Applied Intelligence (2025) 55:45
https://doi.org/10.1007/s10489-024-06008-6

A multi-domain feature alignment and hierarchical cross feature enhancement network for under-sampled magnetic image reconstruction

Qiaohong Liu[1] · Xiaoxiang Han[2] · Yang Chen[3]

Accepted: 30 September 2024

Abstract

Magnetic resonance imaging (MRI) is widely used in clinical diagnosis due to its high resolution and non-invasive scanning capabilities. However, long scanning times limit its development. To reduce acquisition time and obtain high-quality reconstructed images, a novel multi-domain MRI reconstruction network that fully utilizes the image domain, k-space, and wavelet domain is proposed. This network includes a parallel convolutional neural network (CNN) with k-space and wavelet domain branches, as well as a U-shaped image domain network. Following the parallel dual-domain CNN, a dual-domain feature alignment module aligns features from the k-space and wavelet domains into a unified representation space, mitigating artifact impacts. This design enhances the model's understanding of multi-domain signals and improves generalization. Additionally, in the image domain, a hierarchical cross-feature enhancement module, based on Nested UNet, incorporates two cross-attention modules into different hierarchical skip connections of the Nested U-Net to reduce information propagation loss and enhance feature representation. Deep supervision within the image domain network further boosts the network's performance and robustness. Extensive experiments on two public MRI datasets, FastMRI and CC359, as well as the private clinical dataset, validate the proposed method. Compared to several state-of-the-art deep learning methods, our approach demonstrates good reconstruction performance in both numerical assessments and visual effects.

Keywords MRI reconstruction · Wavelet transform · Multi-domain · Feature alignment · Cross-attention

1 Introduction

Magnetic resonance imaging (MRI) is a widely used diagnostic tool in clinical practice due to its non-invasive nature, lack of ionizing radiation, and ability to capture multiple parameters. However, the speed at which k-space is traversed is limited by both physiological and hardware constraints [1], leading to lengthy data acquisition times. This not only reduces the efficiency of MRI equipment but also causes discomfort for patients and can introduce motion artifacts that degrade image quality. Therefore, developing accelerated algorithms to reconstruct high-quality images from under-sampled k-space data is crucial.

Over the past few decades, several advanced image reconstruction techniques have been developed to accelerate the acquisition speed while producing high-quality magnetic resonance images. These techniques include partial Fourier transformation [2], parallel imaging techniques (PI) [3], compressed sensing (CS) techniques [4], low-rank matrix completion techniques [5], and manifold learning techniques [6], among others. Among these, compressed sensing MRI (CS-MRI) [7] stands out as a groundbreaking method that uses prior information about the image to reconstruct high-quality MR images. Commonly used priors include

Qiaohong Liu and Xiaoxiang Han contributed equally to this work.

✉ Yang Chen
gtlinyer@qq.com

Qiaohong Liu
hqllqh@163.com

Xiaoxiang Han
gtlinyer@163.com

1 School of Medical Instruments, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China

2 School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

3 Algorithm Team, ToolSensing Technologies Co., Ltd, Chengdu 610101, China

Published online: 28 November 2024

total variation, wavelet transformation, low rank, and dictionary learning. However, the optimization iterations required to solve the CS-MRI model are computationally expensive. Additionally, there are limitations such as inadequate priors and inflexible parameters. Particularly with high acceleration factors, CS-MRI methods can produce unwanted artifacts and over-smooth anatomical structures.

Recently, the rapid development and widespread use of deep learning technology [8–11] have brought attention to its potential in accelerating MRI reconstruction. Wang et al. [12] were pioneers in using convolutional neural networks (CNN) for reconstructing under-sampled MR images, providing an end-to-end mapping from zero-filled to fully sampled images. They trained an offline CNN to speed up MRI reconstruction. Following their work, many CNN-based MRI reconstruction methods have been proposed, which can be categorized into two types: single-domain and cross-domain methods. Single-domain methods reconstruct MR images from under-sampled k-space data either in the image domain or k-space domain. On the other hand, cross-domain methods leverage the latent relationships between the image and k-space domains, generally achieving better results than single-domain methods. Notable networks in this field include KIKI-Net [13], W-Net [14], Hybrid-Cascade-Net [15], DD-DLN [16], MD-Recon-Net [17], DIMENSION [18], DIIK-Net [19].

Wavelet transform is widely recognized for its application in traditional CS-MRI. Lusting et al. [4] were the pioneers in introducing CS theory to the MRI field, proposing SparseMRI, which integrates the sparse prior of total variation and wavelet transformation for MRI reconstruction. Compared to Fourier transformation, wavelet transformation offers advantages such as multi-directionality, multi-scale, and multi-resolution capabilities, resulting in better sparse image representation. Additionally, wavelet transformation excels at capturing fine image details, which aids in reconstructing high-quality images. Wang et al. [20] were the first to apply wavelet transformation in a deep learning-based method for MRI reconstruction, combining the image domain, k-space, and wavelet domain within convolutional CNNs. However, wavelet transformation remains infrequently used in deep learning-based methods.

Inspired by the works mentioned above, a novel network for MRI reconstruction based on cross-domain method which combined k-space domain, wavelet domain and image domain is proposed. The present study focuses on improving feature alignment in multi-domain feature fusion, as misalignment of features from different domains introduces biases in subsequent processing. Additionally, another focus is on enhancing the removal of artifacts generated during the reconstruction process. The proposed method fully considers the different feature representation in different domain, which can be utilized to improve the reconstruction performance, as shown in Fig. 1. Aiming to reconstruct the high-quality MRI in low acceleration factor, the proposed network consists of a parallel dual-domain CNN contains k-space and wavelet domains and a U-shaped image domain network. After the parallel dual-domains CNN, a dual-domain feature alignment module (DFAM) is designed to align the features from k-space and wavelet domains into a unified representation space and alleviate the impact of artifacts. This design incorporates the strengths of both the frequency domain and the wavelet domain. Frequency domain signals aid in extracting global features from images, such as textures and edges. Wavelet domain signals help capture both spatial and frequency information simultaneously, making them suitable for extracting local features and abrupt changes in images, and also assist the model in recognizing noise. Therefore, this design contributes to a deeper understanding of image signals by the model, enhancing the model's generalization capability. Furthermore, a hierarchical cross feature enhancement module (HCFEM) is used in image domain to achieve the last reconstructed image. To mitigate information propagation loss and enhance feature representation, the design of this module is based on Nested U-Net [21] and two cross-attention modules are embedded into the different hierarchical skip connection of the Nested U-Net. This design can utilize the contextual information from the encoder to preserve more fine details and edges of the reconstructed image. Deep supervision in image domain

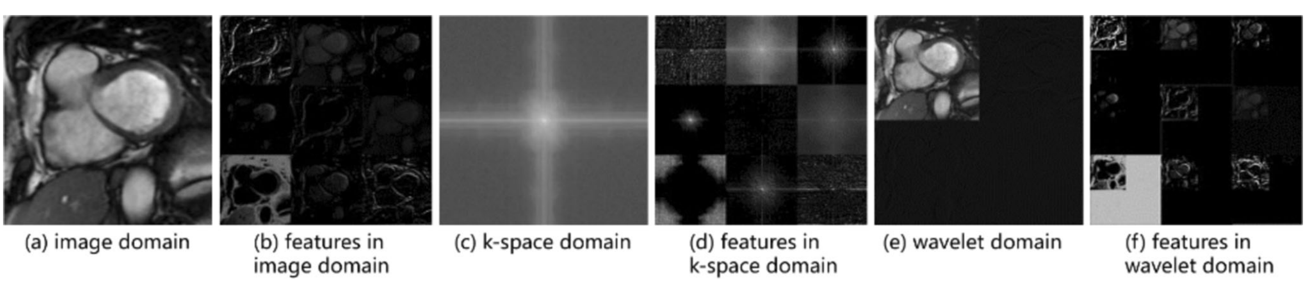


Fig. 1 Images and their features in different domains

network is used to guarantee effective utilization of features at each level and enhance the feature represent ability, while accelerating the network convergence. Overall, the proposed model greatly improves the feature representation capability and enhances the MRI reconstruction performance demonstrated by extensive experiments on three datasets.

The main contributions of this paper are as follows.

1. A novel multi-domain network that combines k-space domain, wavelet domain and image domain, which takes full of the different feature representation in different domains is proposed for fast magnetic resonance reconstruction.
2. A dual-domain feature alignment module (DFAM) which is based on deformable convolution operation is designed to align the two branches features from parallel CNN into a unified representation space. The DFAM can not only fuse the different domain information but also alleviate the artifacts or deformations from different domains.
3. A hierarchical cross feature enhancement module (HC FEM) which can minimize information propagation loss and enhance feature representation in image domain network is introduced reconstruct the final magnetic resonance image. Then, deep supervision is used to promote the network's capability of feature representation and accelerate the network convergence.

2 Related works

2.1 Single domain MRI reconstruction

The first CNN for MRI reconstruction proposed by Wang et al. [12] has promoted the rapid development of deep learning-based reconstruction method. Schlemper et al. [22] designed a deep cascaded of convolutional neural network (DC-CNN) to reconstruct dynamic cardiac MR image in image domain and exploited a data consistency (DC) to fully use the real collected k-space data to remedy the predicted data. Qin et al. [23] proposed a convolutional recurrent neural network (CRNN) which reconstructs high-quality cardiac MR image from highly under-sampled k-space data by jointly utilizing the dependency of time series and the iterative properties of traditional optimization algorithm. Some typical CNN structures such as UNet [24], GAN [25], ResNet [26] are applied to MRI reconstruction in image domain. Aghabiglou et al. [27] firstly introduced the densely connected residual block into UNet in MR image reconstruction make the network deeper and increase the number of model parameters without the consequent training difficulties and vanishing gradient problems. Chatterjee et al. [1] designed a modified ResNet as the network backbone to remove artifacts from the highly under-sampled Cartesian and radial data. Several GAN-based methods were developed for MRI reconstruction, such as DAGAN [28], RefineGAN [29], SEGAN [30], which adopted the UNet-like model as the generators and combined the perceptual, cyclic or structure-enhanced loss functions to recover the image details. However, these methods might not be robust enough for specific types of noise or artifacts. Some general network structures may not be optimized for MRI reconstruction tasks and could be inadequate in restoring certain details. Additionally, some models primarily focus on artifact removal, which may lead to deficiencies in image detail recovery.

With the breakthrough of Transformer [31] in natural language processing, Transformer and its variant version have been applied to computer vision and medical image analysis field. Due to the good global context modeling capability to learn the long-rang dependencies of features effectively, some Transformer-based MRI reconstruction methods are developed. Huang et al. [32] proposed a novel parallel imaging coupled Swin transformer-based model (SwinMR) for fast MRI reconstruction, which utilizes Swin transformer [33] structure to achieve a trade-off for global and local information of images. Furthermore, Huang et al. [34] developed a new Transformer structure for coupling fast MRI problem that coupled Shifted Windows Transformer with UNet to reduce the network complexity and incorporated deformable attention to construe the network explainablity. Additionally, Huang et al. [35] designed a Swin transformer-based GAN (STGAN) model with dual-discriminator structure for promoting the edge and texture preservation of reconstructed MR image. Fabian et al. [36] developed a hybrid architecture of CNN and Transformer to balance computational cost and reconstruction performance. In our previous work, a model combining CNN and Transformer was proposed for cine MRI reconstruction [37]. Guo et al. [38] proposed a lightweight recurrent Transformer to improve the parameter efficiency of the network. Although Transformer-based methods can extract more features and improve performance, some approaches remain complex and may affect practical application efficiency. These methods might be more sensitive to noise and may not effectively remove it. Additionally, Transformer-based methods typically require more diverse data for training; otherwise, they may overfit and affect the model's generalization ability.

These MRI reconstruction models are primarily categorized into four types: CNN-based models, RNN-based models and their variants, GAN-based models, and Transformer-based models and their variants. In terms of structure and complexity, CNN and conventional RNN models are relatively simple, making them easier to train and implement, whereas GAN and Transformer models have more complex structures and are harder to train. Regarding noise and artifact handling, CNN and RNN models may lack robustness in dealing with specific types of noise or artifacts, while GANs excel in detail recovery but are not robust against certain noises. Transformers can extract more features but are sensitive to noise. In terms of data requirements, CNN and RNN models demand relatively low data diversity. GAN and Transformer models usually require more diverse data for training, otherwise, they are prone to overfitting.

Although these single domain MRI reconstruction methods based on deep learning outperform the traditional CS-MRI methods as their data-driven feature extraction and nonlinearity properties, more different features from different domains are not fully leveraged to enhance the reconstruction performance. Therefore, a convolution-based network is employed to extract information from the frequency and wavelet domains. By fully aligning and fusing these features, the model uncovers more information and reconstructs high-quality magnetic resonance images. This approach also avoids using Transformer structures, balancing efficiency and performance.

2.2 Cross domain MRI reconstruction

As for cross domain method, the complementary information from different domains, i.e., image domain and k-space domain, benefits for reconstructing high-quality images. Initially, Eo et al. [13] proposed the KIKI-Net on k-space, image, k-space, and image sequentially for MRI reconstruction, which exhibits superior performance over single domain methods in terms of restoring tissue structures and removing aliasing artifacts. Souza et al. [14] developed the W-Net which is composed of a complex-valued residual UNet in k-space domain and a real-valued UNet in image domain. Souza et al. [15] introduced a cascaded convolutional neural network based on mixed learning of image domain and k-space domain, which consists of six CNN blocks with the same structure and alternate convolutional processing of image domain and frequency domain information. Ran et al. [17] designed a MRI dual-domain reconstruction network (MD-Recon-Net) which contains two parallel and interactive branches simultaneously operating on k-space and spatial-domain data. Liu et al. [19] proposed a full resolution deep interaction framework between image domain and k-space domain (DIIK-Net) with a few parameters for MRI reconstruction. Zhao et al. [39] introduced a Swin Transformer-based dual-domain GAN (SwinGAN), which consists of a frequency-domain generator and an image-domain generator for effectively capturing the long-distance dependencies.

Moreover, there are several methods that combine wavelet transform. Wang et al. [20] first used CNNs in the image domain, k-space and wavelet domain sequentially (IKWI-Net) for under-sampled MRI reconstruction, which can take advantage of the feature representation in different domains and achieve some improvements. Tong et al. [40] presented HIWDNet, which combines the image domain and wavelet domain for MRI reconstruction. Aghabiglou et al. [41] proposed a densely connected wavelet-based autoencoder to fully exploit wavelet domain information.

Cross-domain magnetic resonance reconstruction methods are mainly divided into two categories: frequency domain methods and wavelet domain methods. Frequency domain methods focus on alternating processing between the image domain and k-space, while wavelet domain methods combine information from the image domain and the wavelet domain. In terms of information utilization, frequency domain methods use complementary information from the image and frequency domains, whereas wavelet domain methods utilize multi-scale wavelet information. From the perspective of application scenarios, frequency domain methods are suitable for scenarios requiring the recovery of complex frequency information and long-range dependencies, while wavelet domain methods are suitable for capturing multi-scale features and details.

However, although these methods utilize multi-domain information, most of them simply fuse multi-domain data, which can easily lead to biases. Some methods, such as W-Net and SwinGAN, rely on optimization within specific domains, like k-space or image domain, and may perform well with certain types of data but may not meet expectations with other types. Furthermore, most methods focus solely on reconstruction without addressing the noise generated during the reconstruction process, particularly the biases that may arise from unaligned feature fusion. Therefore, in addition to designing a multi-domain alignment module, the HCFEM module is also designed to eliminate these noises and retain more details and edge information in the reconstructed images.

3 Method

3.1 Problem formulation

The problem of MRI reconstruction from under-sampled data can be described as:

$$y = M \bigodot Fx + \varepsilon \tag{1}$$

where x represents the fully-sampled MR image, F denotes the Fourier transform, M is the under-sampled mask, $\bigodot$ represents element-wise multiplication, and ε is the noise generated during the data acquisition process.

To solve this ill-posed inversion problem, deep learning is introduced to reconstruct the under-sampled MR image. The optimization formulation for this process is as follows:

$$\hat{x} = \arg\min_{x} \frac{1}{2} \left\| y - M \bigodot Fx \right\|_2^2 + \lambda L(\theta) \tag{2}$$

where $\hat{x}$ represents the reconstructed MR image, $L(\theta)$ is the prior regularization term, and λ is the regularization coefficient. In deep learning-based method, the regularization term $L(\theta)$ is defined as follows:

$$L(\theta) = \arg\min_{x} \frac{1}{2} \left\| x - f_{DL}(x_{zf}|\theta) \right\|_2^2 \tag{3}$$

where f_{DL} represents the forward propagation in deep learning, and θ represents the learnable parameters. x_{zf} denotes the zero-filled MR image after under-sampling.

3.2 Proposed methods

To reconstruct the full-sampled data from under-sampled MRI image, a multi-domain network framework with image domain, k-space and wavelet domain is proposed. As shown in Fig. 2., the proposed network consists of two parts, one is a parallel k-space and wavelet domain CNN and another is a U-shaped image domain network. Zero-filled images are used as the input of the proposed network. In the wavelet domain branch, the original input data is decomposed into four different sub-bands represented as W_u via the discrete wavelet transform (DWT), as shown in Fig. 3. Meanwhile, in the k-space domain branch, the original input data is transformed to frequency signal denoted as K_u by fast Fourier transform (FFT). The wavelet signal W_u and frequency signal K_u are separately reconstructed by a high-performance denoising network DIDN [42] to achieve the reconstructed wavelet signal W_r and the reconstructed k-space signal K_r. Then, W_r and K_r are inverse transformed to images I_w and I_k by inverse discrete wavelet transform (iDWT) and inverse fast Fourier transform (iFFT), respectively. To effectively eliminate artifacts and noise, both I_w and I_k are further reconstructed using UNet networks to generate the signals, I_{wr} and I_{kr} respectively. CNN in k-space domain is utilized for k-space completion, while CNN in wavelet domain is used to extract low-frequency and high-frequency features separately at the specific location [20]. To fuse the two branch results and alleviate the artifacts, a multi-domain alignment module is designed to align the different domain signal into a unify representation space I_a. After that, I_a is fed into the hierarchical cross feature enhancement module to obtain the final reconstruction.

3.3 Dual-domain feature alignment module

To fuse the dual-domain CNN and alleviate the artifacts or deformations from different domains, a dual-domain feature alignment module (DFAM) is designed as shown in Fig. 4. Feature alignment techniques [43] based on deformable convolution network (DCN) [44] have been applied to video super-resolution reconstruction, which can perform local deformations on the input during the convolution process. Compared to traditional convolution operations, deformable convolution can better adapt to geometric variations in the input, thereby improving the performance of the model. DFAM receives the reconstructed image I_{wr} from the wavelet domain and the reconstructed image I_{kr} from the k-space domain as the input. I_{wr} is processed by a convolution operation to generate the feature maps with three times the number of channels, which is split into three feature maps o_1, o_2, and m with the same number of channels as I_{wr}. Then, the feature maps o_1 and o_2 are concatenated to form a feature map with twice the number of channels as I_{wr}, which serves as the offset. The feature map generated by applying the sigmoid operation to m is served as the modulation mask. Finally, the input I_{kr}, offset, and modulation mask are processed by a deformable convolutional network to obtain the aligned feature map I_a. The process is formulated as:

$$o_1, o_2, m = \mathcal{C}hunk(\mathcal{C}onv(I_{wr})) \tag{4}$$

$$o = \mathcal{C}at(o_1, o_2) \tag{5}$$

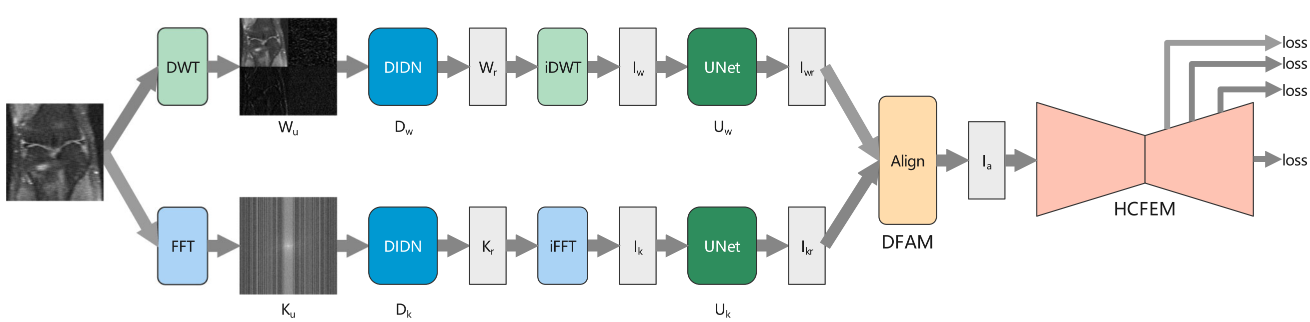


Fig. 2 The overall architecture of the proposed method. The pipeline mainly includes frequency domain branch and wavelet domain branch, dual-domain feature alignment module (DFAM) and Hierarchical cross feature enhancement module (HCFEM)

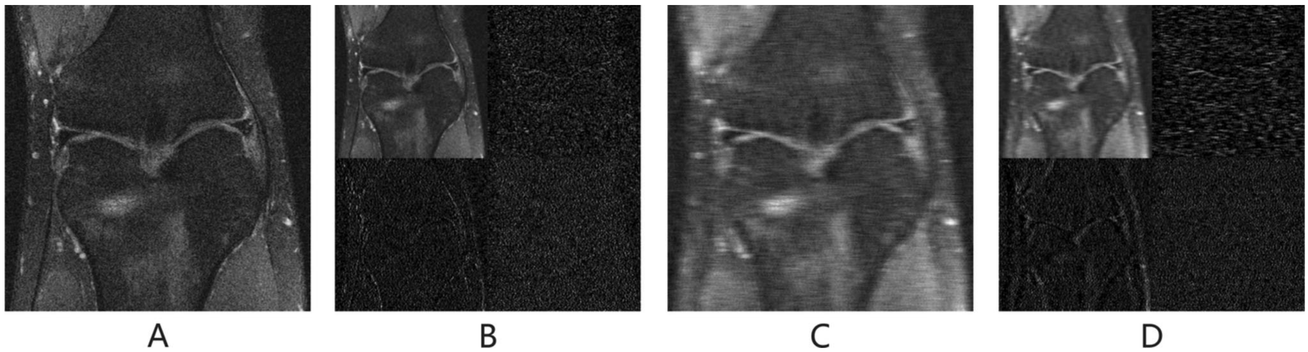


Fig. 3 A is the fully sampled MR image, B is the wavelet sub-bands of A, C is the under-sampled MR image, and D is the wavelet sub-bands of C

$$\hat{a} = Sigmoid(m) = \frac{1}{1+e^{-m}} \tag{6}$$

$$I_a = \mathcal{DCN}(I_{kr}, o, \hat{m}) \tag{7}$$

where $Conv$ represents the convolution operation, $Chunk$ represents the split operation, o_1, o_2, and m are the feature maps after splitting. Cat denotes the concatenation operation, $Sigmoid$ is the logistic function operation. $\mathcal{DCN}$ represents deformable convolutional network.

3.4 Hierarchical cross feature enhancement module

In image domain reconstruction, a hierarchical cross feature enhancement module as shown in Fig. 5. is designed to obtain the final reconstruction results. Nested UNet as an extended version of the traditional UNet, nests the encoder and decoder layers of different scale UNet to form multiple hierarchical sub-networks, which can enhance the network's ability to capture feature information in different scales. Further, to improve information propagation and optimize feature fusion, two cross-attention (CA) modules are introduced in the skip connections at different encoder-decoder hierarchical in Nested UNet.

The cross-attention module comprises two main components: Query and Key-Value mapping. The feature map x from the encoder generates the query feature map $\mathcal{Q}$ by a convolutional operation, while the feature map y from the decoder generates the key feature map $\mathcal{K}$ and the value feature map $\mathcal{V}$ by two convolutional operations, respectively. The similarity matrix $\mathcal{S}$ is calculated by the query feature map $\mathcal{Q}$ and the key feature map $\mathcal{K}$, and then normalized to obtain the attention weight matrix $\mathcal{A}$. Additionally, the matrix $\mathcal{S}$ can be multiplied by a learnable weight Temp to adjust the weights of each attention head. Next, the attention weight matrix $\mathcal{A}$ is multiplied by the value feature map $\mathcal{V}$ to obtain the cross-attention feature map $\mathcal{C}$. The expressions for this process are as follows:

$$\mathcal{A} = Softmax((\mathcal{Q} \bigotimes \mathcal{K}) \times Temp) \tag{8}$$

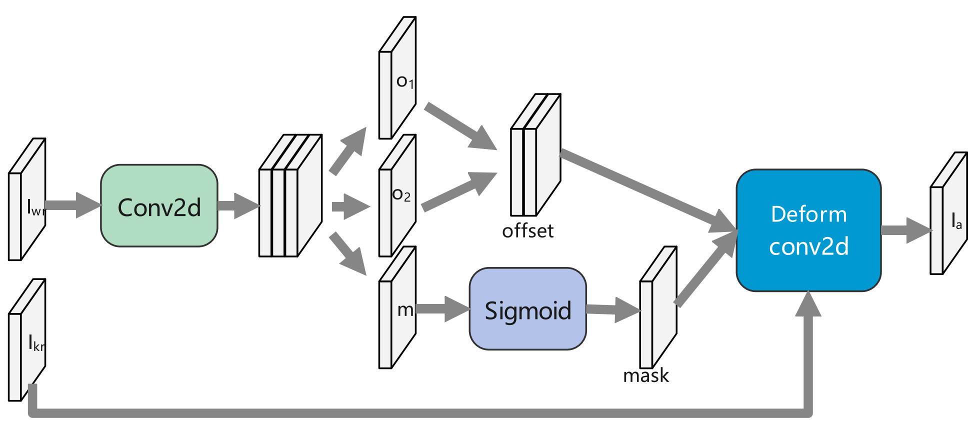
$$\mathcal{C} = \mathcal{A} \bigotimes \mathcal{V} \tag{9}$$

where $\bigotimes$ denotes matrix multiplication.

3.5 Deep supervision

In the image domain network, deep supervision is introduced to guarantee effective utilization of features at each hierarchy of HCFEM and accelerate the model convergence. Each decoder loss $\mathcal{L}_0$, $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$ of the HCFEM is added as the total loss to fuse the multi-scale feature information and

Fig. 4 Dual-domain feature alignment module. The input to this module includes the frequency branch feature map and the wavelet branch feature map. The core module is the deformable convolution, which differs from the regular convolution due to its offsets. The output is the feature map with aligned features



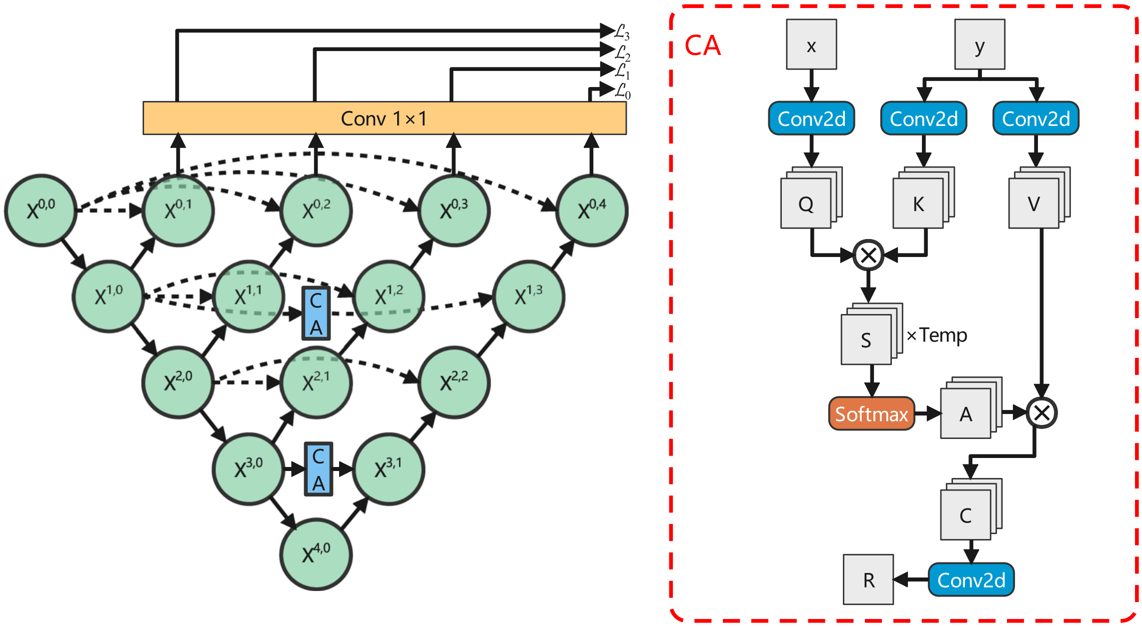


Fig. 5 Hierarchical cross feature enhancement module

understand the image semantics. The total loss is:

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \tag{10}$$

4 Experiment and results

4.1 Datasets

Two publicly available datasets, namely FastMRI [45] and Calgary-Campinas-359 (CC-359) [46], and our private clinical dataset are adopted to demonstrate the reconstruction performance of the proposed method. The FastMRI dataset comprises a large number of knee joint MRI images obtained from over 1,500 fully sampled knee joint MRIs. These images were acquired using 3.0 and 1.5 Tesla magnets, including coronal proton density-weighted images with and without fat suppression. The goal of the CC-359 dataset is to facilitate the development of innovative and fast deep learning models for reconstructing, processing, and analyzing brain magnetic resonance images within the scientific community. This dataset includes data from 359 subjects scanned with scanners from three different vendors (GE, Philips, and Siemens), and it contains T1 volumes acquired at two magnetic field strengths (1.5T and 3.0T). The scans correspond to scans of elderly subjects. Additionally, a private clinical knee MRI dataset is collected from a hospital in Shanghai, which was acquired using a 1.5T United Imaging uMR-588 scanner. This dataset comprises 126 patients with a total of 5,025 images. Due to the slower nature of T2 imaging compared to T1, T2 images are selected for testing, which are scaled to 256×256.

4.2 Implementation details

For FastMRI dataset, the MR data acquisition process is modeled by under-sampling the full sampled k-space data using 1D Gaussian mask. The under-sampling is carried out using 4-fold and 8-fold acceleration factors. And for CC-359 dataset, 2D Gaussian mask is used to under-sample the full sampled k-space data. The under-sampling is carried out using 5-fold and 10-fold acceleration factors. The training is initiated from created real-valued zero-filled images. The private clinical dataset is only adopted for testing, whose under-sampling manner is as CC-359 dataset. Six state-of-art methods, including U-Net [24], DC-CNN [22], KIKI-Net [13], XPD-Net [47], T2-Net [48], and SwinMR [32] are used to compared with the propose method. For a fair comparison, all networks are evaluated with the same setting for training on two public datasets.

The model is designed and implemented in Python 3.8 using PyTorch 1.12.1, a machine learning framework. Additionally, PyTorch-Lightning 1.7.5, an efficient PyTorch-based framework, is employed. The model is trained on a GPU server comprising an Intel Core i9-10900X CPU, 32GB of RAM in total, and Nvidia RTX3080 GPUs with a combined VRAM of 20GB. During the training process, the data batch size is determined based on the data size for optimal memory usage. The number of threads for data reading is set as 16. The initial learning rate is 1e-3, and it is dynamically adjusted with the ReduceLROnPlateau strategy. AdamW is used as the optimizer. The model undergoes 100 epochs of training conducted with double precision.

Additionally, the model is developed in an Ubuntu system environment using the PyCharm editor and an Anaconda virtual environment. The GPU version of PyTorch with the Cuda 11.7 library is used. TorchMetrics calculates the evaluation metrics, and TensorBoard records these metrics during training. OpenCV-Python along with NumPy are employed for image reading and processing. All model training employs five-fold cross-validation, and the error in our comparative experiment results is presented.

4.3 Evaluation metrics

To evaluate the quality of the proposed method for magnetic resonance image reconstruction, three commonly used image quality assessment metrics are employed, namely, Structural Similarity Index (SSIM) , Peak Signal-to-Noise Ratio (PSNR), and Normalized Mean Squared Error (NMSE). SSIM is defined as follows:

$$\mathcal{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{11}$$

where μ_x, μ_y, σ_x, σ_y, and σ_{xy} represent the local mean, standard deviation, and cross-covariance of images x and y, respectively. $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$, where L is the dynamic range of pixel values, and $k_1 = 0.01$ and $k_2 = 0.03$. PSNR is defined as:

$$\mathcal{PSNR} = 10 \ln \frac{\mathcal{R}^2}{\mathcal{MSE}} \tag{12}$$

$$\mathcal{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I(i, j) - R(i, j))^2 \tag{13}$$

where $I(i, j)$ denotes the pixel value of the original image at position (i, j), $R(i, j)$ represents the pixel value of the reconstructed image at position (i, j), and R represents the maximum fluctuation in the input image. NMSE is defined as:

$$\mathcal{NMSE} = \frac{\mathcal{MSE}}{\sum_{i=1}^{M} \sum_{j=1}^{N} I(i, j)^2} \tag{14}$$

4.4 Experimental results on the public dataset FastMRI

Table 1 presents the numerical assessment based on PSNR, SSIM, NMSE values of different methods on FastMRI dataset using 1D Gaussian under-sampling mask. Compared to the zero-filled images, all the seven methods achieve much better performance. The proposed method achieves superior performance over other methods at both 4-fold and 8-fold acceleration factors. For the 4× acceleration, the proposed method enhances the SSIM metric by 0.12% and PSNR metric by 0.64dB compared to the suboptimal model T2-Net model among the other models. For the 8× acceleration, the proposed method improves the SSIM metric by 0.42% and PSNR metric by 0.62dB compared to the suboptimal model T2-Net model. Figure 6 provides a visual representation of the quantitative results under 4× and 8× acceleration factors, respectively. The numerical results means that the proposed method can accelerate the MRI acquisition better than the other state-of-art methods.

The reconstruction MR images of different methods and the absolute differences between reconstruction and the ground truth under 4× and 8× acceleration rates are displayed in Figs. 7 and 8. The local enlarged regions of different methods also are given. From the visual effects, the results of the proposed method are closer to ground truth than the other compared methods. Especially under the low acceleration factor, the absolute difference between reconstruction and the ground truth is less. The proposed method can reconstruct the high-quality MR images with least artifacts and more details.

Table 1 Quantitative comparison results on the FastMRI dataset

Method	4×			8×		
	PSNR (dB) ↑	SSIM (%) ↑	NMSE ↓	PSNR (dB) ↑	SSIM (%) ↑	NMSE ↓
Zero-filled	26.84	63.48	0.0589	24.74	55.36	0.0683
U-Net	30.94 ± 1.23	73.22 ± 2.28	0.0307 ± 0.0142	29.12 ± 1.14	67.32 ± 3.41	0.0461 ± 0.0121
DC-CNN	30.49 ± 1.19	71.13 ± 2.45	0.0345 ± 0.0128	28.94 ± 1.35	66.46 ± 4.53	0.0478 ± 0.0198
KIKI-Net	30.64 ± 1.27	71.79 ± 3.38	0.0328 ± 0.0167	28.87 ± 1.22	66.32 ± 2.43	0.0487 ± 0.0173
XPD-Net	30.89 ± 1.06	72.52 ± 2.52	0.0311 ± 0.0126	29.02 ± 1.12	66.83 ± 2.22	0.0471 ± 0.0139
T2-Net	31.22 ± 1.13	73.94 ± 2.29	0.0286 ± 0.0134	29.84 ± 1.19	67.86 ± 3.44	0.0449 ± 0.0146
SwinMR	30.96 ± 1.25	73.28 ± 3.97	0.0304 ± 0.0150	29.06 ± 1.24	67.25 ± 4.49	0.0459 ± 0.0182
ours	**31.87 ± 1.16**	**74.06 ± 2.16**	**0.0251 ± 0.0133**	**30.46 ± 1.06**	**68.28 ± 2.19**	**0.0431 ± 0.0112**

The bold entries in the table indicate that the value represents the best performance

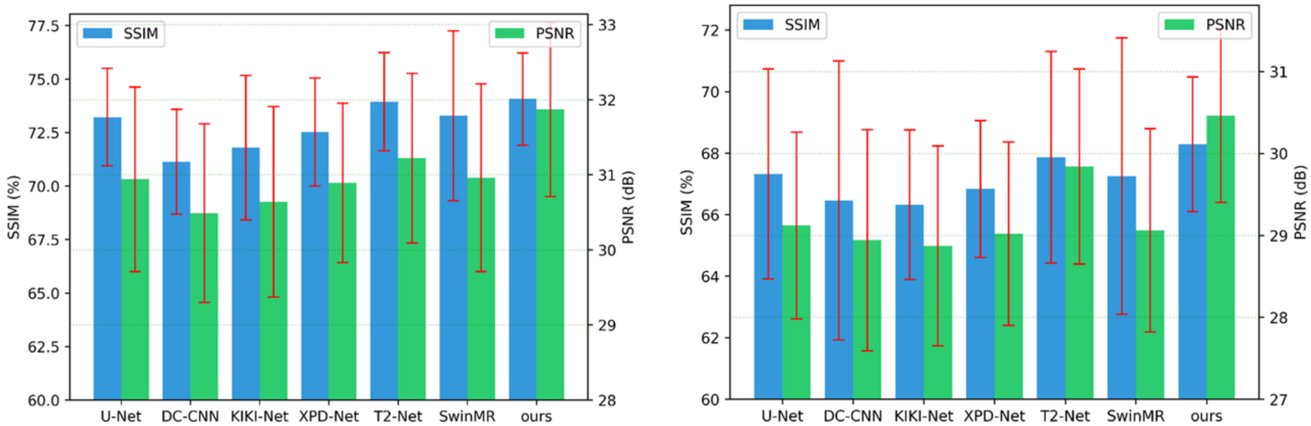


Fig. 6 Visualization of quantitative results of different methods on FastMRI dataset under 4× and 8× acceleration

4.5 Experimental results on the public dataset CC-359

Another compared experiments are conducted on the CC-359 dataset under 5× and 10× acceleration conditions using 2D Gaussian under-sampling mask. Table 2 lists the numerical assessments based on PSNR, SSIM, NMSE values of different methods. The proposed method outperforms the other methods at both 5× and 10× acceleration rates. Specifically, for 5× acceleration, the proposed method exhibits a 3.90% improvement in the SSIM metric and a 1.21dB improvement in the PSNR metric compared to the suboptimal model XPD-Net model. Similarly, for 10× acceleration, our proposed method demonstrates a 3.64% enhancement in the SSIM metric and a 0.85dB improvement in the PSNR metric compared to the suboptimal model XPD-Net model. The quantitative comparison results are illustrated in Fig. 9. All the experiment results highlight the superiority of our proposed method.

The reconstruction MR images of different methods and the absolute differences between reconstruction and the

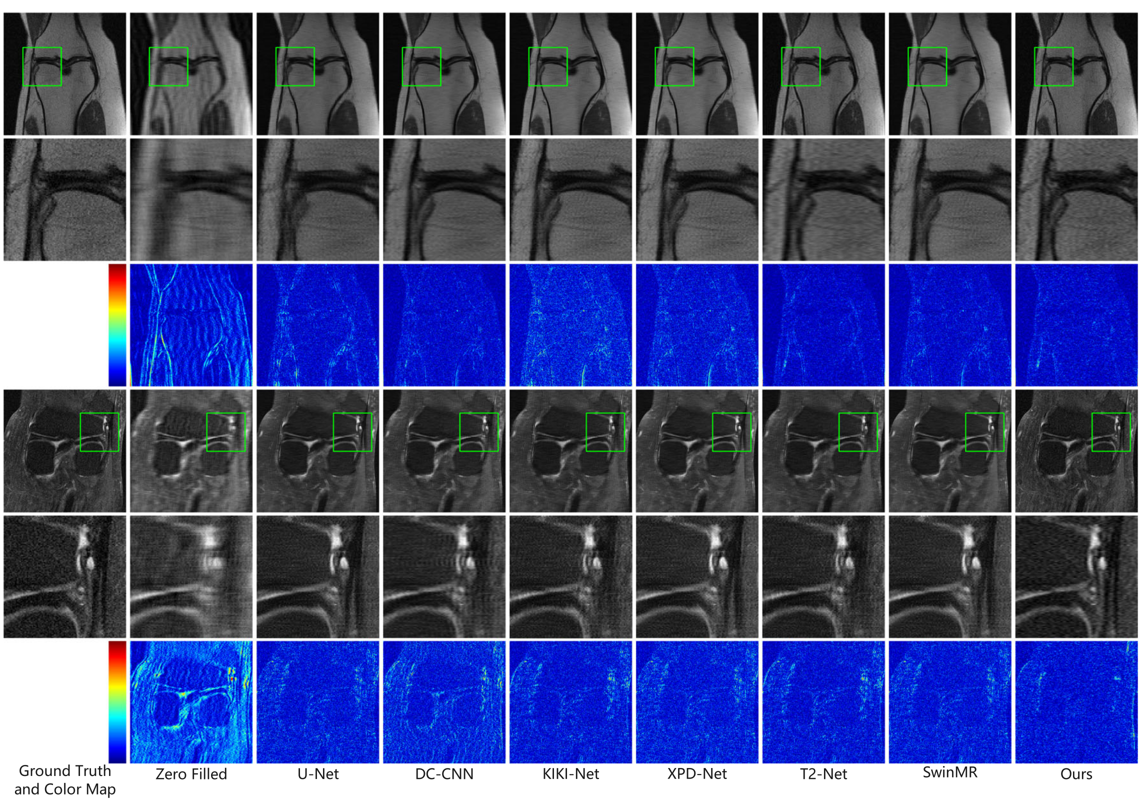


Fig. 7 Qualitative comparison results of different methods on the FastMRI dataset under 4× acceleration

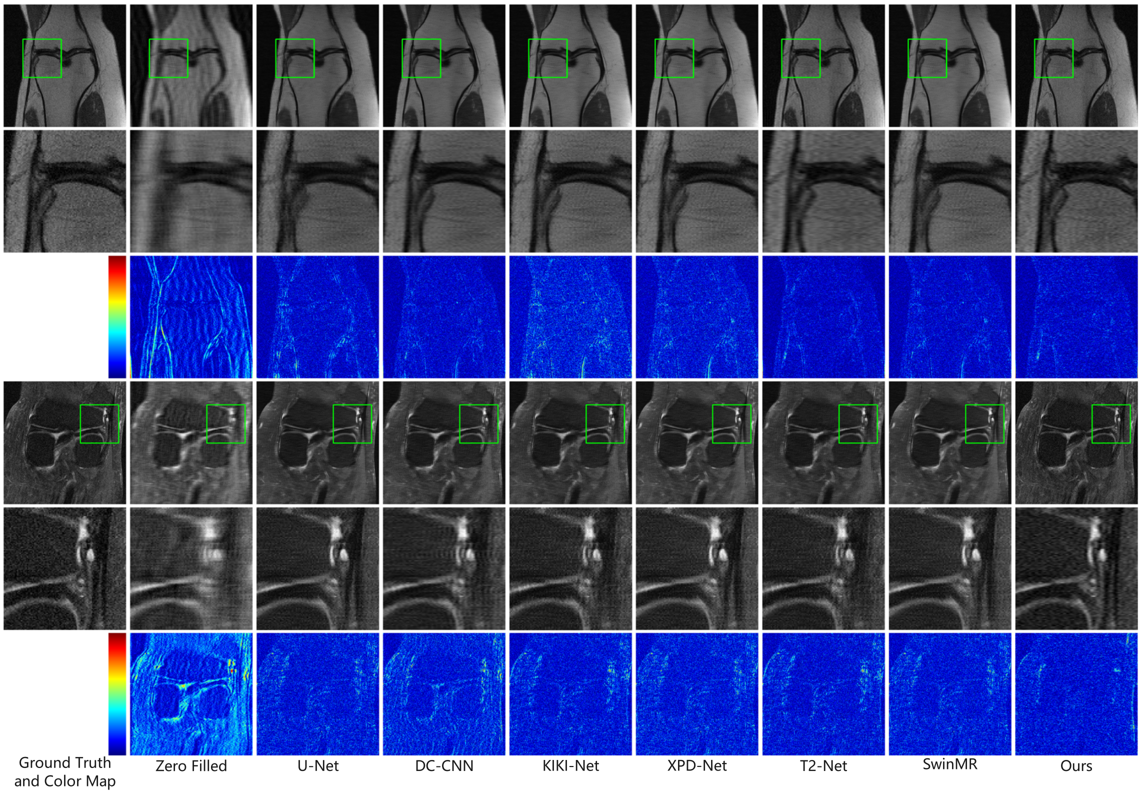


Fig. 8 Qualitative comparison results of different methods on the FastMRI dataset under 8× acceleration

ground truth under 5× and 10× acceleration rates on the CC-359 dataset are displayed in Figs. 10 and 11. The local enlarged regions of different methods also are given. From the visual effects, the proposed method can consistently achieves the superior reconstruction performance, even with the more aggressive under-sampling rates, when compared to other methods. The results in terms of numerical measurement and visual effect demonstrate that the proposed method can accelerate the magnetic resonance imaging with high imaging quality.

4.6 Experimental results on the private clinical dataset

To assess the generalization of the proposed method, all the comparison models trained on the CC-359 dataset are tested by using a separate private clinical dataset. Table 3 presents the numerical evaluation results based on PSNR, SSIM, NMSE under the 5× and 10× acceleration factors. For a 5× acceleration factor, the proposed method demonstrated a 7.8% improvement in the SSIM metric and a 2.59dB

Table 2 Quantitative comparison results on the CC-359 dataset

Method	5×			10×		
	PSNR (dB) ↑	SSIM (%) ↑	NMSE ↓	PSNR (dB) ↑	SSIM (%) ↑	NMSE ↓
Zero-filled	20.17	52.21	0.3568	18.32	45.61	0.5281
U-Net	30.14 ± 2.31	85.18 ± 1.80	0.0328 ± 0.0084	28.69 ± 2.42	81.42 ± 2.28	0.0415 ± 0.0086
DC-CNN	30.46 ± 2.46	86.25 ± 1.94	0.0192 ± 0.0098	29.34 ± 1.45	83.26 ± 1.27	0.0301 ± 0.0033
KIKI-Net	31.51 ± 1.88	86.59 ± 1.16	0.0164 ± 0.0046	29.49 ± 2.05	83.91 ± 1.86	0.0210 ± 0.0057
XPD-Net	31.48 ± 2.17	88.18 ± 1.72	0.0168 ± 0.0074	29.23 ± 2.03	85.91 ± 1.81	0.0230 ± 0.0051
T2-Net	30.42 ± 1.13	85.60 ± 0.85	0.0312 ± 0.0035	29.88 ± 1.98	81.48 ± 1.39	0.0400 ± 0.0046
SwinMR	29.16 ± 2.15	82.88 ± 1.54	0.0349 ± 0.0048	28.09 ± 2.33	79.57 ± 1.89	0.0428 ± 0.0060
ours	**32.69** ± **1.64**	**92.08** ± **1.02**	**0.0162** ± **0.0041**	**30.08** ± **1.43**	**89.55** ± **1.18**	**0.0188** ± **0.0029**

The bold entries in the table indicate that the value represents the best performance

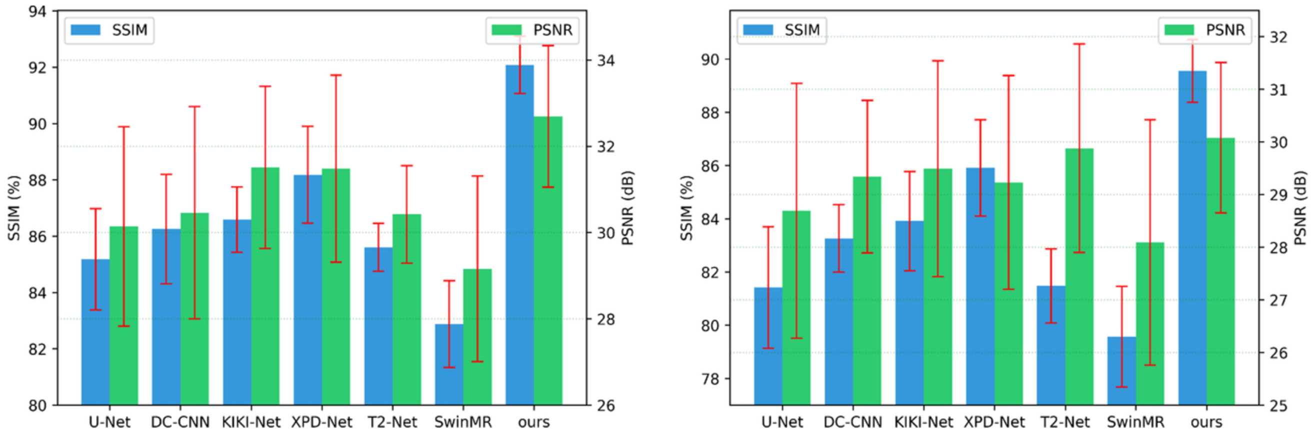


Fig. 9 Visualization of quantitative results of different methods on CC-359 dataset under 5× and 10× acceleration

improvement in the PSNR metric compared to the suboptimal model DC-CNN model. For a 10× acceleration factor, the proposed method exhibited a 6.78% improvement in the SSIM metric and a 4.4dB improvement in the PSNR metric compared to the suboptimal model U-Net model. The quantitative comparison results are depicted in Fig. 12. The reconstruction MR images and the error maps between reconstruction and the ground truth under 5× and 10× acceleration conditions on the private clinical dataset are displayed in Figs. 13 and 14. The numerical measurement and visual effect demonstrate that the proposed method has robust generalization capability and better reconstruction performance.

4.7 Ablation study

To demonstrate the feasibility and effectiveness of dual-domain feature alignment module (DFAM) and the hierarchical cross feature enhancement module (HCFEM) of the proposed network, the ablation studies are conducted using CC-359 dataset, as shown in Table 4. In the proposed model, DFAM aligns the features from k-space domain and the wavelet domain into a shared representation space, which can effectively remove the artifacts and enhance the models understanding capability. CAM in the HCFEM can reduce information transmission loss and strengthen feature

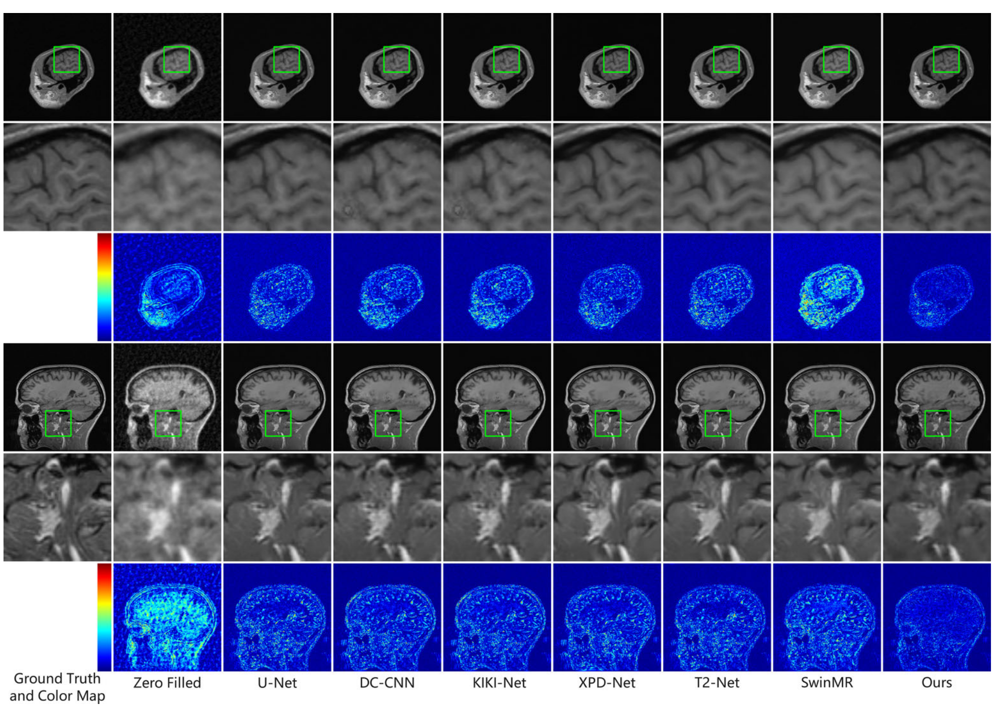


Fig. 10 Qualitative comparison results of different methods on the CC-359 dataset under 5× acceleration

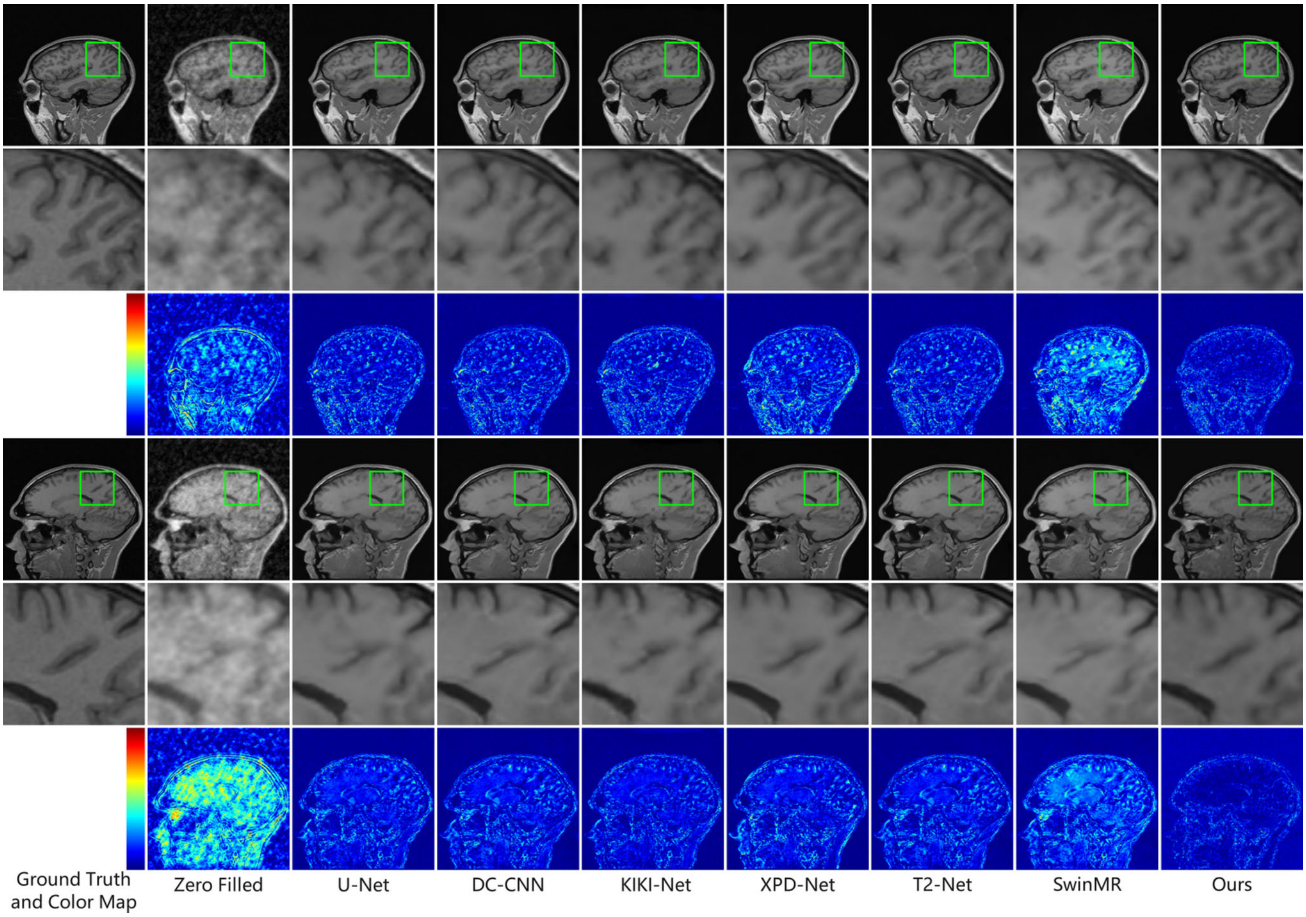


Fig. 11 Qualitative comparison results of different methods on the CC-359 dataset under 10× acceleration

representation. Two modules benefit for enhancing the reconstruction performance.

To take advantage of different domains, the parallel CNNs with k-space domain and wavelet domain is designed. The ablation study about using k-space domain and wavelet domain is conducted on CC-359 dataset, as shown in Table 5. From the numerical results, it can be seen that the dual-domain structure has some improvement in construction performance compared to single domain.

4.8 Experiment comparing HCFEM with other advanced methods

To better demonstrate the superiority of our proposed hierarchical cross feature enhancement module (HCFEM), the related experiments are conducted. Specifically, the HCFEM in the proposed model is replace with simple convolutional blocks and similar convolutional modules from MD-Recon-Net [17] and DIIK-Net [19]. Through this experiment, the

Table 3 Quantitative comparison results on the private clinical dataset

Method	5×			10×		
	PSNR (dB) ↑	SSIM (%) ↑	NMSE ↓	PSNR (dB) ↑	SSIM (%) ↑	NMSE ↓
Zero-filled	22.13	70.29	0.1490	20.69	63.74	0.2260
U-Net	24.37 ± 2.52	78.34 ± 3.16	0.0439 ± 0.0138	24.17 ± 2.94	76.06 ± 2.74	0.0544 ± 0.0107
DC-CNN	26.55 ± 2.11	79.11 ± 2.64	0.0299 ± 0.0109	24.86 ± 3.20	75.92 ± 3.17	0.0452 ± 0.0126
KIKI-Net	26.36 ± 1.86	78.95 ± 2.05	0.0301 ± 0.0043	25.70 ± 1.24	74.11 ± 1.54	0.0391 ± 0.0053
XPD-Net	20.37 ± 1.53	72.97 ± 1.12	0.1021 ± 0.0024	21.56 ± 1.61	66.71 ± 2.31	0.1883 ± 0.0064
T2-Net	20.09 ± 2.85	62.20 ± 3.49	0.1771 ± 0.0163	20.75 ± 3.84	47.63 ± 3.27	0.2627 ± 0.0128
SwinMR	20.21 ± 1.93	62.98 ± 2.41	0.1830 ± 0.0101	19.58 ± 2.39	43.57 ± 2.43	0.2206 ± 0.0082
ours	**29.14** ± **1.64**	**86.91** ± **1.16**	**0.0349** ± **0.0034**	**28.57** ± **2.01**	**82.84** ± **2.55**	**0.0389** ± **0.0088**

The bold entries in the table indicate that the value represents the best performance

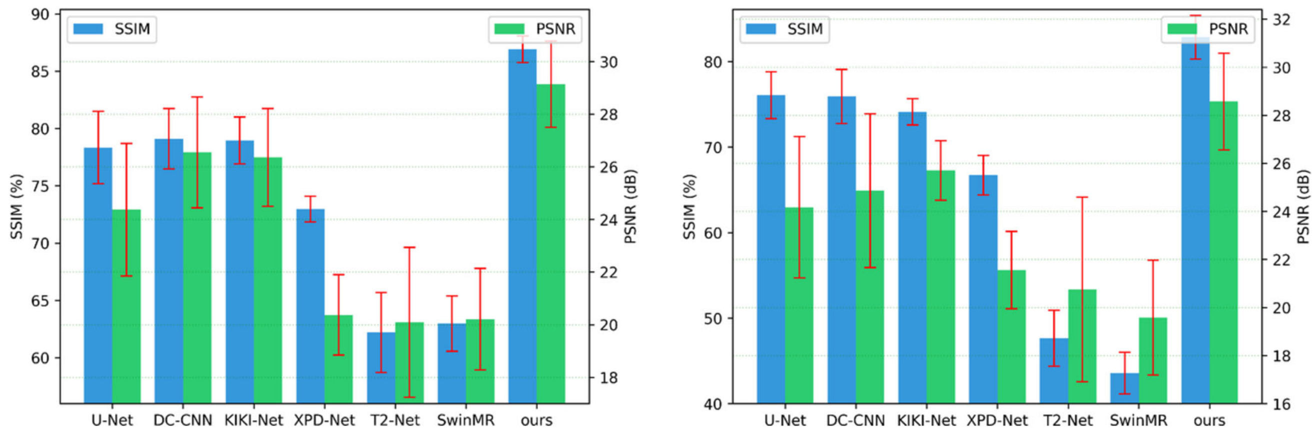


Fig. 12 Visualization of quantitative results of different methods on private clinical dataset under 5× and 10× acceleration

performance of HCFEM is further evaluated. The model in this experiment is trained on the CC359 dataset and tested on the private clinical dataset. The experimental results in Table 6 show that our HCFEM not only outperforms simple convolutional blocks but also performs slightly better than advanced convolutional modules proposed in some MRI reconstruction literature.

4.9 Experiments on loss function weights

Experiments are conducted to verify the effects of the impact of different weights for $\mathcal{L}_0$, $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ on model performance. However, since there are four losses, the final output $\mathcal{L}_0$ with the other three losses generated by deep supervision is juxtaposed. Therefore, the model loss is expressed

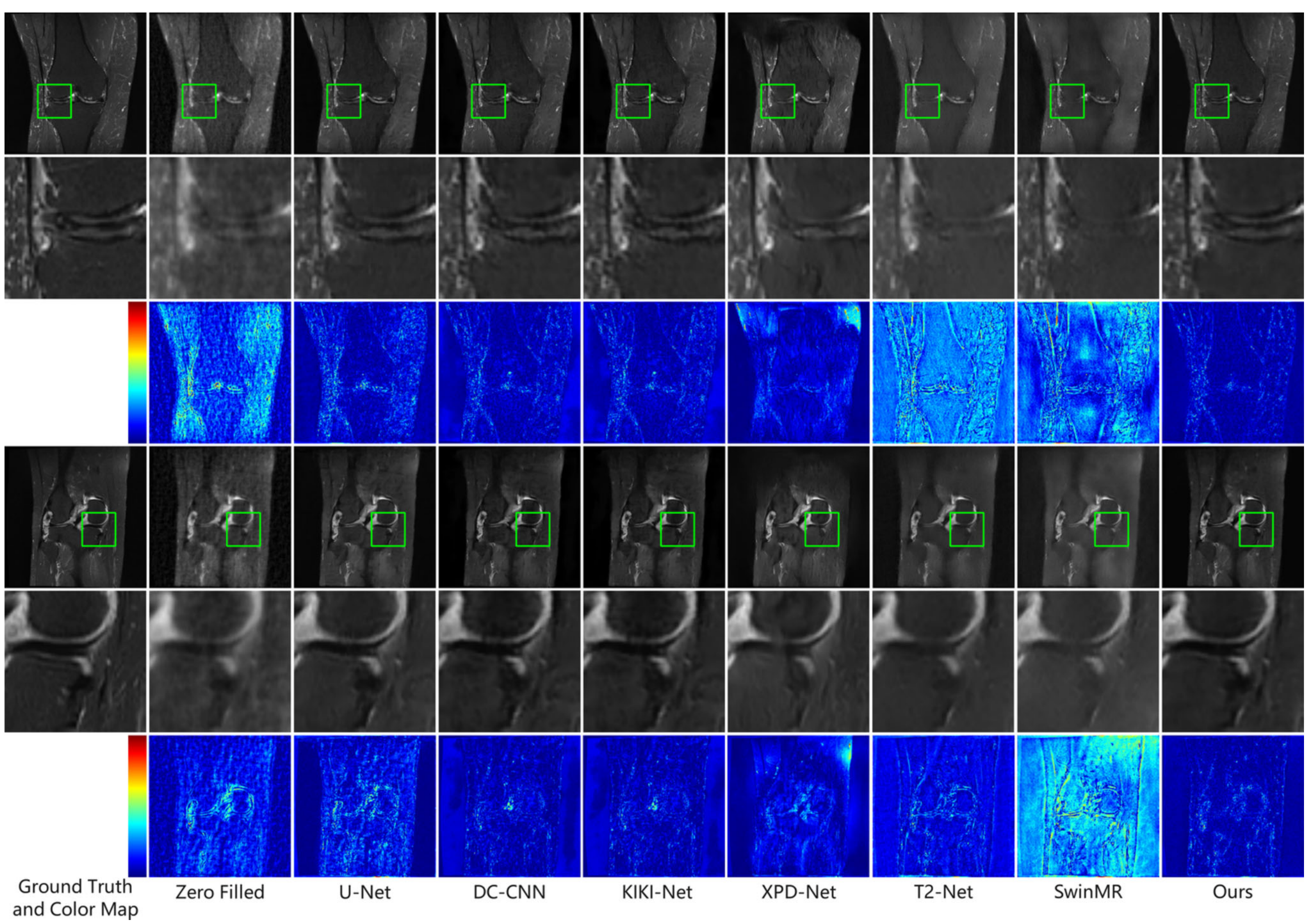


Fig. 13 Qualitative comparison results of different methods on the private clinical dataset under 5× acceleration

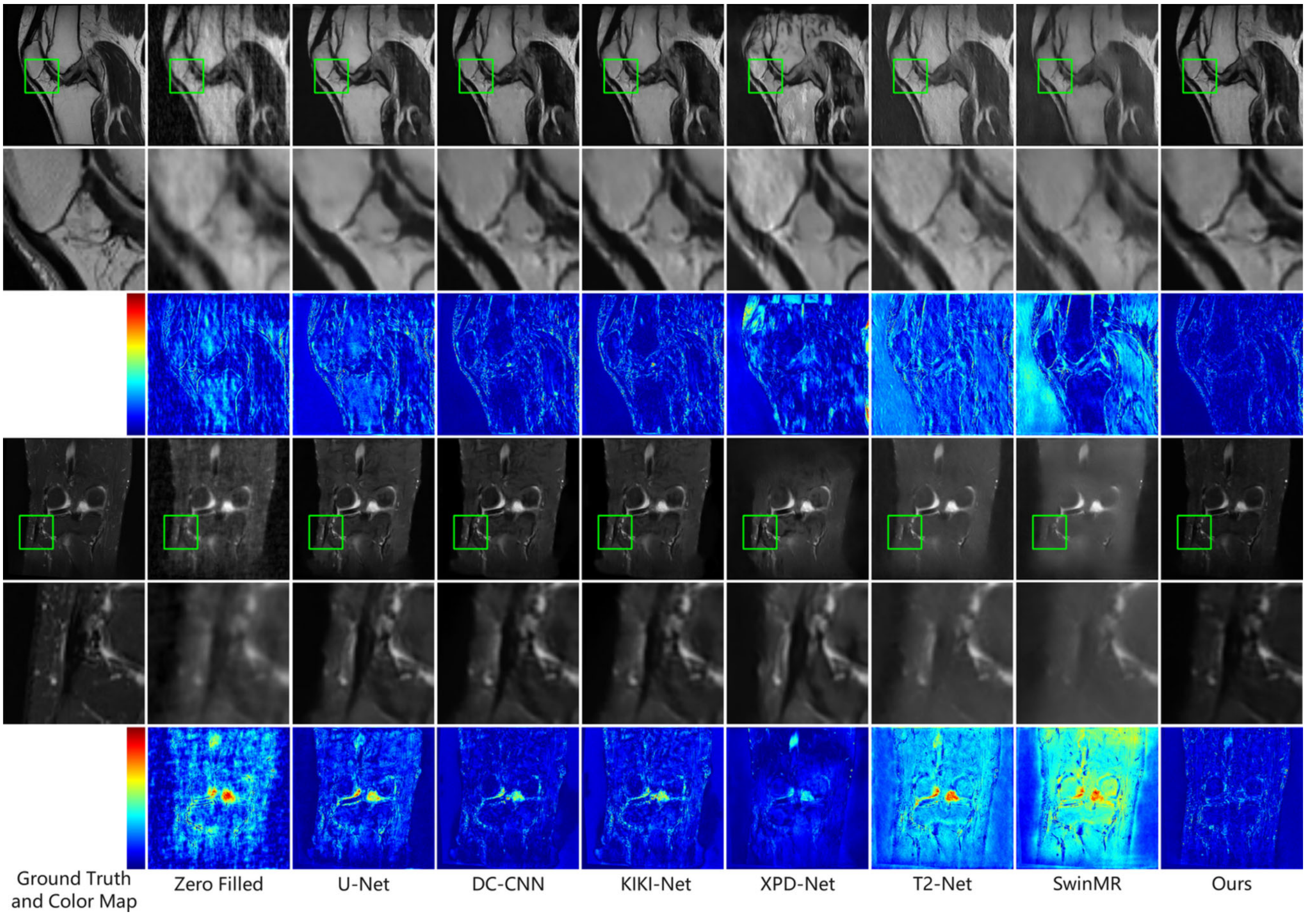


Fig. 14 Qualitative comparison results of different methods on the private clinical dataset under 10× acceleration

Table 4 The effect of DFAM and CAM on the performance of the proposed method on the CC-359 dataset

DFAM	CAM	5×			10×		
		PSNR	SSIM	NMSE	PSNR	SSIM	NMSE
✗	✗	31.97	91.26	0.0174	29.14	87.04	0.0245
✓	✗	32.48	91.58	0.0169	29.74	88.49	0.0228
✗	✓	32.35	91.51	0.0170	29.58	88.15	0.0232
✓	✓	**32.69**	**92.08**	**0.0162**	**30.08**	**89.55**	**0.0188**

The bold entries in the table indicate that the value represents the best performance

Table 5 The effect of using k-space or wavelet domain alone on the reconstruction performance of the proposed method on the CC-359 dataset

K	Wavelet	5×			10×		
		PSNR	SSIM	NMSE	PSNR	SSIM	NMSE
✓	✗	31.88	91.36	0.0176	29.43	88.89	0.0215
✗	✓	31.94	91.54	0.0171	29.51	89.02	0.0206
✓	✓	**32.69**	**92.08**	**0.0162**	**30.08**	**89.55**	**0.0188**

The bold entries in the table indicate that the value represents the best performance

Table 6 Experimental results comparing HCFEM with other advanced methods

Fusion modules	5×			10×		
	PSNR	SSIM	NMSE	PSNR	SSIM	NMSE
CNN block	26.73	79.86	0.0362	25.11	76.24	0.0427
MD-Recon-Net [17]	27.94	82.43	0.0417	27.04	80.59	0.0471
DIIK-Net [19]	28.14	83.27	0.0351	27.42	81.39	0.0395
ours	29.14	86.91	0.0349	28.57	82.84	0.0389

as $\mathcal{L} = \alpha\mathcal{L}_0 + \beta(\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3)$. Experiments focusing on the weights α and β are performed. The model in these experiments is trained on the CC359 dataset and tested on the private clinical dataset. As shown in the experimental results in Table 7, the impact of applying different weights to the loss function on model performance is minimal. However, when the weight of $\mathcal{L}_0$ decreases, there is a slight increase in performance degradation.

4.10 Uncertainty estimation

Despite the improvements in performance achieved by the proposed model, attention needs to be paid to the uncertainty of its predictions. Currently, several advanced methods for uncertainty estimation in deep learning have been proposed [49, 50]. Experiments are conducted on the CC359 dataset and the private clinical dataset using an evidence-based uncertainty estimation method tailored to the task's characteristics. The specific uncertainty estimation method is as follows:

$$e_i = \mathcal{F}_{evi}(\mathcal{F}_{rec}(x_i)) = e^{\frac{\tanh(\mathcal{F}_{rec}(x_i))}{\tau}} \tag{15}$$

$$u_i = \frac{K}{\sum_{k=1}^{K} e_i^k + 1} \tag{16}$$

where $\mathcal{F}_{evi}(\cdot)$ is the transformation function, $\mathcal{F}_{rec}(\cdot)$ is the proposed MRI reconstruction model, x_i is the i-th input sample, $0 < \tau < 1$ is the scaling factor, e_i is the evidence vector, K is the number of categories, and u_i is the uncertainty estimation result.

The uncertainty evaluation experiment is performed on the private clinical dataset using the model trained on the CC359 dataset. The estimation results are shown in Fig. 15. It can be seen from the figure that the reconstruction certainty is lower at the edges of the organs in the image. This is a problem that many related studies highlight as needing attention. Novel methods are planned to be designed in the future to mitigate this issue.

4.11 Testing on images of different sizes

Since our model is trained on a dataset of 256 × 256 images, which is the most common MRI resolution in clinical practice, the model's performance is also tested on test sets of images with different sizes. Specifically, we test the model trained on the CC359 dataset using our private clinical dataset. The experimental results are shown in Table 8. The test dataset is processed in two ways: one set of images is randomly resized to 0.5-1 times the original size, and the other set is randomly resized to 1-2 times the original size. The experimental results show that the proposed model exhibits a certain degree of adaptability to minor scale variations in the images, with minimal impact on performance. For test images smaller than the training images, which are directly upscaled to the training image size, which does not result in any loss of image information and therefore has almost no impact on reconstruction performance. The experimental results even show a slight improvement in performance. For test images slightly larger than the training images, no resizing is necessary, and they can be directly input into the model for reconstruction, with experimental results showing only a small performance loss. However, if the test image size is much larger than the training image size (a rare occurrence in clinical practice), it is likely to affect reconstruction performance. A possible solution for this is to divide the large image into several smaller images, reconstruct each small image, and then stitch them together to form the large image.

4.12 Model inference time

The inference time of a model, which is the time required for the model to process input data and generate output

Table 7 Experimental results of different loss function weights

Fusion modules		5×			10×		
α	β	PSNR	SSIM	NMSE	PSNR	SSIM	NMSE
0.5	1.0	29.01	86.75	0.0368	28.43	82.65	0.0418
1.0	0.5	29.09	86.84	0.0362	28.51	82.78	0.0401
1.0	1.0	29.14	86.91	0.0349	28.57	82.84	0.0389

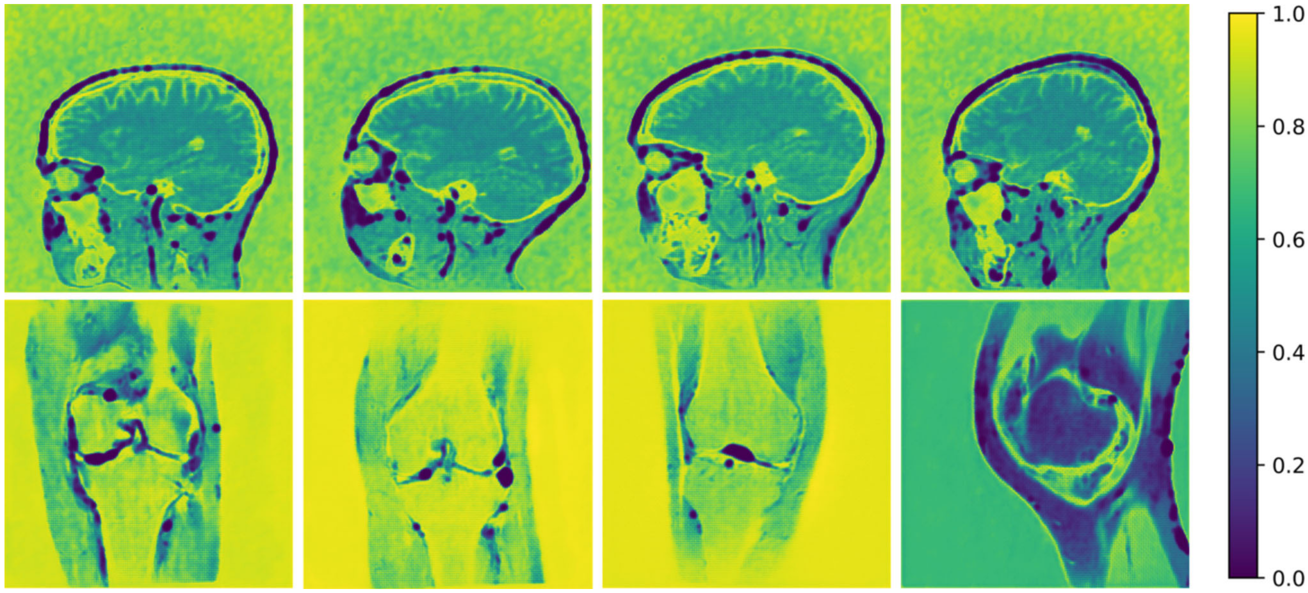


Fig. 15 Experiments on the proposed model using the CC359 dataset and our private clinical dataset employ the evidence-based uncertainty estimation method

results, directly impacts the performance and user experience of an application. For MRI, it is essential for the model to have a certain level of real-time capability. In other words, the model's inference time should be less than the signal acquisition time of the MRI equipment to ensure maximum efficiency. Therefore, the inference time of our model is examined under different hardware conditions and various modes. Specifically, experiments are conducted on three different GPUs, including the NVIDIA GeForce RTX 3080, RTX 2080 Ti, and GTX 1080 Ti. Additionally, the inference time is investigated under parallel conditions and with model quantization. The specific experimental results shown in Table 9 indicate that the quantized model running on the RTX 3080 performs the best. Given the slight performance loss incurred by quantization, parallel inference on the RTX 3080 GPU is deemed to offer an optimal cost-performance ratio.

5 Discussion

The qualitative and quantitative results from the extensive experiments demonstrate the superiority and generalization capability of the proposed method. The proposed method can achieve the good performance with three reasons. First, the proposed method take full advantages of different features from the image domain, k-space domain, and wavelet domain, which benefits the model to obtain better reconstruction MR images with more details and structures. The multi-domain architecture design also can enhance the model's understanding of the underlying image representation, thereby improving its generalization. Second, the deformable alignment module reduces the disparity in feature representations across different domains by aligning the k-space domain and wavelet domain features into a unified representation space. This alignment enables the model to better understand and utilize information from different domains. Third, the introduction of the HCFEM and the design of cross-attention modules between different hierarchies of encoders and decoders facilitate the transfer and integration of information across various levels. This module reduces information loss during transmission, enhances feature representation, and improves the model's robustness.

Specifically, the experimental results show that Transfor mer-based such as T2-Net and SwinMR perform well on the large-scale FastMRI dataset but poorly on the smaller CC-359 dataset. It means that Transformer-based methods

Table 8 The impact of different test data sizes on model performance

Image Size	5×			10×		
	PSNR	SSIM	NMSE	PSNR	SSIM	NMSE
256 × 256	29.14	86.91	0.0349	28.57	82.84	0.0389
Random 1 2×	28.46	86.05	0.0487	27.32	81.92	0.0497
Random 0.5 1×	29.21	86.98	0.0337	28.68	82.92	0.0368

Table 9 The inference time of the proposed model under different hardware conditions and various modes

GPU device	Inference time (ms per image)	Parallel inference time (ms per image)	Quantized inference time (ms per image)
RTX 3080	42.8 ± 0.6	16.6 ± 0.3	8.1 ± 0.8
RTX 2080Ti	52.4 ± 0.5	24.2 ± 5.3	19.6 ± 2.1
GTX 1080Ti	78.7 ± 1.3	36.8 ± 3.7	28.4 ± 4.3

commonly require a larger scale of training data. However, when tested on our private clinical test set with a domain gap from the training data, Transformer-based methods exhibit significantly poor performance, indicating their limited generalization capabilities. Additionally, the experimental results also indicates that CNN-based methods outperform Transformer-based methods in terms of generalization.

Although the proposed method shows its advantages, it still has limitations. The first is the k-space domain CNN branch and the wavelet domain CNN branch is relatively parallel. The interactive relationship between each domain are not well studied. Second, 1D Gaussian and 2D Gaussian sampling masks are used to generate the under-sampled images in this paper. Some other masks, like Cartesian, radial, Possion and learnable masks should be utilized to generate the under-sampled images and improve the performance.

6 Conclusion

In this paper, a novel network for MRI reconstruction based on cross-domain method which combined k-space domain, wavelet domain and image domain to reconstruct MRI from under-sampled k-space data is proposed. The proposed network is composed of a parallel CNN with k-space domain branch and wavelet domain branch, and a U-shaped image domain network, which can exploit the feature representation from different domain to achieve the high-quality MR images. A dual-domain feature alignment module is introduced to align the feature representation from parallel CNNs with k-space and wavelet domain into a unified space and remove the artifacts. Additionally, a hierarchical cross feature enhancement module in image domain is designed to achieve the final reconstructed images. Besides, deep supervision is employed in HCFEM to enhance network performance and robustness. Experimental results demonstrate that the proposed method achieves good reconstruction and generalization performance under various acceleration conditions for brain and knee MRI. In future work, further improvements will be made to enhance the proposed method's performance and computational efficiency. However, this study trains and tests model only on brain and knee data. In the future, it is necessary to collect multi-center clinical data from different anatomical sites, devices, and acquisition methods. This study does not consider motion artifacts or other artifacts generated during the acquisition process. Future research can design models to address these challenges. Additionally, more efficient inference models can be developed for scenarios with limited hardware conditions.

Acknowledgements The authors would like to thank the anonymous reviewers and the associate editor for their constructive comments and suggestions that helped to improve both the technical content and the presentation quality of this paper. This work is supported by the National Natural Science Foundation of China under grant No.61801288.

Author Contributions Qiaohong Liu is responsible for methodological research, paper writing, and funding acquisition. Xiaoxiang Han is responsible for methodological research, conducting experiments, and data analysis. Yang Chen oversees conceptual guidance, project organization, and paper review within the team.

Data availability and access The data and code of the study are available from the corresponding author upon reasonable request.

Declarations

Ethical and informed consent for data used This study was approved by the Institutional Review Board of Shanghai University of Medicine and Health Sciences with a patient exemption applied. All publicly available data has been subject to an exemption.

Competing Interests The authors declare no potential conflicts of interest regarding any financial support, research, authorship, and publication of this article.

References

1. Chatterjee S, Breitkopf M, Sarasaen C, Yassin H, Rose G, Nürnberger A, Speck O (2022) Reconresnet: regularised residual learning for mr image reconstruction of undersampled cartesian and radial data. Comput Biol Med 143:105321. https://doi.org/10.1016/j.compbiomed.2022.105321
2. Patel MR, Klufas RA (1999) Gradient-and spin-echo mr imaging of the brain. Am J Neuroradiol 20(7):1381–1383
3. Deshmane A, Gulani V, Griswold MA, Seiberlich N (2012) Parallel mr imaging. J Magn Reson Imaging 36(1):55–72. https://doi.org/10.1002/jmri.23639
4. Lustig M, Donoho D, Pauly JM (2007) Sparse mri: the application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for

Magnetic Resonance in Medicine 58(6):1182–1195. https://doi.org/10.1002/mrm.21391
5. Hu Y, Li P, Chen H, Zou L, Wang H (2021) High-quality mr fingerprinting reconstruction using structured low-rank matrix completion and subspace projection. IEEE Trans Med Imaging 41(5):1150–1164. https://doi.org/10.1109/TMI.2021.3133329
6. Eo T, Shin H, Jun Y, Kim T, Hwang D (2020) Accelerating cartesian mri by domain-transform manifold learning in phase-encoding direction. Med Image Anal 63:101689. https://doi.org/10.1016/j.media.2020.101689
7. Qu X, Guo D, Ning B, Hou Y, Lin Y, Cai S, Chen Z (2012) Undersampled mri reconstruction with patch-based directional wavelets. Magn Reson Imaging 30(7):964–977. https://doi.org/10.1016/j.mri.2012.02.019
8. Han X, Liu Y, Liu G, Lin Y, Liu Q (2023) Loanet: a lightweight network using object attention for extracting buildings and roads from uav aerial remote sensing images. PeerJ Comp Sci 9:1467. https://doi.org/10.7717/peerj-cs.1467
9. Liu Y, Han X, Liang T, Dong B, Yuan J, Hu M, Liu Q, Chen J, Li Q, Zhang Y (2023) Edmae: an efficient decoupled masked autoencoder for standard view identification in pediatric echocardiography. Biomed Signal Process Control 86:105280. https://doi.org/10.1016/j.bspc.2023.105280
10. Liu Y, Huang Q, Han X, Liang T, Zhang Z, Lu X, Dong B, Yuan J, Wang Y, Hu M et al (2024) Atrial septal defect detection in children based on ultrasound video using multiple instances learning. J Imaging Inform Med, 1–11. https://doi.org/10.1007/s10278-024-00987-1
11. Theerthagiri P, Ruby AU (2023) Seasonal learning based arima algorithm for prediction of brent oil price trends. Multimed Tools Appl 82(16):24485–24504. https://doi.org/10.1007/s11042-023-14819-x
12. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, Feng D, Liang D (2016) Accelerating magnetic resonance imaging via deep learning. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE, pp 514–517. https://doi.org/10.1109/ISBI.2016.7493320
13. Eo T, Jun Y, Kim T, Jang J, Lee H-J, Hwang D (2018) Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn Reson Med 80(5):2188–2201. https://doi.org/10.1002/mrm.27201
14. Souza R, Frayne R (2019) A hybrid frequency-domain/image-domain deep network for magnetic resonance image reconstruction. In: 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, pp 257–264. https://doi.org/10.1109/SIBGRAPI.2019.00042
15. Souza R, Lebel RM, Frayne R (2019) A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction. In: International conference on medical imaging with deep learning. PMLR, pp 437–446
16. Sun L, Wu Y, Shu B, Ding X, Cai C, Huang Y, Paisley J (2020) A dual-domain deep lattice network for rapid mri reconstruction. Neurocomputing 397:94–107. https://doi.org/10.1016/j.neucom.2020.01.063
17. Ran M, Xia W, Huang Y, Lu Z, Bao P, Liu Y, Sun H, Zhou J, Zhang Y (2020) Md-recon-net: a parallel dual-domain convolutional neural network for compressed sensing mri. IEEE Trans Radiat Plasma Med Sci 5(1):120–135. https://doi.org/10.1109/TRPMS.2020.2991877
18. Wang S, Ke Z, Cheng H, Jia S, Ying L, Zheng H, Liang D (2022) Dimension: dynamic mr imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. NMR Biomed 35(4):4131. https://doi.org/10.1002/nbm.4131
19. Liu Y, Pang Y, Liu X, Liu Y, Nie J (2023) Diik-net: a full-resolution cross-domain deep interaction convolutional neural network for mr image reconstruction. Neurocomputing 517:213–222. https://doi.org/10.1016/j.neucom.2022.09.048
20. Wang Z, Jiang H, Du H, Xu J, Qiu B (2020) Ikwi-net: a cross-domain convolutional neural network for undersampled magnetic resonance image reconstruction. Magn Reson Imaging 73:1–10. https://doi.org/10.1016/j.mri.2020.06.015
21. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2019) Unet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 39(6):1856–1867. https://doi.org/10.1109/TMI.2019.2959609
22. Schlemper J, Caballero J, Hajnal JV, Price A, Rueckert D (2017) A deep cascade of convolutional neural networks for mr image reconstruction. In: Information processing in medical imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25. Springer, pp 647–658. https://doi.org/10.1007/978-3-319-59050-9_51
23. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D (2018) Convolutional recurrent neural networks for dynamic mr image reconstruction. IEEE Trans Med Imaging 38(1):280–290. https://doi.org/10.1109/TMI.2018.2863670
24. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
25. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Inf Process Syst 27:2672–2680. https://doi.org/10.48550/arXiv.1406.2661
26. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778. https://doi.org/10.1109/CVPR.2016.90
27. Aghabiglou A, Eksioglu EM (2021) Mr image reconstruction using densely connected residual convolutional networks. Comput Biol Med 139:105010. https://doi.org/10.1016/j.compbiomed.2021.105010
28. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, Liu F, Arridge S, Keegan J, Guo Y et al (2017) Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. IEEE Trans Med Imaging 37(6):1310–1321. https://doi.org/10.1109/TMI.2017.2785879
29. Quan TM, Nguyen-Duc T, Jeong W-K (2018) Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. IEEE Trans Med Imaging 37(6):1488–1497. https://doi.org/10.1109/TMI.2018.2820120
30. Zhang T, Wan P, Zhang D (2019) Segan: structure-enhanced generative adversarial network for compressed sensing mri reconstruction. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 1012–1019. https://doi.org/10.1609/aaai.v33i01.33011012
31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30. https://doi.org/10.48550/arXiv.1706.03762
32. Huang J, Fang Y, Wu Y, Wu H, Gao Z, Li Y, Del Ser J, Xia J, Yang G (2022) Swin transformer for fast mri. Neurocomputing 493:281–304. https://doi.org/10.1016/j.neucom.2022.04.051
33. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986
34. Huang J, Xing X, Gao Z, Yang G (2022) Swin deformable attention u-net transformer (sdaut) for explainable fast mri. In: International

conference on medical image computing and computer-assisted intervention. Springer, pp 538–548. https://doi.org/10.1007/978-3-031-16446-0_51

35. Huang J, Wu Y, Wu H, Yang G (2022) Fast mri reconstruction: how powerful transformers are? In: 2022 44th Annual international conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp 2066–2070. https://doi.org/10.1109/EMBC48229.2022.9871475
36. Fabian Z, Tinaz B, Soltanolkotabi M (2022) Humus-net: hybrid unrolled multi-scale network architecture for accelerated mri reconstruction. Adv Neural Inf Process Syst 35:25306–25319. https://doi.org/10.48550/arXiv.2203.08213
37. Han X, Chen Y, Liu Q, Liu Y, Chen K, Lin Y, Zhang W (2024) Reconstruction of cardiac cine mri using motion-guided deformable alignment and multi-resolution fusion. Int J Imaging Syst Technol 34(4):23131. https://doi.org/10.1002/ima.23131
38. Guo P, Mei Y, Zhou J, Jiang S, Patel VM (2023) Reconformer: accelerated mri reconstruction using recurrent transformer. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2023.3314747
39. Zhao X, Yang T, Li B, Zhang X (2023) Swingan: a dual-domain swin transformer-based generative adversarial network for mri reconstruction. Comput Biol Med 153:106513. https://doi.org/10.1016/j.compbiomed.2022.106513
40. Tong C, Pang Y, Wang Y (2022) Hiwdnet: a hybrid image-wavelet domain network for fast magnetic resonance image reconstruction. Comput Biol Med 151:105947. https://doi.org/10.1016/j.compbiomed.2022.105947
41. Aghabiglou A, Eksioglu EM (2022) Densely connected wavelet-based autoencoder for mr image reconstruction. In: 2022 45th International conference on Telecommunications and Signal Processing (TSP). IEEE, pp 212–215. https://doi.org/10.1109/TSP55681.2022.9851354
42. Yu S, Park B, Jeong J (2019) Deep iterative down-up cnn for image denoising. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 0–0. https://doi.org/10.1109/CVPRW.2019.00262
43. Tian Y, Zhang Y, Fu Y, Xu C (2020) Tdan: temporally-deformable alignment network for video super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3360–3369. https://doi.org/10.1109/CVPR42600.2020.00342
44. Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y (2017) Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision, pp 764–773. https://doi.org/10.1109/ICCV.2017.89
45. Zbontar J, Knoll F, Sriram A, Murrell T, Huang Z, Muckley MJ, Defazio A, Stern R, Johnson P, Bruno M et al (2018) fastmri: an open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839. https://doi.org/10.48550/arXiv.1811.08839
46. Souza R, Lucena O, Garrafa J, Gobbi D, Saluzzi M, Appenzeller S, Rittner L, Frayne R, Lotufo R (2018) An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. Neuroimage 170:482–494. https://doi.org/10.1016/j.neuroimage.2017.08.021
47. Ramzi Z, Ciuciu P, Starck J-L (2020) Xpdnet for mri reconstruction: an application to the 2020 fastmri challenge. arXiv preprint arXiv:2010.07290. https://doi.org/10.48550/arXiv.2010.07290
48. Feng C-M, Yan Y, Fu H, Chen L, Xu Y (2021) Task transformer network for joint mri reconstruction and super-resolution. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. Springer, pp 307–317. https://doi.org/10.1007/978-3-030-87231-1_30
49. Abbaszadeh Shahri A, Chunling S, Larsson S (2023) A hybrid ensemble-based automated deep learning approach to generate 3d geo-models and uncertainty analysis. Eng Comput, 1–16. https://doi.org/10.1007/s00366-023-01852-5
50. Abbaszadeh Shahri A, Shan C, Larsson S (2022) A novel approach to uncertainty quantification in groundwater table modeling by automated predictive deep learning. Nat Resour Res 31(3):1351–1373. https://doi.org/10.1007/s11053-022-10051-w

Qiaohong Liu was born in China. She received the Ph.D. degree from the School of Mechatronic Engineering and Automation, Shanghai University, in 2015. From 2016 to 2017, she was a Lecturer with the School of Health Information Technology and Management, Shanghai University of Medicine and Health Sciences. Since 2017, she has been an Associate Professor with the Medical Instrumentation College, Shanghai University of Medicine and Health Sciences. She is currently an Associate Professor and a Master's Tutor. She is mainly engaged in scientific research in medical image processing, medical imaging big data technology, computer simulation, and other related fields. In recent years, she has presided over some research projects, including one project of the Youth Science Fund of the National Natural Science Foundation of China in 2018. More than 50 academic articles have been published in various journals, including 15 in SCI. The Deputy Editor-in-Chief compiled four Chinese national textbooks.

Xiaoxiang Han was born in Yancheng, Jiangsu, P.R.China in 1998. He received the B.S. degree in computer science and technology from the Jinling Institute of Technology in 2021, and the master's degree in electronic information from the University of Shanghai for Science and Technology in 2024. From 2021 to 2024, he was a graduate student specializing in medical imaging and image processing with the School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, P.R.China. He has published multiple academic papers indexed by SCI or EI. His research interests encompass pattern recognition, medical image analysis, intelligent ultrasound diagnosis, semi-/weakly-/self-supervised learning, multimodal learning, etc.

Yang Chen a master's degree holder in Communication and Information Systems from the University of Electronic Science and Technology, is currently an MBA student at Peking University's Guanghua School of Management. He has been deeply engaged in the research and development of wireless communication technology for many years, with over a decade of experience in R&D and management in the fields of DSP, algorithms, system design, integration, and testing for both the network and terminal sides of wireless communication systems.