI super-resolution network based on frequency and spatial domains, where the inputs of the network are multi-contrast MR images that provide high-resolution HR images.First, we propose a dual domain feature alignment module, including a Multi-Scale Spatial Attention Block(MSAB) and a Multi-Scale Frequency Attention Block(MFAB).The MSAB performs feature align ment in the spatial domain and the MFAB performs feature alignment in the frequency domain.In addition, the multi-scale feature fusion module (MSFB) can adequately align multi scale features, and a channel selection block is proposed to better fuse features at different scales.Extensive experiments on IXI and FastMRI datasets show that our method achieves good results in both quantitative and qualitative evaluations.

REFERENCES

[1] E. Plenge, D. H. Poot, M. Bernsen, G. Kotek, G. Houston, P. Wielopolski, L. van der Weerd, W. J. Niessen, and E. Meijering, "Super-resolution methods in MRI: can they improve the trade-off between resolution,signal-to-noise ratio, and acquisition time?" Magnetic Resonance in Medicine, vol. 68, no. 6, pp. 1983–1993, 2012.

[2] E. Van Reeth, I. W. Tham, C. H. Tan, and C. L.Poh, "Super-resolution in magnetic resonance imaging: a review," Concepts in Magnetic Resonance Part A, vol. 40, no. 6, pp. 306–325, 2012.

[3] Wei Chen, Jun Zhao, Yaming Wen, Bin Xie, Xuanling Zhou,Lin Guo,Liu Yang, Jian Wang, Yongming Dai, and DaiquanZhou. Accuracy of 3-tmri using susceptibility-weightedimaging to detect meniscal tears of the knee. Knee Surgery,Sports Traumatology, Arthroscopy, 23(1):198–204,2015.

[4] W. Chen, J. Zhao, Y. Wen, B. Xie, X. Zhou, L. Guo,L. Yang, J. Wang, Y. Dai, and D. Zhou, "Accuracy of 3-t mri using susceptibility-weighted imaging to detect meniscal tears of the knee," Knee Surgery, Sports Traumatology, Arthroscopy, vol. 23, pp. 198–204, 2015.8

[5] Kun Zeng, Hong Zheng, Congbo Cai, Yu Yang, Kaihua Zhang, and Zhong Chen. 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, 2018.

[6] Lyu, Q., Shan, H., Steber, C., Helis, C., Whitlow, C., Chan, M., Wang,G.: Multi-contrast super-resolution mri through a progressive network.IEEE transactions on medical imaging 39(9), 2738–2749 (2020)

[7] Feng, C.M., Fu, H., Yuan, S., Xu, Y.: Multi-contrast mri super-resolution via a multi-stage integration network. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg,France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 140–149.Springer (2021)

[8] Li, G., Lv, J., Tian, Y., Dou, Q., Wang, C., Xu, C., Qin, J.: Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20636–20645(2022)

[9] Y. Liu et al., "Flexible Alignment Super-Resolution Network for Multi-Contrast Magnetic Resonance Imaging," in IEEE Transactions on Multi-media, vol. 26, pp. 5159-5169, 2024, doi: 10.1109/TMM.2023.3330085.
[10] Lixin Zheng. Super-resolution reconstruction of knee mag-netic resonance imaging based on deep learning. Computer methods and programs in biomedicine, 187:105059, 2020. 2
[11] C. Ledig, L. Theis, F. Husz´ar, J. Caballero, A. Cunningham, A.Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., "Photo-realistic single image super-resolution using a generative adversarial network,"in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
[12] Guangyuan Li, Jun Lv, Xiangrong Tong, Chengyan Wang,and Guang Yang. High-resolution pelvic mri reconstruction using a generative adversarial network with attention and cyclic loss. IEEE Access,9:105951–105964, 2021. 2
[13] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C.Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in Proceedings of the European Conference on Computer Vision (ECCV) workshops, 2018, pp. 0–0.
[14] J. Liang, H. Zeng, and L. Zhang, "High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9392–9400.
[15] Q. Gao, Y. Zhao, G. Li, and T. Tong, "Image super- resolution using knowledge distillation," in Asian Conference on Computer Vision.Springer, 2018, pp. 527–541.
[16] W. Lee, J. Lee, D. Kim, and B. Ham, "Learning with privileged information for efficient image super-resolution," in European Conference on Computer Vision. Springer, 2020, pp. 465–482.
[17] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi,"Image super-resolution via iterative refinement," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
[18] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," Neuro-computing, vol. 479, pp. 47–59, 2022.
[19] H. Zheng, M. Ji, H. Wang, Y. Liu, and L. Fang,"Crossnet: An end-to-end reference-based super resolution network using cross-scale warping," in Proceedings of the European Conference on Computer Vision (ECCV),2018, pp. 88–104.
[20] G. Shim, J. Park, and I. S. Kweon, "Robust reference-based super-resolution with similarity-aware deformable convolution," in Proceedings of the IEEE/CVF Confer-ence on Computer Vision and Pattern Recognition, 2020,pp. 8425–8434.
[21] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7982–7991.
[22] Y. Jiang, K. C. Chan, X. Wang, C. C. Loy, and Z. Liu, "Robust reference-based super-resolution via c2-matching," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2103–2112.
[23] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution,"in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020, pp. 5791–5800.
[24] L. Lu, W. Li, X. Tao, J. Lu, and J. Jia, "Masa-sr: Matching acceleration and spatial adaptation for reference- based image super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6368–6377.
[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition,"arXiv preprint arXiv:1409.1556, 2014.
[26] Tang Y, Han K, Guo J, et al. An image patch is a wave: Phase-aware vision mlp[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10935-10944.
[27] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017,pp.3883–3891.
[28] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015,pp.769–777.
[29] IXI. [Online]. https://brain-development.org/ixi-dataset.
[30] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley,A. Defazio, R. Stern, P. Johnson, M. Bruno et al., "FastMRI: An open dataset and benchmarks for accelerated MRI," arXiv preprint arXiv:1811.08839, 2018.
[31] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio,M. Parente, K. J. Geras, J. Katsnelson,H. Chandarana et al., "fastMRI: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning," Radiology:Artificial Intelligence, vol. 2, no. 1, p. e190007, 2020