IDDNet: a deep interactive dual-domain convolutional neural network with auxiliary modality for fast MRI reconstruction
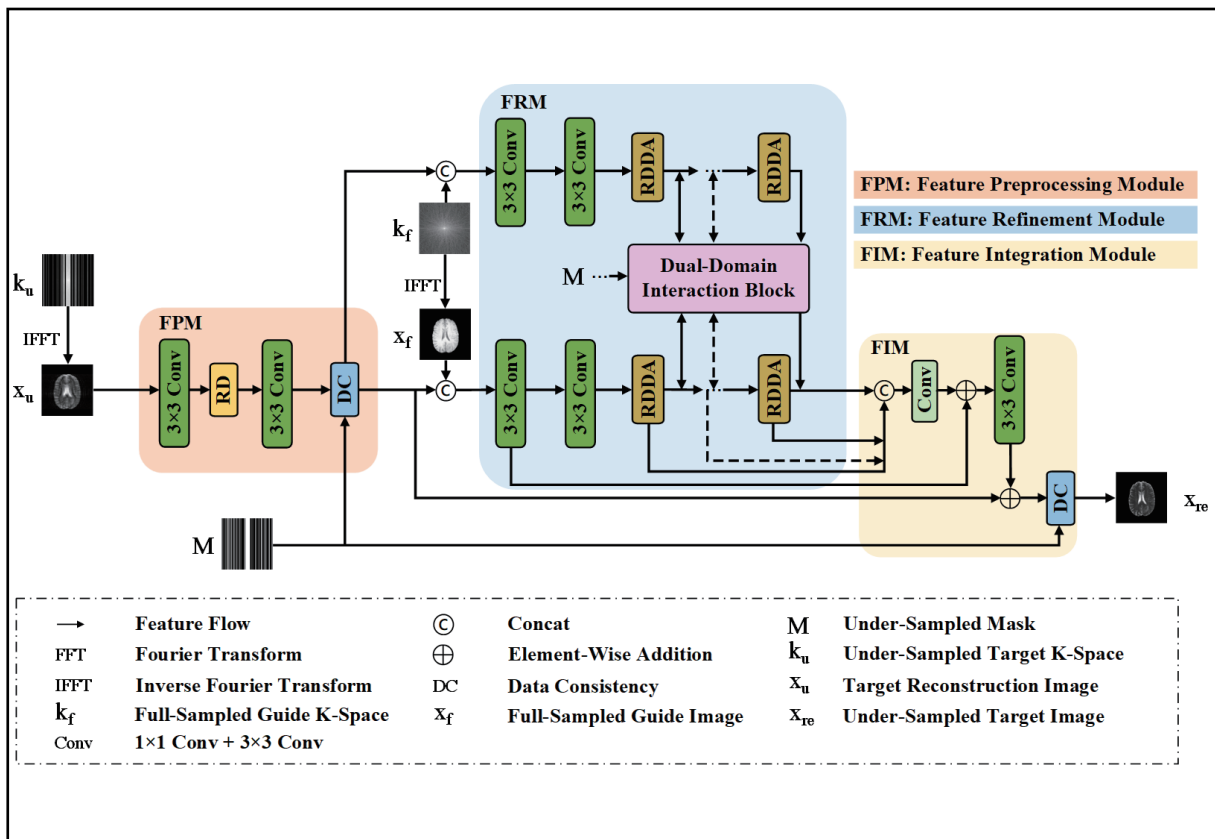
Yi Cao, and Hongwei Du ✉

School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Hongwei Du, E-mail: duhw@ustc.edu.cn

Graphical abstract

We propose IDDNet, a parallel dual-domain interaction framework with an auxiliary modality, to accelerate MRI reconstruction.

Public summary

■ We extract and interactively utilize relevant information from different MR domains by designing a hierarchical feature extraction and dual-domain parallel interaction network structure.

■ We combine multimodal auxiliary imaging technology and leverage similar information between different MR modalities to enhance the reconstruction effect.

■ The experimental results show that our model provides better visual and quantitative reconstruction results at both low and high acceleration rates with various sampling masks.

Article

IDDNet: a deep interactive dual-domain convolutional neural network with auxiliary modality for fast MRI reconstruction

Yi Cao, and Hongwei Du ✉

School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Hongwei Du, E-mail: duhw@ustc.edu.cn

Abstract: Reconstructing a complete image accurately from an undersampled k-space matrix is a viable approach for magnetic resonance imaging (MRI) acceleration. In recent years, numerous deep learning (DL)-based methods have been employed to improve MRI reconstruction. Among these methods, the cross-domain method has been proven to be effective. However, existing cross-domain reconstruction algorithms sequentially link the image domain and k-space networks, disregarding the interplay between different domains, consequently leading to a deficiency in reconstruction accuracy. In this work, we propose a deep interactive dual-domain network (IDDNet) with an auxiliary modality for accelerating MRI reconstruction to effectively extract pertinent information from multiple MR domains and modalities. The IDDNet first extracts shallow features from low-resolution target modalities in the image domain to obtain visual representation information. In the following feature processing, a parallel interactive architecture with dual branches is designed to extract deep features from relevant information of dual domains simultaneously to avoid redundant priority priors in sequential links. Furthermore, the model uses additional information from the auxiliary modality to refine the structure and improve the reconstruction accuracy. Numerous experiments at different sampling masks and acceleration rates on the MICCAI BraTS 2019 brain and fastMRI knee datasets show that IDDNet achieves excellent accelerated MRI reconstruction performance.

Keywords: MRI reconstruction; deep learning; dual-domain; multimodal

CLC number: R318; TP18   **Document code:** A

1 Introduction

Magnetic resonance imaging (MRI) is a commonly utilized modality for medical diagnostic imaging and is renowned for its exceptional soft tissue contrast, high-resolution anatomical depiction, and noninvasive and radiation-free nature. Nevertheless, MRI acquisition takes a long time due to the requirement for complete sampling of the raw k-space data. Such a slow data acquisition process causes patient discomfort and is prone to motion artifacts, limiting the diagnostic capabilities of diseases.

Various sophisticated acceleration algorithms have been suggested to shorten the imaging time. Previously, significant advancements were achieved by utilizing two techniques: hardware-based parallel imaging (PI)[1] and compressed sensing (CS)-based magnetic resonance imaging (CS-MRI)[2]. The PI uses special phase array coils to spatially encode the difference in sensitivity of the phased array coil at the spatial position, thereby reducing the k-space data acquisition time. The main categories of PI reconstruction algorithms include GRAPPA[1], SENSE[3], SPIRiT[4] and E-SPIRiT[5]. In conventional CS-MRI[6], the image is assumed to be a sparse representation after sparse transformation, and the underdetermined problem is solved iteratively by combining regularization parameters. Typical transformations include wavelet transform[7], low-rank transform[8], and total variation (TV)[9].

Despite the significant achievements of traditional methods, their development in clinical practice is impeded by issues such as rigid parameter adjustments, inaccurate prior conditions and time-consuming iterative processes.

In recent years, various deep learning (DL)-based methods have been proposed to accelerate MRI reconstruction to recover high-fidelity images from undersampled data[10–22]. As shown in Fig. 1, MR data can be obtained in two distinct forms: the k-space and the image domain. Therefore, two methods are available: single-domain and cross-domain methods.

With respect to single-domain methods, Wang et al.[10] first applied a convolutional neural network (CNN) framework for MRI reconstruction in 2016. Schlemper et al.[11] used a deep cascaded CNN structure and proposed a DC layer to use prior k-space information. Dilated convolutions and residual connections were applied to the recursive dilated network (RDN)[12], the U-Net[23] structure-based network (U-RDN)[13] and the multiscale dilated network (MDN)[14] to balance the reconstruction performance and the network parameters. Several GAN-based networks have also been proposed, such as the DA-GAN[15] and cyclic loss GAN[16]. Han et al.[22] used U-Net and the annihilating filter-based low-rank Hanel matrix approach (ALOHA) only in the k-space. As shown in Fig. 2, information is consistent across different domains, yet it is presented in diverse formats. Thus, extracting features only in
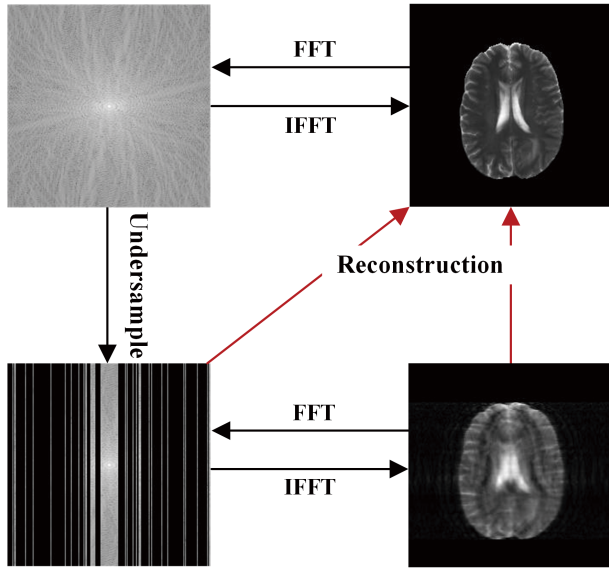
Fig. 1. Relationships among the different MRI modalities used for reconstruction. The FFT operator refers to the 2D fast Fourier transform, and the IFFT operator refers to the 2D inverse fast Fourier transform.
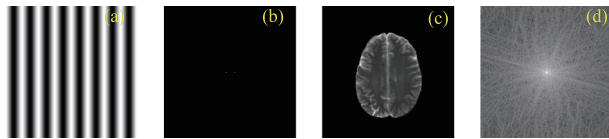

Fig. 2. Visualization of data in different domains. (a) Spatial (image) domain visualization of cosine waves; (b) frequency domain visualization of cosine waves; (c) image domain visualization of MR data; (d) frequency domain (k-space) visualization of MR data.

a single domain fails to use the potential information between multiple domains, thus affecting the reconstruction performance, such as excessive smoothness and lack of texture details.

Cross-domain methods leverage additional information from multiple domains. The cross-domain reconstruction methods can be divided into sequential cascade methods and dual-domain parallel methods. The earlier developed cascade method named KIKI-Net[17] used an interleave cascade framework of k-space and image data connected by fast Fourier transform (FFT) units and first verified the advantages of cross-domain methods over single-domain methods. On this basis, Wang et al.[18] proposed IKWI-Net to further extend the cascade framework by using the wavelet transform. Subsequently, Zhou et al.[19] proposed a dual-domain recurrent network, DuDoRNet, which added T1 prior to the cascade frame. It can be found that the cascade method based on the supplementary information of multiple domains is better than single domain reconstruction, but the order of different domains has always been an unavoidable problem. To avoid unnecessary a priori priorities and engage with information from two distinct domains simultaneously, Ran et al.[20] proposed MD-Recon-Net, a dual-domain parallel framework that treats k-space and image domains as two parallel processing networks. This approach allows for the utilization of dual-domain information and facilitates interaction without causing interference between sequences. Additionally, Liu et al.[21]

proposed DIIK-Net as a lightweight solution based on this parallel framework.

However, the existing parallel methods have two deficiencies: efficient feature extraction and a lack of comprehensive utilization of different MR modalities. First, the association between the information present in distinct data streams is frequently observed at an elevated level[24,25]. According to the aforementioned study, it is evident that shallow network layers have limited correlation extraction. The existing parallel methods use the direct integration of multidomain or multimodal information into the network, which leads to redundant and inefficient feature extraction. Additionally, the acquisition of multiple magnetic resonance (MR) sequences, including T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI), is often crucial in clinical settings to enhance the analysis and differentiation of lesions. Different MR sequences exhibit a strong correlation, and full-sampled T1 imaging can be used to assist undersampled T2 imaging, effectively shortening the overall imaging time[26,27]. The existing parallel dual-domain networks do not use multimodal auxiliary reconstruction technology, which affects the reconstruction accuracy.

In this work, we propose a parallel dual-domain interaction framework with an auxiliary modality to accelerate MRI reconstruction. To effectively utilize multidomain and multimodal information, we opt to perform feature extraction in two steps and utilize various feature extractors. Considering the widespread application of superresolution imaging[28] and denoising techniques[29], we use residual dense convolutional blocks for visual feature extraction. As the network deepens and the need for extracting multidomain and multimodal feature correlation information increases, we incorporate dilated convolutions[12,14] to expand the receptive field and integrate attention mechanisms[30,31] to associate contextual information. We compare our method with state-of-the-art methods and conduct ablation studies to validate the effectiveness of our network design in reducing artifacts and restoring texture structure.

The main contributions of our work are as follows:

● We propose a dual-domain parallel MRI reconstruction network, which consists of a single-domain preprocessing module, a dual-domain feature refinement module, and a feature integration module, to accelerate the reconstruction of multimodal MR images.

● According to the different feature extraction capabilities of the deep and shallow layers of the network, an image domain feature preprocessing module is added before the dual-domain parallel processing module to prerestore some aliased images, and the data consistency layer is used to constrain the sampled real data. Furthermore, feature extractors with different structures are designed to meet the needs of different feature processing methods.

● The feature refinement module is structured as a dual-branch parallel architecture to extract k-space and image domain features. The dual-domain interaction block is embedded in the parallel framework as a bridge for deep information interaction.

● The dual-domain information of the auxiliary modality is integrated into the dual-branch parallel framework to provide

additional feature information.

2 Methods

2.1 Notations

Denoting the 2D-complex values of full-sampled k-space and image as $y_f \in \mathbb{C}^{M \times N}$ and $x_f \in \mathbb{C}^{M \times N}$, the MRI reconstruction problem is equivalent to solving $x_f \in \mathbb{C}^{M \times N}$ from the undersampled k-space values $y_u \in \mathbb{C}^{M \times N}$ or the undersampled image values $x_u \in \mathbb{C}^{M \times N}$, such as:

$$y_u = M \odot y_f + \varepsilon = M \odot F x_f + \varepsilon = F_u x_f + \varepsilon, \qquad (1)$$

$$x_u = F^{-1} y_u, \qquad (2)$$

where ε is noise, $M \in \mathbb{C}^{M \times N}$ is the binary undersampling mask, $\odot$ represents elementwise multiplication, F is the 2D FFT operator, F^{-1} is the 2D inverse fast Fourier transform (IFFT) operator and $F_u \in \mathbb{C}^{M \times N}$ represents the undersampled Fourier encoding matrix. Because the undersampling process does not follow Nyquist's law, for the ill-posed optimization problem shown in Eq. (1), CS-MRI limits the potential solutions by utilizing prior knowledge, and the corresponding optimization problem can be formulated as:

$$\underset{x_f}{\mathrm{argmin}} \| F_u x_f - y_u \|_2^2 + \lambda R(x_f), \qquad (3)$$

where $\| F_u x_f - y_u \|_2^2$ represents the data fidelity and R represents regularization terms, which are typically the l_0 norm and l_1 norm in the sparsifying transform field of x_f. $\lambda \geq 0$ adjusts the balance of the two terms. In DL-based CS-MRI methods[11,32], parameters θ are optimized through network training to obtain an optimal reconstruction result close to the fully sampled image x_f. Then, the method can be expressed as:

$$\underset{x_f}{\mathrm{argmin}} \| F_u x_f - y_u \|_2^2 + \lambda \| x_f - f_{\mathrm{CNN}}(I|\theta) \|_2^2, \qquad (4)$$

where f_{CNN} is the forward mapping of the CNN network with parameters θ, I represents the input of the network and $f_{\mathrm{CNN}}(I|\theta)$ is the output of the model, which represents the reconstruction result. In our method, we use the strategy shown in expression (4), where I includes dual-domain data y_u and x_u. The training parameters are under the guidance of dual-domain full-sampled auxiliary data $x_{f,\mathrm{T1}}$ and $y_{f,\mathrm{T1}}$.

2.2 Data consistency

Data consistency (DC)[11] is a constraint strategy based on the raw data. We use the DC layer to constrain the information extraction and reintegration. The DC layer in the k-space domain can be defined as:

$$y_{\mathrm{dc}}(j) = \begin{cases} \bar{y}(j), & \text{if } j \notin \Omega, \\ \dfrac{\bar{y}(j) + \alpha y(j)}{1 + \alpha}, & \text{if } j \in \Omega, \end{cases} \qquad (5)$$

where $\bar{y}$ denotes the k-space data predicted by the network, y denotes the undersampled k-space data, j denotes the index of the k-space matrix, Ω denotes the set of sampled positions of the k-space matrix, and α is a hyperparameter.

For undersampled MR original data, we hope that the CNN network can effectively predict and fill the unsampled data points without performing any processing on the sampled real data points. The DC layer uses a linear weighted combination between the predicted and original sampled values. If the index of the k-space value does not fall into the sampling area ($j \notin \Omega$), the value predicted by the CNN modules is used; otherwise, the original value is used. Specifically, the DC layer is used at the end of feature extraction and domain transformation to ensure that the original sampled real data are not mispredicted by the CNN network, which is an effective a priori constraint method for the network. When it is necessary to implement DC layer constraint operations on image domain data, the domain transformation of the FFT and IFFT can be performed before and after the above formula.

2.3 Method overview

An overview of our method is illustrated in Fig. 3. It mainly consists of three parts: (i) the feature preprocessing (FP) module for the target undersampled image, which extracts the shallow image features of the target modality; (ii) the feature refinement (FR) module consisting of dual-domain information refinement branches and interaction modules, which extracts and interacts deep dual-domain features based on the additional information of the fused auxiliary modality in parallel; and (iii) the feature integration (FI) module, which integrates information extracted from each module to reconstruct the target modality image. The input and output of the dual-domain data in our experiments are both complex-valued matrices.

2.4 Feature preprocessing module

To avoid training with too much redundant information, we choose to perform a more comprehensive interaction in the dual domain at a deeper network level. Therefore, before proceeding with the process of multimodal fusion and multidomain correlation feature extraction, we propose the FP module to extract specific shallow features from the undersampled target modality T2 only in the image domain. Additionally, to avoid retraining with the initial sampling data, we use a DC layer to constrain the preprocessing procedure. The process can be formulated as:

$$x_{\mathrm{pre1}} = Conv(x_u), \qquad (6)$$

where $Conv$ denotes 32 3×3 kernel convolution operations. Then, x_{pre1} is sent to the RD block for stacking and merging shallow features of the target modality, and we can obtain x_{pre2} as follows:

$$x_{\mathrm{pre2}} = Conv(H_{\mathrm{RD}}(x_{\mathrm{pre1}})), \qquad (7)$$

where H_{RD} is the residual dense convolution operation by the RD block.

As shown in Fig. 4a, we use the RD block as the feature extractor to mine the visual features of images. Considering the utilization of preprocessing operations solely in shallow networks and exclusively in the T2 image domain, the complexity of the RD block was intentionally minimized. The dense blocks consist of multiple layers of 3×3 convolution
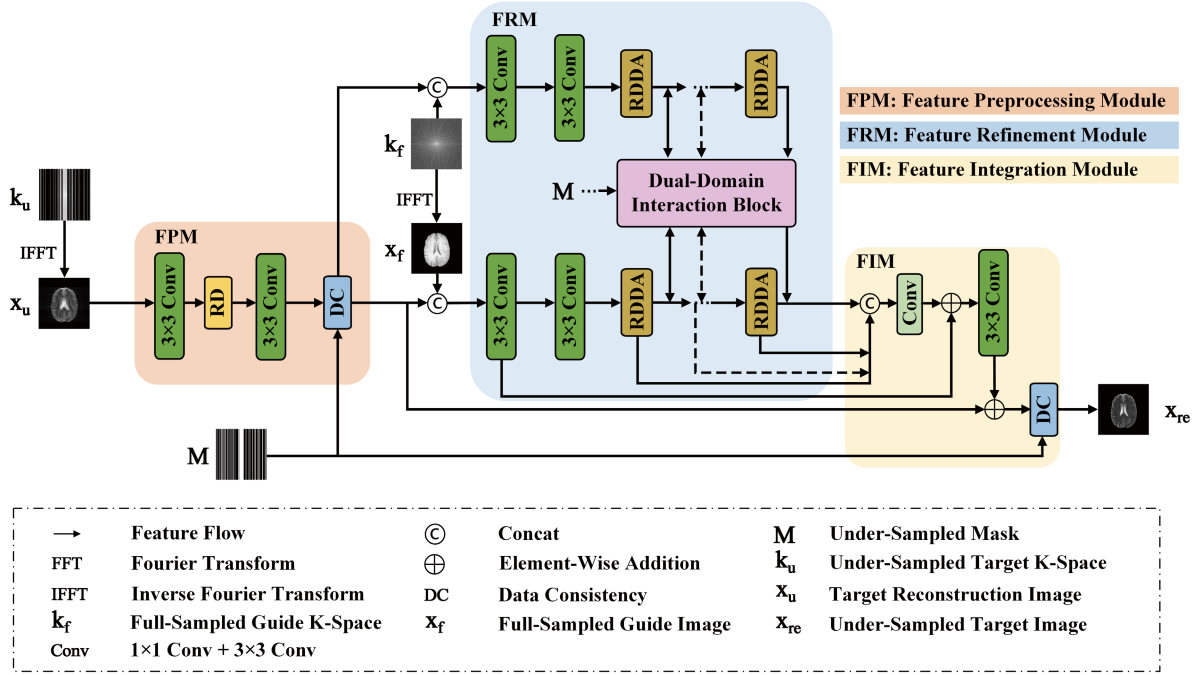
Fig. 3. Overview of the architecture of the proposed IDDNet.

and LeakyReLU units and global residual connections. The dense blocks transfer the feature extracted by the current layer to subsequent layers. Features of different depths are joined by a concatenation operation, and parameters are reduced by a 1×1 convolution layer. We add a long skip connection for global residual learning to preserve the initial characteristic properties of the target modality. After the prior constraints of H_{DC}, the image result of the FP module x_{pre} can be defined as:

$$x_{pre} = F^{-1}y_{pre} = H_{DC}(Fx_{pre2}). \qquad (8)$$

2.5 Feature refinement module

The core idea of IDDNet is deep interaction. We propose the feature refinement (FR) module to effectively leverage the complementary and associated multidomain and multimodal information, thereby enabling deep feature extraction. The bridge for the interaction between k-space and image domains is the dual-domain interaction (DDI) block, which simultaneously utilizes frequency domain features and image visual features to enhance reconstruction accuracy.

2.5.1 Dual-branch parallel frame

In the FR module, two branches, namely, K-Net and I-Net, operate simultaneously to process the dual-domain data. We splice the result of the preprocessed target modality T2 x_{pre} with the auxiliary modality T1, and the input x_{fuse} of I-Net is defined as:

$$x_{fuse} = H_{concat}(x_{pre}, x_{f,T1}). \qquad (9)$$

For K-Net, the input y_{fuse} is defined as:

$$y_{fuse} = H_{concat}(y_{pre}, y_{f,T1}). \qquad (10)$$

Then, the initial feature extraction of the fusion modality with two sequential 32 3×3 convolution layers is as follows:

$$x_{initial2} = Conv(x_{initial1}) = Conv(Conv(x_{fuse})), \qquad (11)$$

$$y_{initial2} = Conv(y_{initial1}) = Conv(Conv(y_{fuse})), \qquad (12)$$

where $x_{initial1}$ and $x_{initial2}$ denote the shallow features of the fusion in the image domain and $y_{initial1}$ and $y_{initial2}$ denote the shallow features of the fusion in the k-space domain.

2.5.2 Residual dilated dense attention block

With the deepening of network layers and the increasing demand for extracting multidomain and multimodal relevant information, we design residual dilated dense attention (RDDA) blocks as deep feature extractors. As shown in Fig. 4c, the RDDA block consists of two components, the residual dilated dense (RDD) block and the fusion attention (FA) block. Generally, we use I_{RDDA}^{i} as the input of the ith RDDA block. The process of feature extraction for one RDDA block is formulated as follows:

$$I_{RDDA}^{i'} = Conv(H_{concat}(H_{RDD}(I_{RDDA}^{i}), H_{FA}(I_{RDDA}^{i}))) + I_{RDDA}^{i}, \qquad (13)$$

$$H_{RDD} = Conv(H_{concat}(I_{RDDA}^{i}, I_{RDDA}^{mid})), \qquad (14)$$

where $I_{RDDA}^{i'}$ is the output of the ith RDDA block, H_{RDD} denotes the RDD block, H_{FA} is the fusion attention operation in the FA block and I_{RDDA}^{mid} denotes the output of the mid-results in the process of sequentially connecting dense layers with different dilation rates.

To design the RDD block, we incorporate dilated convolutions with different dilation rates into the dense convolution process of the RD block. This modification allows the model to have broader receptive fields and combines different receptive fields to extract features from multiple scales.

The FA block shown in Fig. 4b, which consists of a channel attention block and a spatial attention block, is used to ac-
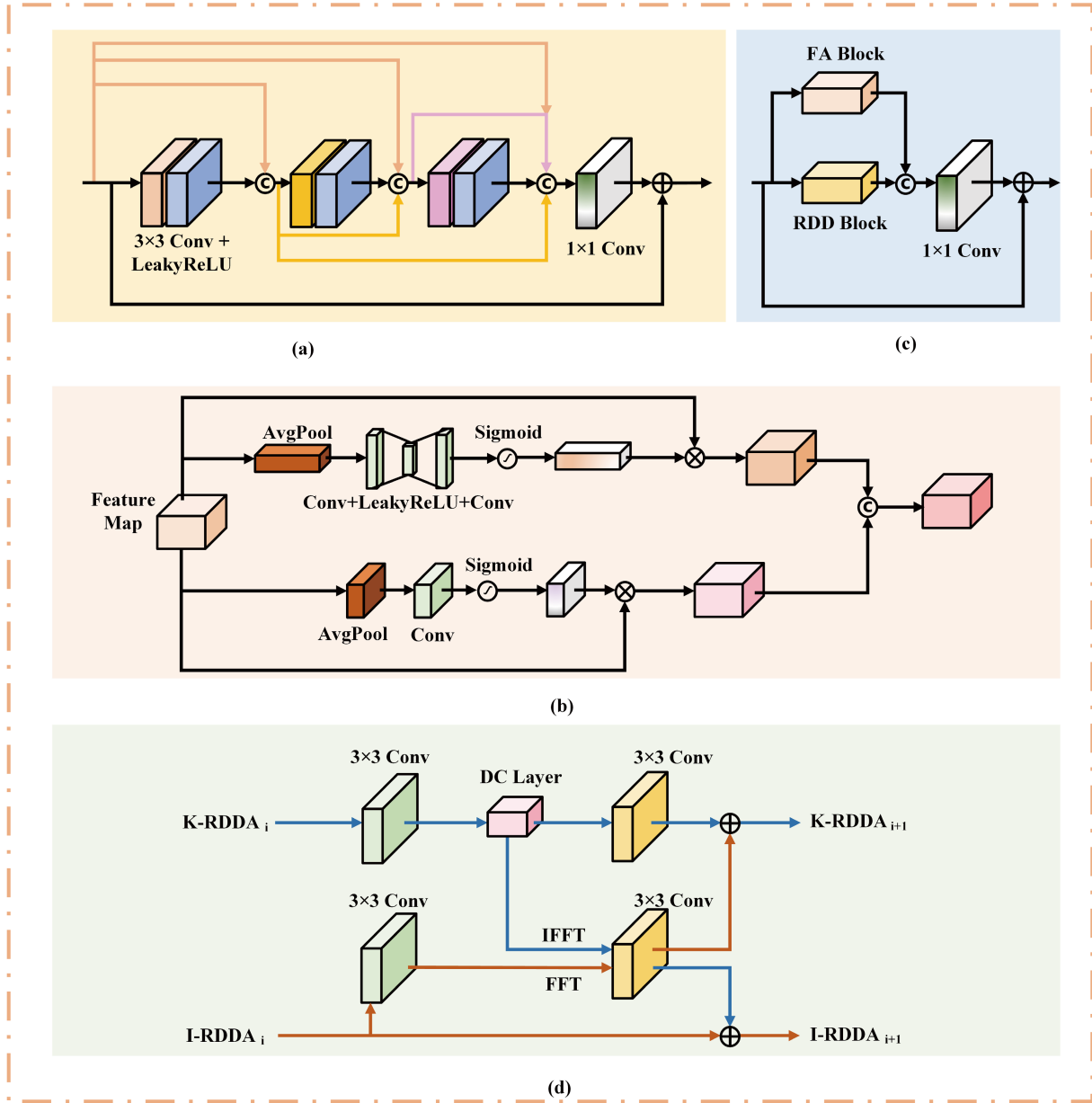
Fig. 4. Block diagrams of component modules. (a) The residual dense (RD) block, which contains several 3×3 convolution and LeakyReLU combined convolution layers and one 1×1 convolution layer for dimensionality reduction. The growth rate of the dense blocks is set to 32, and the number of combined convolution layers is set to 6. (b) Fusion attention (FA) block. (c) The residual dilated dense attention (RDDA) block, which uses the residual dilated dense (RDD) block and FA block for feature extraction, and 1×1 convolution for channel dimensionality reduction. The RDD module replaced the ordinary convolution in Fig. 4(a) with dilated convolution of different dilation rates. (d) Dual-domain interaction (DDI) block.

quire key similar features. The operation of H_{FA} is formulated as:

$$H_{\text{concat}}(I_{\text{RDDA}}^i \otimes \sigma(H_{\text{mlp}}(H_{\text{pool}}(I_{\text{RDDA}}^i))), I_{\text{RDDA}}^i \otimes$$
$$\sigma(Conv(H_{\text{pool}}(I_{\text{RDDA}}^i)))). \tag{15}$$

2.5.3 Dual-domain interaction block

One of the main steps of this paper is the utilization of an interactive mechanism to extract deep features from paired dual-domain data rather than processing frequency domain and image domain data independently. We propose embedding dual-domain interaction (DDI) blocks between RDDA blocks to achieve dual-domain interaction.

As shown in Fig. 4d, we perform a DC operation in the interaction block. We apply the DC layer on K-Net to enforce constraints based on the original *k*-space data. Subsequently, the *k*-space features are connected and fused with the image domain features after applying the DC constraint and FFT. Additionally, the image domain features are residually connected to the *k*-space domain features after applying the DC constraint and IFFT. Since the domain transformation needs to reduce the data with 2 channels, we perform channel dimension scaling through the convolution operation before and after the domain transformation.

In summary, the interaction process of K-Net with I-Net can be formulated as follows:

$$I_k^{i'} = Conv(H_{DC}(I_{RDDA-k}^{i'})), \qquad (16)$$

$$I_{RDDA-img}^{i+1} = Conv(F^{-1}I_k^{i'}) + I_{RDDA-img}^{i'}. \qquad (17)$$

Similarly, the interaction process of I-Net with K-Net can be formulated as follows:

$$I_{img}^{i'} = Conv(I_{RDDA-img}^{i'}), \qquad (18)$$

$$I_{RDDA-k}^{i+1} = Conv(FI_{img}^{i'}) + Conv(I_k^{i'}). \qquad (19)$$

In this way, the information from the image domain and frequency domain complements each other to achieve interactive feature extraction.

2.6 Feature integration module

In the FI module, we integrate the previously extracted features and reconstruct the final T2 image. The module first integrates the dual-domain interaction features extracted by the FR module. We only concatenate the intermediate results of all RDDA blocks in I-Net because the frequency domain features from each layer of the RDDA block have already been combined with each layer of I-Net. The initial feature of the FR block $x_{initial1}$ is added as a residual connection to the channel-compressed dual-domain refined feature. The fused feature is then compressed into 2 channels. The process can be formulated as follows:

$$x_{deep} = Conv(Conv(H_{concat}(I_{RDDA-img}^{1,2,\cdots,n})) + x_{initial1}), \qquad (20)$$

where $I_{RDDA-img}^{1,2,\cdots,n}$ denotes the intermediate results for all RDDA blocks.

Subsequently, under the constraints of the DC layer on the original data, the fused image feature x_{deep} and the preextracted feature x_{pre} establish a global residual connection to obtain the final reconstructed image. The reconstruction result x_{recons} is formulated as:

$$x_{recons} = H_{DC}(x_{deep} + x_{pre}). \qquad (21)$$

3 Experiments and results

3.1 Datasets and masks

MICCAI BraTS 2019[33,34], which consists of MR images of the brain, includes 460 pairs of publicly available multi-institutional MR sequences. Each group had four modalities

(T1-weighted, T2-weighted, T1-weighted with contrast enhancement, and FLAIR), and the matrix size was 155× 240×240. We split 335 pairs of data for training, 74 pairs for validation, and 51 pairs for testing.

The fastMRI dataset[35] is the largest publicly available raw MR dataset. We use knee single-coil data from the single-coil track of the fastMRI dataset. Following Ref. [36], 227 and 24 pairs of proton-density (PD) and fat-suppressed PD (PDFS) knee volumes, respectively, are selected for training and testing. As an auxiliary modality, PDs are used to guide the reconstruction of target modality PDFSs. We crop out an input of size 320 × 320 for each slice from the center region.

The original full k-space data are undersampled via elementwise multiplication with binary masks. Random and equispaced undersampling masks with 4× (8% central sampling rate) and 8× (4% central sampling rate) acceleration rates are used. Fig. 5 illustrates the masks used in our experiments.

3.2 Implementation details

Our experiments are performed on an NVIDIA GeForce GTX 2080Ti GPU with a memory size of 11 GB under the PyTorch framework. We use the Adam optimizer with parameters β_1=0.9 and β_2=0.999, and the learning rate is set to 1E−4. The model uses the L1 loss as the loss function. We train the model for 100 epochs with a batch size of 1. The number of RDDA blocks is set to 3. Referring to the hybrid dilated convolution (HDC)[37], the matrix of the dilation factor is set to [1, 2, 4, 1, 2, 4].

3.3 Comparisons with other methods

To evaluate the model's effectiveness, we compare our method on two datasets with three state-of-the-art learning-based methods, namely, U-Net, MD-Recon-Net, and DuDoRNet. The zero-filled reconstruction method (ZF) serves as the baseline for both datasets. The comparison methods are derived from the code of the fastMRI dataset and the original research papers. To ensure a fair evaluation, the training parameters of the competing methods are set to their default values.

3.3.1 Quantitative evaluation

The reconstruction results are evaluated by computing the NMSE, PSNR and SSIM between the reconstructed and ground truth images, which are effective ways to evaluate the quality of the reconstruction results. The quantitative evaluation results of the different methods are shown in Table 1 and Table 2. Our method yields the best results among al-
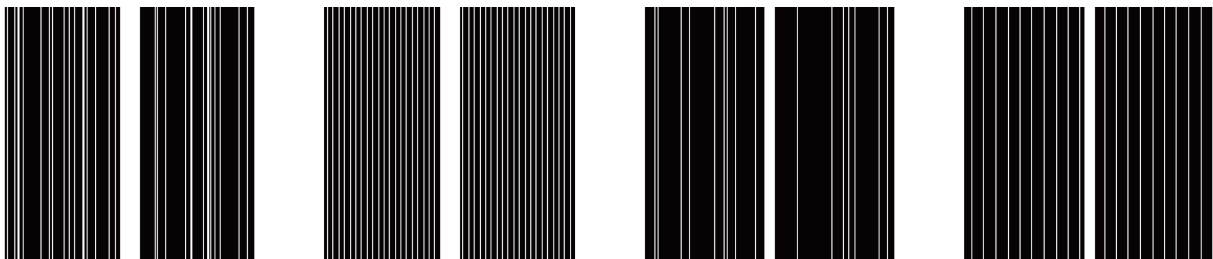

Fig. 5. The masks used in the experiments. The masks displayed in the sequence from left to right are labeled 4× Random mask, 4× Equispaced mask, 8× Random mask and 8× Equispaced mask.

most all methods with different acceleration rates and masks in the test dataset. For the single-domain method U-Net, our model achieves significant improvement. At high acceleration rates, the degradation is more obvious, and the artifacts of the zero-filled images are more obvious; however, our model still achieves better index results. In brain image reconstruction, our model improves the PSNR by nearly 7 dB and the SSIM by 0.0225 dB in comparison with the parallel interaction framework MD-Recon-Net under the 4× random sampling strategy. Compared with the domain-cascaded method DuDoRNet, which is the current best approach, the metrics of our model are also improved. In the reconstruction of knee images, only our model and DuDoRNet achieve PSNR values greater than 30 dB at 4× acceleration. All of the above findings prove that IDDNet is more effective than other dual-domain methods in information recovery.

3.3.2 Qualitative evaluation

Fig. 6 and Fig. 7 show the visual reconstruction results. We show the reconstruction results and error maps at different acceleration rates and for different templates. We also provide partial magnification to facilitate the observation of the details. The zero-filled reconstructed image has many artifacts, and many texture details are missing. Other methods can effectively suppress artifacts; however, IDDNet has the fewest artifacts and provides better visual reconstruction results at both low and high acceleration rates.

3.4 Ablation studies

Six experiments were conducted to assess the comparative

performance of various composition structures within our method. The outcomes were assessed using the BraTS 2019 test dataset. To mitigate the impact of randomness, experiments were performed with varying acceleration rates. The quantitative findings of the ablation experiments are presented in Table 3, while the qualitative results are depicted in Fig. 8.

3.4.1 Effectiveness of dual-domain Information

These experiments, referred to as "w/o Kspace" and "w/o Cross" in Table 3, aim to confirm the complementary validity of dual-domain information and the effectiveness of the proposed deep dual-domain interaction (DDI) block.

The exclusion of k-space information in the w/o Kspace operation results in a significantly lower overall index compared to the dual-domain architecture. Visualization results also demonstrate that k-space can offer more detailed structure and texture information, and the two-branch network effectively learns the representation of various domain features. The w/o Cross operation alters the cross-depth information exchange of K-I parallel networks. K-Net only provides input to I-Net, and I-Net solely receives K-Net features without reciprocal interactions. Quantitative indicators reveal that the removal of cross-operation impacts reconstruction performance, suggesting that DDI blocks play a crucial role in leveraging complementary information from different domains to enhance reconstruction.

3.4.2 Effectiveness of feature preprocessing

We verify the function of the feature preprocessing (FP) module, as indicated by the operation w/o FP in Table 3. Within

Table 1. Quantitative evaluation of reconstruction under different undersampling patterns in the BraTS 2019 dataset.

Method	Random 4×			Random 8×			Equispaced 4×			Equispaced 8×		
	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑
Zero-Filled	0.0545	26.56	0.5840	0.1110	23.41	0.5158	0.0561	26.46	0.5775	0.1113	23.38	0.5115
U-Net[23]	0.0114	33.46	0.9480	0.0352	28.55	0.8977	0.0087	34.77	0.9568	0.0315	29.11	0.9030
MD-Recon-Net[20]	0.0076	35.18	0.9588	0.0305	29.21	0.9038	0.0067	35.87	0.9634	0.0287	29.54	0.9075
DuDoRNet[19]	0.0021	41.86	0.9742	0.0090	35.49	0.9375	0.0020	42.15	0.9771	0.0076	36.03	0.9385
Ours	**0.0020**	**42.13**	**0.9813**	**0.0089**	**35.50**	**0.9445**	**0.0016**	**42.81**	**0.9852**	**0.0074**	**36.20**	**0.9495**

The best results are marked in bold. ↑ and ↓ indicate that a higher or lower value is better.

Table 2. Quantitative evaluation of reconstruction under different undersampling patterns in the fastMRI dataset.

Method	Random 4×			Random 8×			Equispaced 4×			Equispaced 8×		
	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑
Zero-Filled	0.0428	27.05	0.6056	0.0645	25.26	0.4781	0.0428	27.05	0.6064	0.6471	25.25	0.4770
U-Net[23]	0.0317	28.27	0.6486	0.0447	26.73	0.5377	0.0306	28.43	0.6566	0.0428	26.92	0.5425
MD-Recon-Net[20]	0.0320	28.27	0.6501	0.0441	26.80	0.5361	0.0300	28.52	0.6594	0.0438	26.82	0.5334
DuDoRNet[19]	0.0205	30.26	0.7873	0.0335	28.10	0.6458	0.0201	30.21	0.7878	0.0331	28.15	0.6528
Ours	**0.0201**	**30.34**	**0.7896**	**0.0326**	**28.20**	**0.6520**	**0.0199**	**30.39**	**0.7913**	**0.0323**	**28.26**	**0.6544**

The best results are marked in bold. ↑ and ↓ indicate that a higher or lower value is better.
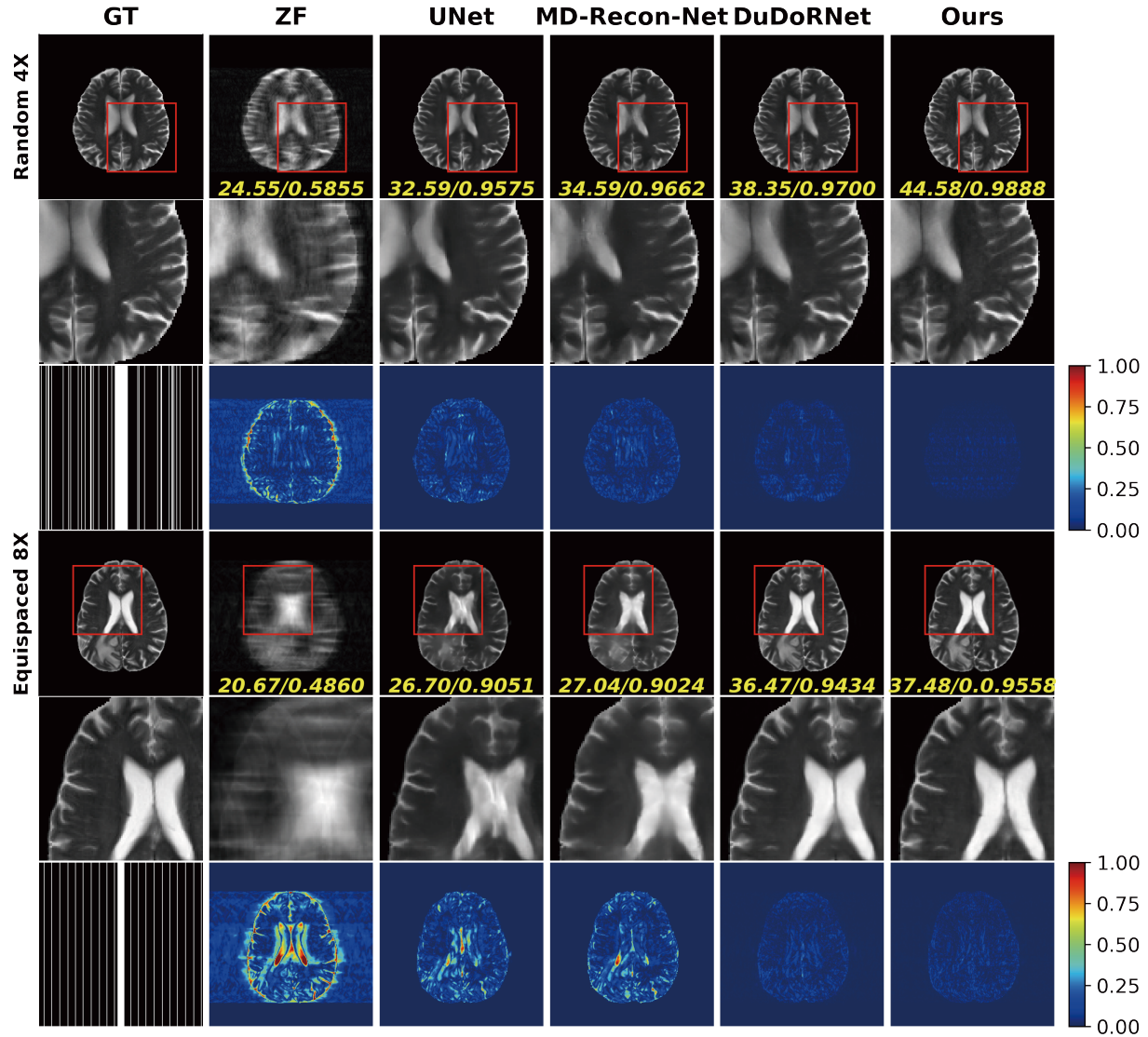
Fig. 6. Visual comparisons and error maps of IDDNet with the state-of-the-art methods on the BraTS 2019 dataset.

the FP module, we perform the extraction of the visual feature of the target modality and leverage the original prior information through the constraint of the DC layer. In this experiment, we remove the entire FP module and replace the input of the FR module with zero-filled T2 images and initial undersampled k-space data, while the global residual connection of the FI module is also replaced with zero-filled T2 images. The results in Table 3 and Fig. 8 prove that the separate shallow feature extraction of the target modality by the FP module reduces artifacts.

3.4.3 Effectiveness of multimodal information

The coupling relationship between multimodal features provides another kind of prior information for the recovery of undersampled target modalities. In our model, the auxiliary modality information also has dual-domain complementary interactions. In the w/o T1-GA operation, we remove the T1 input and retain only the undersampling modality T2 as the input. As the results show, the indicators decrease, and T1 assistance combined with dual-domain information provides a more detailed subtle structure for reconstruction and achieves better quantization performance.

3.4.4 Effectiveness of feature refinement

In the w/o H-conv operation, dilated convolutions are removed, and in the w/o AM operation, we remove the fusion attention blocks. After removing the components, there are small decreases in both the quantitative and qualitative results.

4 Conclusions

In this work, we propose a deep interactive dual-domain network (IDDNet) with auxiliary modality for MRI reconstruction. We employ a shallow-to-deep feature extraction approach to facilitate the reconstruction of the T2 modality. Initially, the distinctive data from T2 are isolated through the shallow feature preprocessing module. Subsequently, a dual-domain parallel structure is employed to combine similar information from T1, enabling the extraction of diverse representations from both the k-space and image domains. Ultimately, feature integration is accomplished through the global residual connection. Extensive comparison and ablation experiments demonstrate the effectiveness of
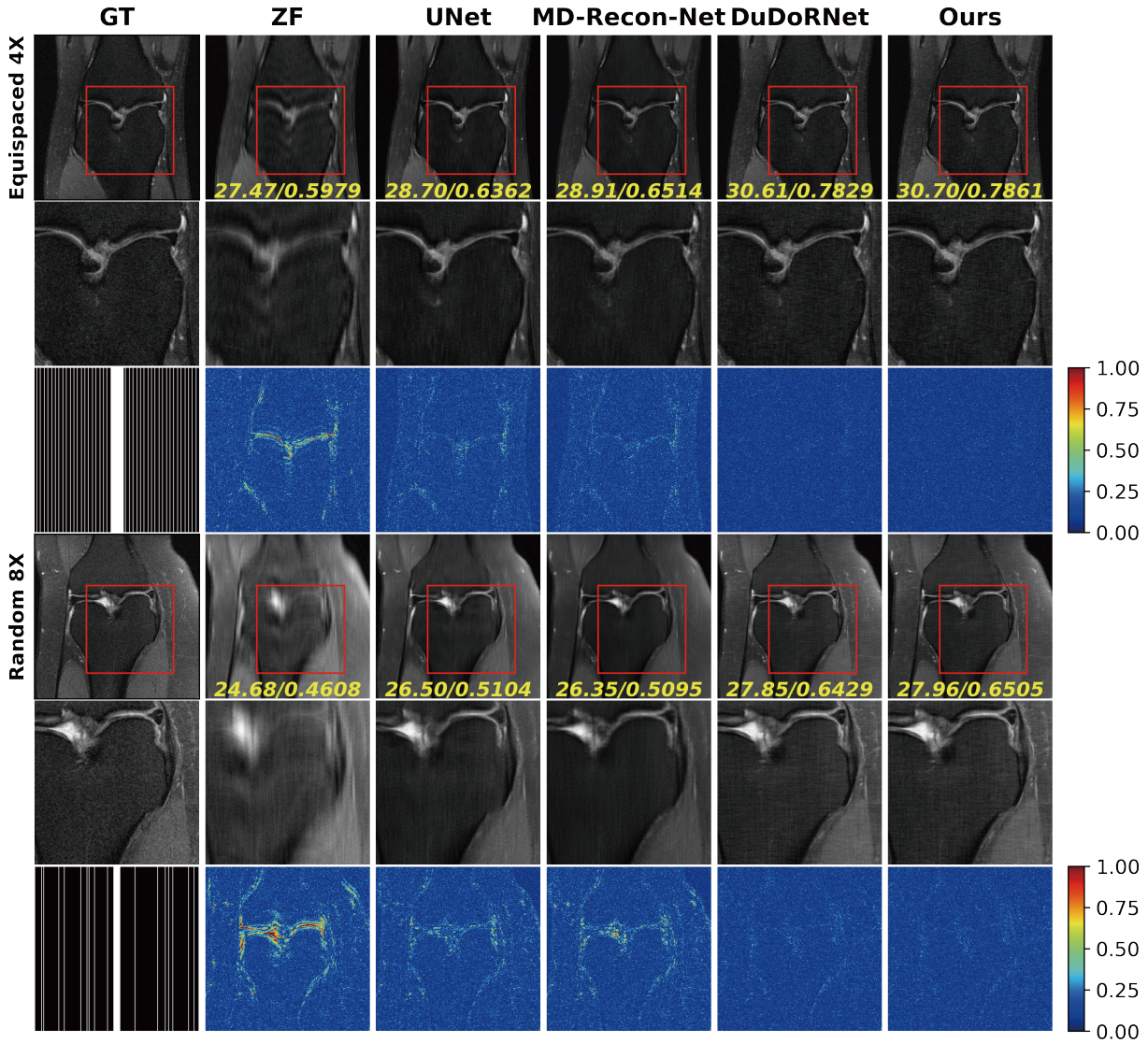
Fig. 7. Visual comparisons and error maps of IDDNet with the state-of-the-art methods on the fastMRI dataset.

Table 3. Quantitative evaluations of ablation studies on the BraTS 2019 dataset.

Method	Random 4×			Random 8×		
	NMSE↓	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑
w/o Kspace	0.0052	37.28	0.9724	0.0158	32.62	0.9391
w/o Cross	0.0037	39.97	0.9628	0.0101	34.90	0.9149
w/o FP	0.0031	40.14	0.9696	0.0102	34.98	0.9270
w/o T1-GA	0.0026	40.25	0.9749	0.0119	33.61	0.9347
w/o H-conv	0.0022	41.73	0.9777	0.0098	35.14	0.9386
w/o AM	0.0021	41.80	0.9787	0.0092	35.32	0.9410
Ours	**0.0020**	**42.13**	**0.9813**	**0.0089**	**35.50**	**0.9445**

The best results are marked in bold. ↑ and ↓ indicate that a higher or lower value is better.

IDDNet in reducing artifacts and restoring texture structure. Furthermore, the inclusion of an auxiliary modality leads to further enhancements in both reconstructed image quality and perception quality.
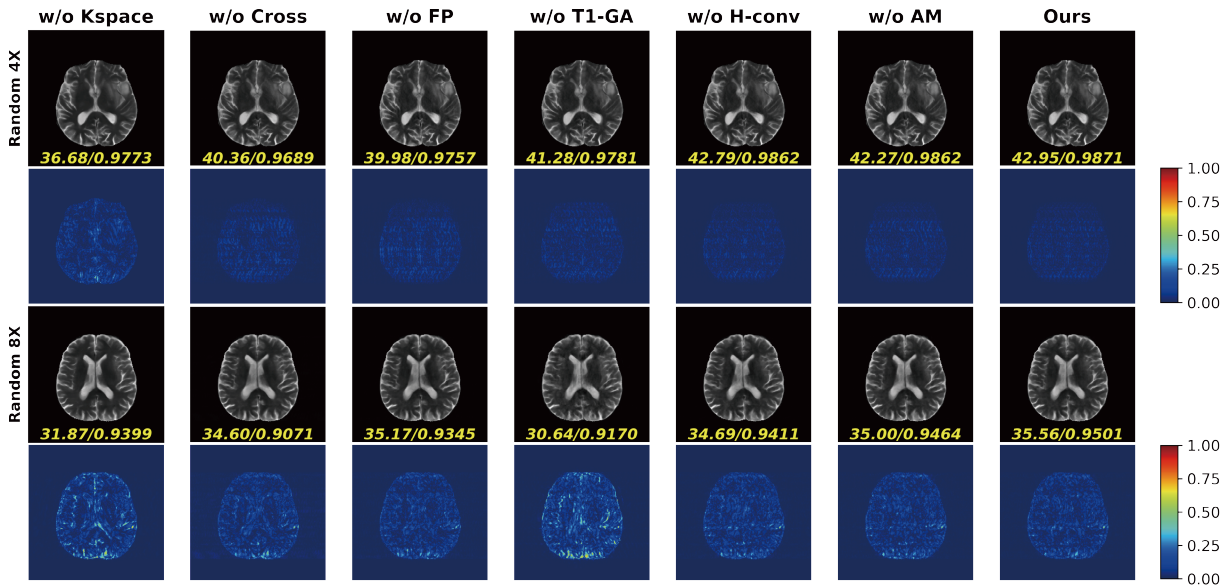
Fig. 8. Visual comparisons of IDDNet and its variations on the BraTs2019 dataset.

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Yi Cao is currently pursuing a master's degree at the University of Science and Technology of China. Her research focuses mainly on fast MRI reconstruction based on deep learning.

Hongwei Du is currently an Associate Professor at the School of Information Science and Technology, University of Science and Technology of China (USTC). He received his Ph.D. degree from the Department of Electronic Science and Technology, USTC, in 2007. His research interests include medical image processing based on deep learning, fast MRI imaging principle and algorithm, and advanced biomedical imaging system.

References

[1] Griswold M A, Jakob P M, Heidemann R M, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, **2002**, *47* (6): 1202–1210.

[2] Lustig M, Donoho D, Pauly J M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, **2007**, *58* (6): 1182–1195.

[3] Pruessmann K P, Weiger M, Scheidegger M B, et al. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, **1999**, *42* (5): 952–962.

[4] Lustig M, Pauly J M. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary *k*-space. *Magnetic Resonance in Medicine*, **2010**, *64* (2): 457–471.

[5] Uecker M, Lai P, Murphy M J, et al. ESPIRiT —an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, **2014**, *71* (3): 990–1001.

[6] Lustig M, Donoho D L, Santos J M, et al. Compressed sensing MRI. *IEEE Signal Processing Magazine*, **2008**, *25* (2): 72–82.

[7] Qu X, Cao X, Guo D, et al. Combined sparsifying transforms for compressed sensing MRI. *Electronics Letters*, **2010**, *46* (2): 121–123.

[8] Haldar J P, Zhuo J W. P-LORAKS: Low-rank modeling of local *k*-space neighborhoods with parallel imaging data. *Magnetic Resonance in Medicine*, **2016**, *75* (4): 1499–1514.

[9] Block K T, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*, **2007**, *57* (6): 1086–1098.

[10] Wang S S, Su Z H, Ying L, et al. Accelerating magnetic resonance imaging via deep learning. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). Prague, Czech Republic: IEEE, **2016:** 514–517.

[11] Schlemper J, Caballero J, Hajnal J V, et al. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, **2018**, *37* (2): 491–503.

[12] Sun L Y, Fan Z W, Huang Y, et al. Compressed sensing MRI using a recursive dilated network. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). New Orleans, LA: AAAI, **2018:** 2444–2451.

[13] Ding P L K, Li Z Q, Zho Y X, et al. Deep residual dense U-Net for resolution enhancement in accelerated MRI acquisition. In: Proc. SPIE 10949, Medical Imaging 2019: Image Processing. SPIE, **2019**, 10949: 109490F.

[14] Dai Y X, Zhuang P X. Compressed sensing MRI via a multi-scale dilated residual convolution network. *Magnetic Resonance Imaging*, **2019**, *63*: 93–104.

[15] Yang G, Yu S M, Dong H, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, **2018**, *37* (6): 1310–1321.

[16] Quan T M, Nguyen-Duc T, Jeong W K. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Transactions on Medical Imaging*, **2018**, *37* (6): 1488–1497.

[17] Eo T, Jun Y, Kim T, et al. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, **2018**, *80* (5): 2188–2201.

[18] Wang Z L, Jiang H T, Du H W, et al. IKWI-net: A cross-domain convolutional neural network for undersampled magnetic resonance image reconstruction. *Magnetic Resonance Imaging*, **2020**, *73*: 1–10.

[19] Zhou B, Zhou S K. DuDoRNet: learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, **2020:** 4272–4281.

[20] Ran M S, Xia W J, Huang Y Q, et al. MD-Recon-Net: a parallel dual-domain convolutional neural network for compressed sensing MRI. *IEEE Transactions on Radiation and Plasma Medical Sciences*, **2021**, *5* (1): 120–135.

[21] Liu Y, Pang Y W, Liu X H, et al. DIIK-Net: A full-resolution cross-domain deep interaction convolutional neural network for MR image reconstruction. *Neurocomputing*, **2023**, *517*: 213–222.

[22] Han Y, Sunwoo L, Ye J C. *k*-space deep learning for accelerated MRI. *IEEE Transactions on Medical Imaging*, **2020**, *39* (2): 377–386.

[23] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, et al. , editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer, **2015**: 234–241.

[24] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, **2006**, *313* (5786): 504–507.

[25] Martínez H P, Yannakakis G N. Deep multimodal fusion: combining discrete events and continuous signals. In: ICMI '14: Proceedings of the 16th International Conference on Multimodal Interaction. New York: ACM, **2014**: 34–41.

[26] Xiang L, Chen Y, Chang W T, et al. Deep-learning-based multi-modal fusion for fast MR reconstruction. *IEEE Transactions on Biomedical Engineering*, **2018**, *66* (7): 2105–2114.

[27] Xuan K, Xiang L, Huang X Q, et al. Multimodal MRI reconstruction assisted with spatial alignment network. *IEEE Transactions on Medical Imaging*, **2022**, *41* (9): 2499–2509.

[28] Zhang Y L, Tian Y P, Kong Y, et al. Residual dense network for image super-resolution. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, **2018**: 2472–2481.

[29] Kim D W, Chung J R, Jung S W. GRDN: grouped residual dense network for real image denoising and GAN-based real-world noise modeling. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, **2019**: 2086–2094.

[30] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, **2018**: 7132–7141.

[31] Woo S H, Park J, Lee J Y, et al. CBAM: convolutional block attention module. In: Ferrari V, Hebert M, Sminchisescu C, et al. , editors. Computer Vision – ECCV 2018. Cham: Springer, **2018**: 3–19.

[32] Qin C, Schlemper J, Caballero J, et al. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, **2019**, *38* (1): 280–290.

[33] Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv: 1811.02629, **2018**.

[34] Menze B H, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, **2015**, *34* (10): 1993–2024.

[35] Zbontar J K F, Sriram A. fastMRI: an open dataset and benchmarks for accelerated MRI. arXiv: 1811.08839, **2018**.

[36] Xuan K, Sun S, Xue Z, et al. Learning MRI *k*-space subsampling pattern using progressive weight pruning. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer, **2020**: 178–187.

[37] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, **2018**: 1451–1460.