TRANS-Net: Transformer-Enhanced Residual-Error AlterNative Suppression Network for MRI Reconstruction

Dianlin Hu, Yikun Zhang, Jianfeng Zhu, Qiegen Liu, Yang Chen

TRANS-Net: Transformer-enhanced Residual-error AlterNative Suppression Network for MRI Reconstruction

Dianlin Hu, Yikun Zhang, Jianfeng Zhu, Qiegen Liu, *Senior Member, IEEE*, Yang Chen, *Senior Member, IEEE*

***Abstract*—Since deep priors could exploit more intrinsic features than hand-crafted prior knowledge, unrolled reconstruction methods significantly improve image quality for fast MRI reconstruction with the combination of iterative optimization and deep neural network-based regularization terms. One popular way for unrolled methods is to employ the regularization penalty on the reconstructed result within the image space. Dissimilarly, in this paper, we innovatively propose a reconstruction framework termed Transformer-enhanced Residual-error AlterNative Suppression Network (TRANS-Net) for MRI reconstruction to extend the unrolled methods. Unlike some existing model-based algorithms, TRANS-Net additionally deploys the regularization term on the error maps in the residual image domain, which greatly emphasizes the high-frequency information contained in the MR images. Besides, to further explore the global spatial correlation of the reconstructed image, the transformer module is adopted to enlarge the receptive field of deep priors regularized on the reconstructed images for better tissue recovery. The quantitative and qualitative results show that the proposed TRANS-Net method has superior performance to the state-of-the-art reconstruction algorithms in anatomical structure restoration and perceptual detail preservation.**

***Index Terms*—MRI imaging, unrolled reconstruction method, deep prior, residual image domain, transformer.**

This work was supported in part by the Science and Technology Program of Guangdong (2018B030333001), in part by the State Key Project of Research and Development Plan under Grant 2022YFC2401600, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and development Programs in Jiangsu Province of China under Grant BE2021703 and BE2022768. Yang Chen is with the Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China, and also with the Laboratory of Image Science and Technology, the School of Computer Science and Engineering, Southeast University, Nanjing 210096, China (Corresponding authors: Qiegen Liu; Yang Chen).

Dianlin Hu, Yikun Zhang, and Yang Chen are with the Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China, and also with the Centre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), Rennes, F-3500 France, and also with the Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China (e-mail: dianlin@seu.edu.cn; yikun@seu.edu.cn; chenyang.list@seu.edu.cn).

Qiegen Liu is with the Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China (e-mail: liuqiegen@ncu.edu.cn).

Jianfeng Zhu is with the Xingaoyi Medical Equipment Co., Ltd., Yuyao, China (e-mail: zhujf@china-mri.com).

I. Introduction

MAGNETIC Resonance Imaging (MRI) is a widely used medical imaging tool for numerous clinical applications due to its non-invasive ability to reveal structural, anatomical and functional information [1]. However, the main drawback of MRI is the slow acquisition process, which will lead to degraded images because of the patient movement and physiological motion. Meanwhile, the prolonged acquisition time also restricts the large-scale usage of MRI since the expensive cost and less comfort [2]. Therefore, how to accelerate the acquisition process for MRI is a hot research topic. One possible way for the MR imaging time reduction is to undersample the measurement data in k-space. However, the undersampled k-space data will generate corrupted images inundated with severe aliasing artifacts because of the Nyquist-Shannon theorem [1, 3]. To tackle this issue, many reconstruction algorithms have been developed to improve image quality.

Compressed Sensing (CS)-based reconstruction methods have attracted great attention in medical imaging [4]. Noticing the sparseness property of MRI images in some transform domains, some methods have been proposed for fast MRI reconstruction. Specifically, Block *et al.* introduced the total variation (TV) as the regularization constraint for multi-coils MRI reconstruction to suppress streaking artifacts and remove noise [5]. Later, the total generalized variation (TGV) was proposed to extend TV theory and achieved better performance in noise reduction and edge preservation over the conventional TV minimization [6]. Besides the mentioned constraint functions, sparse representation was another commonly used regularization prior [7]. Ravishankar *et al.* employed dictionary learning as the penalty function to enforce the sparsity for the local structure, which demonstrated dramatic improvements compared to previous CS methods [8]. Further, Huang *et al.* incorporated bayesian statistics with dictionary learning to formulate a novel nonparametric model (BPFA) for MR imaging [9]. BPFA gave an impressive reconstruction result that benefited from adaptively learning the hyper-parameters. As an extension of sparsity prior, low rank has been widely applied for dynamic MRI by exploiting the correlations between spatiotemporal MR images [10, 11]. Realizing the non-local similarity in MR images, non-local means (NLM)

filter was induced for noise suppression and gained perceptually improved images [12, 13]. Although CS-based MR imaging (CS-MRI) methods exhibited remarkable performance in aliasing artifact removal and anatomical structure preservation from the undersampled k-space data, they still had some defects in the sensitive hyper-parameter selection, high computational cost and hand-designed priors.

Recently, deep learning (DL)-based methods have been successfully applied in inverse problems [14, 15]. Particularly, convolutional neural networks (CNN) have become the state-of-the-art technique for various medical imaging tasks [16, 17]. Assisted by the large-scale training datasets, Wang *et al.* designed a convolutional neural network to learn the mapping relationship between the corrupted image and the reference image [18]. Experiments implied that this method could successfully restore fine structures and details. Then, to stabilize the training process and accelerate the convergence, residual learning was utilized for fast MRI reconstruction [19]. These CNN-based methods only exploited the objective function from a single domain to optimize network weights, which may fail to use prior knowledge in other domains. Therefore, Yang *et al.* developed an innovative CNN model, which involved image domain loss, frequency domain loss, feature space loss and adversarial loss, resulting in superior reconstruction images with visually promoted details and textures [1]. Unlike the cost function defined independently in various transform domains, a cyclic data consistency loss was deployed in [20] to simultaneously underline the consistency between the reconstructions and references in k-space and the image space. Based on the [20], Liu *et al.* adopted the data augmentation training strategy to strengthen the robustness of the proposed method (SANTIS) against undersampling pattern discrepancy [21]. Compared to CS-MRI methods, SANTIS could provide better images with sharper edges. For the comprehensive information contained in frequency and image domains, the cross-domain-based framework (KIKI-Net) was proposed, leading to improved results over the single-domain neural network [22]. Further, a multi-domain CNN (MD-CNN) was investigated for cardiac MR cine imaging [23]. Despite these methods possessing outstanding performance and short imaging time, the acquisition of large-scale noise-free datasets limited their applications in clinical practice.

Replacing the explicitly hand-crafted priors with the implicitly deep neural networks, unrolled reconstruction methods have presented great potential in image reconstruction. Zhang *et al.* employed the neural network to replace the ℓ_1-norm constraint based on the ISTA method and gained impressive results for image compressive sensing [15]. Ma *et al.* directly unfolded the ADMM optimization for hyperspectral image super-resolution [24]. To accelerate the convergence of subproblems, MetaInv-Net adopted gradient descent with a learned initializer for sparse-view CT reconstruction [25]. Unlike regularizing the image by the classical CNN model, Xia *et al.* introduced graph convolution to boost the denoising performance for low-dose CT [26]. In MRI reconstruction, there are also many unrolled methods. Schlemper *et al.* designed sequential CNNs and employed data consistency layers to boost the reconstruction image quality [3]. At the moment, to overcome the drawbacks of CS-MRI in hyper-parameter settings and hand-designed regularization constraints, an unrolled variational model (VN) was performed by gradient descent optimization [27]. Further, Aggarwal *et al.* established a model-based reconstruction method (MoDL) to solve data fidelity minimization in more complex MR imaging using a conjugate gradient optimization strategy [28]. Data fidelity term was a necessary component in unrolled reconstruction methods, and it was usually optimized by ℓ_2-minimization. However, ℓ_2-distance might not be the optimal metric function because of ignoring the real distribution of noise [29]. Hence, to resolve this issue, Cheng *et al.* employed the neural network to learn the data consistency term. Most existing unrolled methods only underlined the sparse prior. Therefore, Ke *et al.* combined low rank and sparsity to promote the reconstruction results for dynamic MR cine imaging [30]. These unrolled methods were derived from CS-MRI, so they were more interpretable than CNN-based methods. Meanwhile, the regularization terms were automatically learned from the datasets. Hence, they usually outperformed hand-designed priors [31, 32].

Excluding the k-space domain-based processing methods [22, 23, 29], both CS-MRI and DL-based methods merely deployed regularization constraints on the reconstructed results in the image domain to improve the image quality. However, this scheme may limit further improvement for MRI reconstruction because it is hard to directly promote high-frequency information based on the reconstructed image. Therefore, this paper proposes a Transformer-enhanced Residual-error AlterNative Suppression Network (TRANS-Net) for single-coil MRI reconstruction. It is an unrolled method that employs deep neural networks as the regularization constraints. The main contributions of this work are three-fold. First, most existing unrolled methods stress the image property of the reconstructed results, e.g., image gradient sparsity or low rank [27, 30], to remove aliasing artifacts and recover anatomical structures. Differing from these methods, TRANS-Net also regularizes the error MR images on the residual domain, reducing the difficulty of finding the optimal solution for detail preservation. Second, most unrolled methods usually utilize small neural networks in each iteration, focusing only on the local features. To overcome this problem, TRANS-Net applies the transformer module to encode the global information effectively. Third, TRANS-Net conducts two light U-shape-based neural networks to improve the representation capacity of the deep priors. Two datasets were performed to validate the superior performance of TRNAS-Net in detail preservation and artifact removal.

The rest of this paper is organized as follows. Section II introduces the background of unrolled methods. The mathematical model of TRANS-Net and the architecture of the neural network are given in Section III. Section IV presents the experiments and analyses for the proposed method. In Section V, we will discuss some related issues and make conclusions.

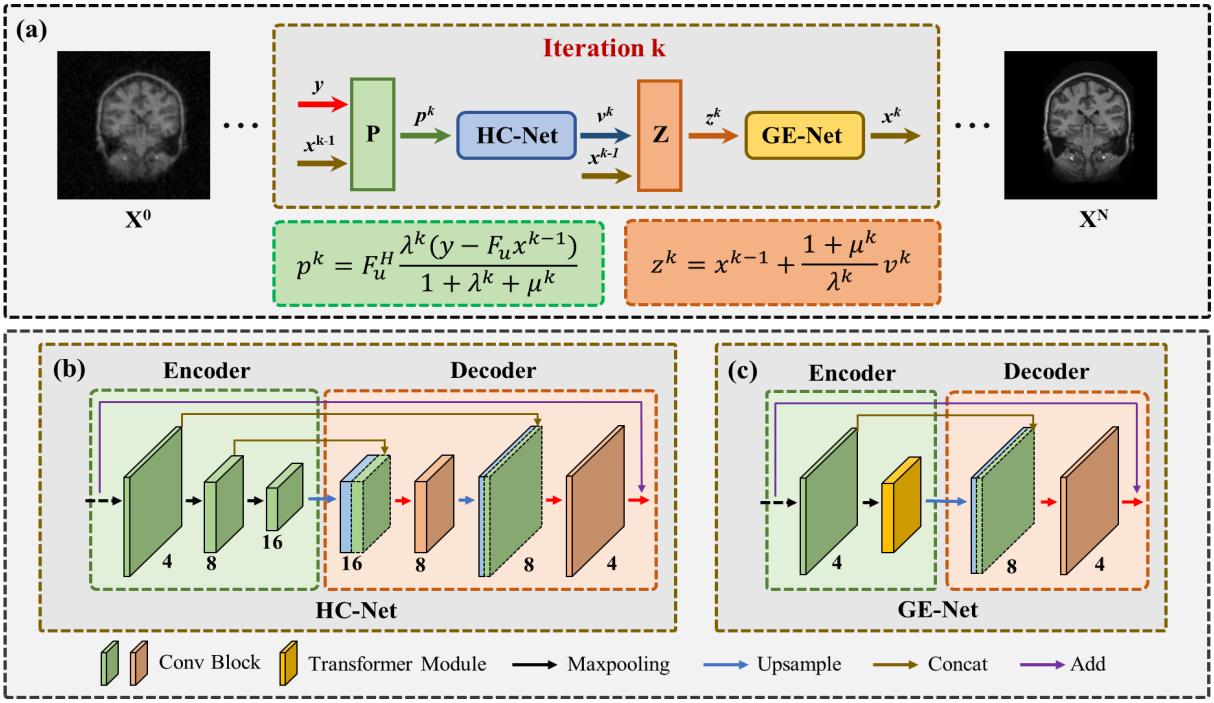

Fig. 1. The flowchart of TRANS-Net. (a) The workflow of the proposed method, (b) The architecture of HC-Net, (c) The architecture of GE-Net.

II. Background

A. Compressed Sensing-based MRI Model

The classic CS-MRI model can be expressed as follows:

$$\min_x \frac{1}{2}\|F_u x - y\|_2^2 + \lambda R(x) \tag{1}$$

where the first term is data fidelity term, F_u denotes the undersampled Fourier transform, y presents the undersampled k-space measurement data, x stands for the reconstructed MR image, $R(\cdot)$ expresses the regularization function on the x, λ indicates the hyper-parameter to balance the trade-off between the data fidelity term and the regularization term. By solving the optimization problem Eq. (1) [5, 6, 8], a high-quality image x can be obtained.

B. Unrolled MRI Reconstruction Model

The hand-designed prior constraint R is the main obstacle of CS-MRI because it cannot intrinsically expose the image property. Hence, some unrolled methods are investigated to overcome the shackle of CS-MRI via using the deep priors trained by MR images themselves [33-35]. The typical unrolled method can be formulated as follows:

$$\min_x \frac{\lambda}{2}\|F_u x - y\|_2^2 + \frac{1}{2}\|x - f_\theta(x_u)\|_2^2 \tag{2}$$

here $x_u = F_u^H y$ is the MR image reconstructed from the zero-filled undersampled k-space data y, f_θ represents the trained neural network with the weighting parameter θ. Eq. (2) can be iteratively solved by alternatively updating the regularization term and data consistency term, which are briefly expressed as [3]:

$$z^{k+1} = f_{\theta^{k+1}}(x^k) \tag{3}$$

$$u^{k+1} = \begin{cases} s_{z^{k+1}}(j) & \text{if } j \in \Omega \\ \frac{s_{z^{k+1}}(j) + \lambda y(j)}{1+\lambda} & \text{if } j \notin \Omega \end{cases} \tag{4}$$

$$x^{k+1} = F^H u^{k+1} \tag{5}$$

where $f_{\theta^{k+1}}$ is the deep neural network at $k+1^{\text{th}}$ iteration, Ω indicates the subset of indices sampled in k-space, $s_{z^{k+1}} = Fz^{k+1}$ is the frequency data of z^{k+1}, F denotes the Fourier transform, j stands for the j^{th} coefficient in Ω, F^H presents the inverse Fourier transform. Eqs. (3)-(5) give the classical workflow of the unrolled method, which is simple but effective [3].

III. TRANS-Net Model

A. Mathematical Model of TRANS-Net

Most unrolled methods make great efforts to improve the MR image quality according to Eq. (1). However, those methods only regularize the reconstructed results x on the image domain, which may ignore the residual components in the k-space domain as follows.

$$r^k = y - F_u x^k \tag{6}$$

where x^k is the reconstructed image at the k^{th} iteration and r^k stands for the residual map between the x^k and y in the frequency domain. Compared to x^k, r^k indicates the high-frequency error and can be used to compensate for the x^k in respect to subtle detail missing. Therefore, inspired by [36], Eq. (1) can be extended by adding one extra regularization term on the k-space residual map r^k, whose objective function is written as:

$$\min_{x,r} \frac{\lambda}{2}\|y - (F_u x + r)\|_2^2 + \frac{1}{2}\|x - \Gamma(F_u x + r)\|_2^2 + \frac{\mu}{2}\|r\|_2^2 + \zeta\|Wx\|_1 \tag{7}$$

where Γ stands for a complex function to accurately transform the k-space data into the image data. W indicates the sparsity transform, e.g., wavelet framework. $\lambda > 0$, $\mu \geq 0$, and $\zeta > 0$ are hyper-parameters to control the proportion of different terms. Differing from Eq. (1), Eq. (7) has two data fidelity terms in the k-space domain and image domain by additionally introducing the residual map r. Specifically, the first term is to ensure that the sum of $F_u x$ in k-space and r is close to the measurement data y. With the assumption that the Γ can accurately transform the k-space data into image data, the second term aims to push the current reconstructed image x to the reference image. These two terms can further improve the constrain capacity as opposed to the one data fidelity in Eq. (1). The third term aims to minimize the residual map r and the last term represents the sparsity-based regularization term to constrain the reconstructed image x in the transform domain. Since the optimization problem Eq. (7) has two variables, then we can use the alternative optimization strategy to divide it into the following sub-problems:

$$r^{k+1} = \underset{r}{\operatorname{argmin}} \frac{1}{2}\|x^k - \Gamma(F_u x^k + r)\|_2^2 + \frac{\lambda}{2}\|y - (F_u x^k + r)\|_2^2 + \frac{\mu}{2}\|r\|_2^2 \quad (8)$$

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \frac{1}{2}\|x - \Gamma(F_u x + r^{k+1})\|_2^2 + \frac{\lambda}{2}\|y - (F_u x + r^{k+1})\|_2^2 + \zeta\|Wx\|_1 \quad (9)$$

Regarding Eq. (8), the solution of r^{k+1} should be satisfied with the condition as follows:

$$\left(\frac{\partial\Gamma(F_u x^k + r)}{\partial(F_u x^k + r)}\right)^T (\Gamma(F_u x^k + r) - x^k) + \lambda(F_u x^k + r - y) + \mu r = 0 \quad (10)$$

According to [36], $\left(\frac{\partial\Gamma(F_u x^k + r)}{\partial(F_u x^k + r)}\right)^T$ can be approximated by F_u and $F_u\big(\Gamma(F_u x^k + r)\big)$ can be simplied with $F_u x^k + r$ (more interprelation of these two apprioximation can be referred to [36]). Therefore, Eq. (8) can be solved as:

$$r^{k+1} = \frac{\lambda(y - F_u x^k)}{1 + \lambda + \mu} \quad (11)$$

For Eq. (9), the solution of x^{k+1} can be given as (more details can be found in supplementary material):

$$x^{k+1} = W^* S_\delta(W(x^k + \frac{1+\mu}{\lambda}\Gamma(r^{k+1}))) \quad (12)$$

here S_δ is the soft-thresholding operation, W^* denotes the adjoint of W, and $\delta = \frac{\zeta}{\lambda}$.

There are two operators, Γ and $\mathrm{W}^* S_\delta W$ in Eq. (12), can be replaced by deep priors Φ_R and Φ_I, respectively. Meanwhile, the hyperparameters λ and μ are also optimized in different iterations. Hence, we have the TRANS-Net model as shown in Fig. 1(a):

$$x^{k+1} = \Phi_\mathrm{I}^{\mathrm{k}+1}(x^k + \frac{1+\mu^{k+1}}{\lambda^{k+1}}\Phi_\mathrm{R}^{\mathrm{k}+1}(F_u^H r^{k+1})) \quad (13)$$

In our work, the transform function Γ is composed of $\Phi_\mathrm{R} F_u^H$ for simplification.

B. K-space Error Map

The k-space error map r^k in Eq. (11) is the key element of TRANS-Net, which exactly estimates the undersampled k-space difference between the reconstructed image x^{k-1} and the measurement data y. Even though the reconstructed image x^{k-1} completes the task of artifact reduction and tissue restoration, r^k still remains the high-frequency information that x^{k-1} missed. Therefore, the reconstructed image x^{k-1} can be gradually improved following the direction pointed by r^k via the iterative process, which is superior to existing unrolled methods.

C. High-frequency Correction Neural Network (HC-Net)

To obtain the promoted reconstructed image x^k based on the previous intermediate result x^{k-1}, the k-space error map r^k needed to be transformed into the residual image first. In this paper, we adopt the inverse Fourier transform to get the residual image by $p^k = F_u^H r^k$. However, p^k cannot provide an accurate difference evaluation from the undersampled k-space data r^k.

So, unlike the classical unrolled method, TRANS-Net additionally conducts a deep neural network $\Phi_\mathrm{R}^\mathrm{k}$ to correct the residual image p^k for the precise evaluation of the difference between the reconstructed image x^k and the reference image. This step persistently underlines the high-frequency component, e.g., edges and fine structures, resulting in encouraging improvement than [3, 27].

Specifically, we apply the high-frequency correction neural network (HC-Net) to enhance the performance of TRANS-Net in detail preservation, as shown in Fig. 1(b). HC-Net is a lightweight U-shape network rather than a stack of successive convolutional layers [16, 37], which could offer a more powerful ability in feature extraction because of having more learnable parameters with the same computational resources. HC-Net has two sub-components: encoder and decoder. The target of the encoder is to compress the spatial information of feature maps and extract the essential features lying in the low-dimensional manifold, which has the robustness to noise and perturbation [38]. The decoder aims to correct the residual image p^k to get the accurate difference map v^k from the encoded feature space. Skip connection and global residual connection are both necessary components in U-shape networks [39, 40], which can accelerate information flow, avoid detail missing and speed up convergence [14, 41]. Also, BN [42] and ReLu [43] are employed to expedite the training process, prevent overfitting and improve the non-linear representation capacity. More architecture details of HC-Net can be found in Fig. 1(b).

D. Global-feature Enhancement Neural Network (GE-Net)

After obtaining the corrected residual image v^k via HC-Net, the updated intermediate result z^k can be achieved by the weighted sum of x^{k-1} and v^k. Again, another deep prior $\Phi_\mathrm{I}^\mathrm{k}$ is employed to regularize z^k to get the final reconstructed image x^k at the k^{th} iteration.

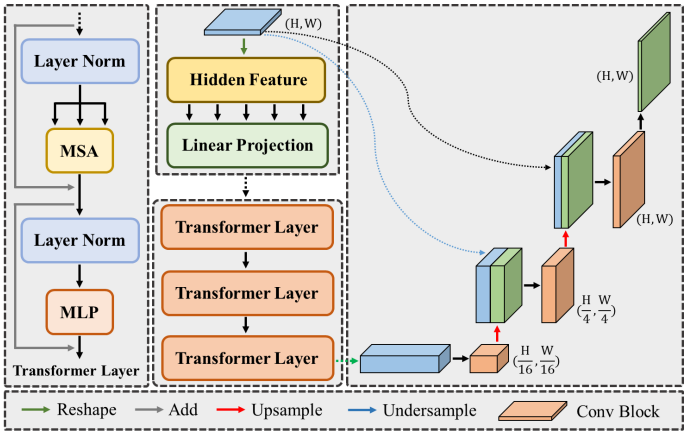

Fig. 2. Overview of transformer module in GE-Net.

Φ_I^k also uses a U-shape network smaller than HC-Net, termed global-feature enhancement neural network (GE-Net), as illustrated in Fig. 1(c). Meanwhile, to further enlarge the receptive field of GE-Net, the transformer technique is adopted to explore the spatial correlations within the global feature space [44], which has been applied to image processing and gets impressive results [45-48].

The workflow of the transformer module in GE-Net is depicted in Fig. 2 [48]. First, we need to create a 1D sequence by flattening non-overlapped patches from the input feature map $M_I \in R^{H\times W\times C}$ into the serialized signal $M_S \in R^{L\times N^2}$, where each flattened patch has N^2 elements and $L = (H \times W)/N^2$ is the number of patches. Then, a linear projection layer is used to embed M_S into the latent space and get the $M_L \in R^{L\times K}$. Subsequently, successive transformer layers are utilized to strengthen the exploitation of the global relationship contained in the feature map. It can be noticed that the transformer layer mainly consists of two sublayers, which are multi-head self-attention (MSA) and multi-layer perceptron (MLP), respectively [49]. Noticing that the output sequence of transformer layers M_T has the same dimension with M_L. Next, M_T is reshaped into $M_F \in R^{\frac{H}{N}\times\frac{W}{N}\times K}$, which is lied in the 2D feature map space. Last, several convolutional layers and upsampled operations are employed to gain the final result M_E. The detailed implementation of the transformer module can be found at https://github.com/lonelyatu/TRANS-Net.

E. Loss function

In this work, we use ℓ_1-minimization as the loss function to optimize the TRANS-Net, which is defined as follows:

$$L = \|f_\theta(y) - I_{FD}\|_1 \quad (14)$$

where f_θ is the TRANS-Net with learnable parameters θ, $f_\theta(y)$ in Fig.1 (a) presents the generated result at the final iteration block of the proposed method, I_{FD} indicates the reference image reconstructed from fully sampled k-space data. In the training stage, Eq. (14) is minimized to get the optimal parameters θ. In the test stage, all the parameters are fixed to get the reconstructed result directly.

IV. Experiments

A. Setup

1) Datasets

First, the MICCAI 2013 grand challenge dataset was performed to validate the performance of different reconstruction methods. It has more than twenty subjects that contain T_1-weighted and T_2-weighted images. We randomly selected 14978 2D brain images from 100 T_1-weighted MRI datasets for training and 200 2D brain images from another independent 50 datasets for testing. The corresponding imaging parameters are as follows: the image size is $256\times256\times N$ ($x\times y\times z$) and each voxel presents an area of $0.9\times0.9\times3$ mm^3 and TR/TE=7.9ms/4.5ms.

Second, to evaluate the performance of the proposed method on different anatomical tissues, we collected cine MR images from Second Data Science Bowl Cardiac Challenge, which has more than 1000 patients. And each sequence has 30 time periods. The image quality varies greatly between different patients, with differing image resolution, contrast, and aspect ratio. Specifically, 4830 2D cardiac images were randomly selected from 55 cases for training and another 200 2D cardiac images from 10 cases for testing.

All the images were cropped into the size of 256×256 and normalized to the range [0, 1].

2) Sampling Masks

There were three undersampling patterns performed in this work, including 2D Gaussian (G2D), 2D Radial (R2D) and 1D Gaussian (G1D). All the datasets were undersampled with the acceleration factors of 10, 5, 3.3, respectively. Fig.3 depicts some examples of undersampling masks.

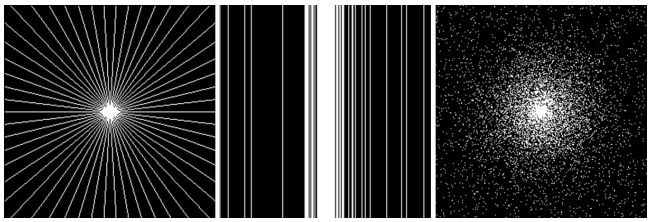

Fig. 3. Representative undersampling patterns.

3) Comparison Methods

In this paper, all the DL-based methods were conducted in the TensorFlow framework, and traditional methods were performed by MATLAB 2018a. All the experiments were implemented on a PC (CPU was Inter(R) Xeon(R) E5-2683, 2 GHz, GPU was NVIDIA GTX 1080Ti with 11G memory, RAM was 64GB) with Windows 10 operating system.

The proposed TRANS-Net was optimized using the Adam algorithm [50] with β_1=0.9 and β_2=0.999. The learning rate was initially set to 10^{-4} and slowly decreased to 10^{-6}. It took about 15 hours for 50 epochs to train the proposed method on the brain dataset and 6 hours on the cardiac dataset. The batch size was 8 and needed about 8GB of memory. And the iteration number was set to 10 for the trade-off between the performance and computational cost.

There are two hyperparameters λ and μ in the proposed method as shown in Fig. 1(a). Differing from fixing them in

traditional methods, hyperparameters of TRANS-Net in different iterations have different values, because they are optimized at the training stage to find the optimal values. At the training step, they were initially set to 1 and 0, respectively. All the training and testing details can be found at https://github.com/lonelyatu/TRANS-Net.

To assess the performance of TRANS-Net, some advanced MRI reconstruction methods, including zero filling (ZF), TGV[6], DLMRI [8], DAGAN [1], DCCNN [3], VN [27], and LDC [29] were treated as comparisons. Specifically, TGV and DLMRI present traditional iterative methods with hand-designed regularization terms. DAGAN is the image-domain-based method. DCCNN, VN and LDC stand for the unrolled methods.

Besides, we adopted the root mean square error (RMSE), peak single-to-noise ratio (PSNR) and structural similarity index (SSIM) [51] to evaluate the reconstruction results of various methods quantitatively. RMSE and PSNR aim to evaluate the pixel-wise distance between the reconstructed image and the reference image. Lower RMSE and higher PSNR mean better results. Meanwhile, SSIM measures the similarity of the feature contents between the reconstructed image and the ground truth. A larger SSIM value corresponds to a more competitive result. These three assessments are widely used for different tasks in image restoration [14, 30, 52, 53].

B. Brain Data Results

Table I lists the average quantitative evaluations of the 200 2D T_1-weighted brain images by various reconstruction methods with different undersampling patterns and acceleration factors. It can be seen that ZF inevitably leads to the worst assessment scores in all cases because of the zero-filling operation. Compared to ZF, TGV and DLMRI bring obvious improvement, which means the regularization constraint in the conventional CS-MRI is effective. Particularly, DLMRI achieves better results than DAGAN in some cases, which indicates the traditional iterative method is still competitive when the regularization term is chosen appropriately. Benefiting from the deep learning and rigorous CS theory, unrolled methods can consistently expose the relationship between undersampled k-space data and MR images, achieving better evaluations than the CNN-based method. Under the guidance of the learned data fidelity term, LDC can evaluate the distance from the reconstructed result to the reference image in a more accurate way and leads to better quantitative scores than DCCNN and VN. Further, our TRANS-Net gains the lowest RMSE, highest PSNR and SSIM scores that indicate the proposed method can lead to more accurate MR images in pixel value and anatomical structures, which confirms that the residual-domain-based regularization term and the transformer module can improve the reconstruction image.

Table I
QUANTITATIVE EVALUATIONS OF DIFFERENT METHODS ON THE BRAIN DATASET WITH DIFFERENT UNDERSAMPLING PATTERNS AND ACCELERATION FACTORS

Method	Mask	G2D(R=10)	G2D(R=5)	G2D(R=3.3)	R2D(R=10)	R2D(R=5)	R2D(R=3.3)	G1D(R=10)	G1D(R=5)	G1D(R=3.3)
ZF	RMSE	0.0446	0.0351	0.0256	0.0528	0.0301	0.0180	0.0639	0.0355	0.0311
	PSNR	27.10	29.17	31.90	25.64	30.52	35.00	24.00	29.09	30.24
	SSIM	0.4073	0.4548	0.5221	0.3849	0.5231	0.6480	0.5741	0.6888	0.7176
TGV	RMSE	0.0204	0.0101	0.0052	0.0255	0.0093	0.0047	0.0487	0.0294	0.0254
	PSNR	34.30	40.46	46.30	32.49	41.35	47.15	26.35	31.29	32.38
	SSIM	0.7738	0.9742	0.9860	0.8618	0.9784	0.9889	0.8010	0.9105	0.9396
DLMRI	RMSE	0.0102	0.0081	0.0045	0.0208	0.0083	0.0046	0.0467	0.0148	0.0148
	PSNR	39.92	41.87	47.58	33.81	42.25	47.30	26.72	36.70	36.69
	SSIM	0.9312	0.9765	0.9889	0.8897	0.9801	0.9892	0.8627	0.9736	0.9390
DAGAN	RMSE	0.0110	0.0064	0.0036	0.0218	0.0078	0.0042	0.0331	0.0153	0.0113
	PSNR	39.25	44.04	48.91	33.40	43.41	47.76	29.68	36.39	39.04
	SSIM	0.9698	0.9795	0.9897	0.9303	0.9815	0.9908	0.8938	0.9630	0.9691
DCCNN	RMSE	0.0093	0.0054	0.0032	0.0184	0.0071	0.0036	0.0328	0.0143	0.0103
	PSNR	40.77	45.59	50.09	34.83	44.20	48.99	29.75	36.98	39.83
	SSIM	0.9785	0.9846	0.9929	0.9389	0.9836	0.9925	0.9129	0.9645	0.9769
VN	RMSE	0.0056	0.0037	0.0023	0.0135	0.0055	0.0029	0.0277	0.0100	0.0082
	PSNR	45.21	48.74	52.75	37.60	45.44	50.97	31.24	40.08	41.79
	SSIM	0.9879	0.9913	0.9965	0.9462	0.9882	0.9969	0.9205	0.9726	0.9855
LDC	RMSE	0.0049	0.0030	0.0019	0.0119	0.0050	0.0026	0.0250	0.0089	0.0073
	PSNR	46.23	50.50	54.15	38.68	46.20	51.83	32.15	41.06	42.78
	SSIM	0.9910	0.9921	0.9974	0.9654	0.9916	0.9978	0.9157	0.9816	0.9892
TRANS-Net	RMSE	**0.0042**	**0.0027**	**0.0018**	**0.0114**	**0.0047**	**0.0024**	**0.0241**	**0.0080**	**0.0067**
	PSNR	**47.44**	**51.24**	**55.05**	**39.05**	**46.78**	**52.70**	**32.43**	**41.98**	**43.52**
	SSIM	**0.9918**	**0.9935**	**0.9985**	**0.9770**	**0.9948**	**0.9985**	**0.9383**	**0.9896**	**0.9920**

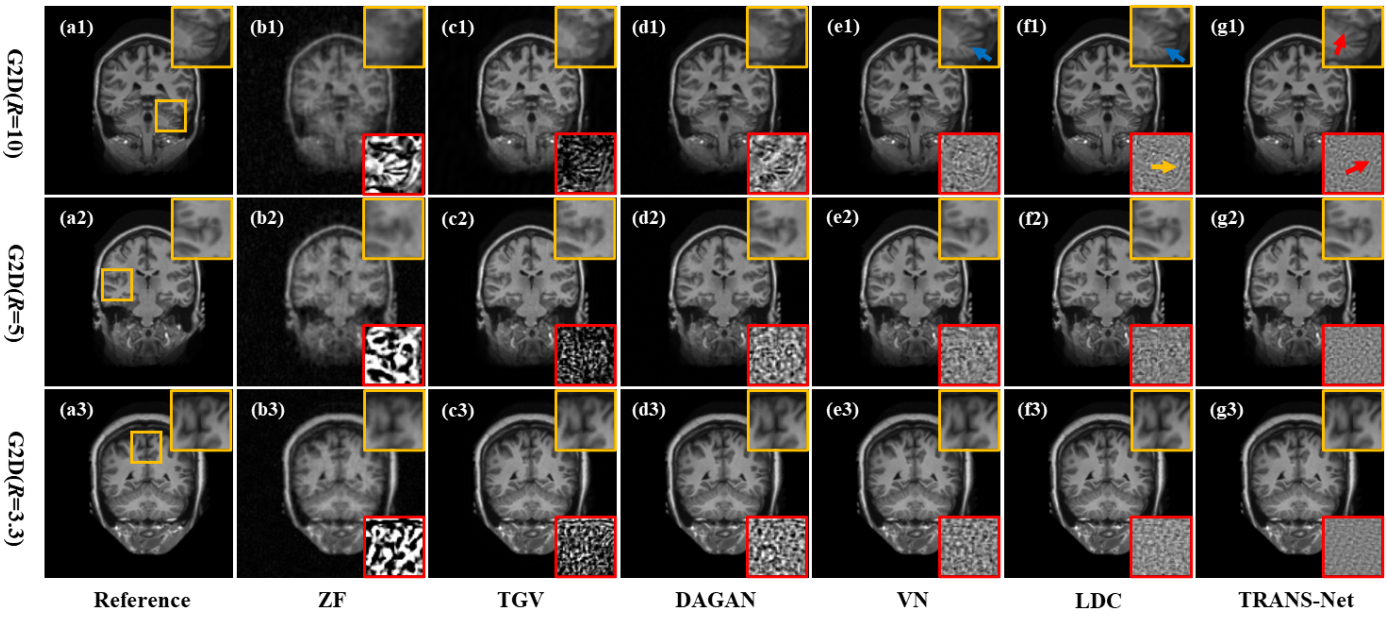

Fig. 4. Reconstruction results from the brain dataset for different methods on the G2D undersampling pattern with different acceleration factors, including reconstructed images, magnified ROIs and the corresponding difference maps. The display windows of the reconstructed images and ROIs are [0, 1], the display windows of the difference maps with the acceleration factors R=10, R=5, and R=3.3 are [-0.05, 0.05], [-0.025, 0.025], and [-0.015, 0.015], respectively.

To demonstrate the performance of different reconstruction methods in the visual aspect, experiments on the G2D undersampling pattern with acceleration factors R=10, R=5 and R=3.3 were performed, respectively. Fig. 4 gives the reconstruction results, regions-of-interest (ROIs), and the corresponding difference maps of different algorithms. It can be observed that ZF brings aliasing artifacts and blurring tissues in Fig. 4(b1)-(b3) due to the loss of k-space data. Noticing the sparse property of MR images in the gradient domain, the TV-based method can effectively remove the artifacts and restore tissue structures. However, TGV still misses some subtle details compared to reference images. This implies the limitation of CS-MRI that the hand-designed regularization term cannot accurately model the high-quality MR image space. By automatically learning the intrinsic feature from the MR images, DL-based methods can provide more promising results than CS-MRI, which is consistent with the observations in Table I. More specifically, DAGAN outperforms TGV by predicting the artifacts from the degraded images directly. Nevertheless, it fails to guarantee the consistency between the reconstruction image and the undersampled measurement data in k-space. Therefore, the data consistency layer is induced in the unrolled method. Consequently, VN and LDC can produce improved images with more detailed information (as pointed out by the blue arrow in Fig. 4(e1)(f1)). Further, LDC provides more tissue details than VN by introducing a more exact data fidelity term, as indicated by the yellow arrow in Fig. 4(f1). It is highly noticed that the proposed TRANS-Net can generate more clear edges (as pointed by red arrows in Fig. 4(g1)) and more precise images (as demonstrated by difference maps in Fig. 4(g2)-(g3)).

To further investigate the visual performance of the proposed method on different undersampling patterns, numerous experiments were conducted. Fig. 5 depicts the reconstruction images of different methods on the R2D and G1D undersampling patterns with the acceleration factor R=5. Again, ZF still results in severely degraded reconstruction images. Meanwhile, from the reconstruction results in Fig. 4(b1)-(b3) and Fig. 5(b1)-(b2), we can conclude that different undersampling patterns will induce various artifacts with diverse manifestations. It can be noticed in Fig. 5 (c1)-(c2) that DLMRI can provide improved perceptual images compared to ZF. DCCNN, VN, and LDC remain the superior performance in detail restoration (as pointed by the blue arrows in Fig. 5(d1)(e2)(f2)). This means the performance of unrolled methods is stable in terms of different undersampling patterns and acceleration factors. Last but not least, the proposed TRANS-Net has the best visual results in subtle structure preservation, as indicated by the red arrows in Fig. 5(g1)-(g2).

C. Cardiac Data Results

To further inspect the performance of the proposed method on different organ tissues, experiments on the cine cardiac dataset with different undersampling patterns and acceleration factors were performed.

Table II gives average assessments of 200 images for different methods. The same conclusion in Table I can also be observed in Table II that the proposed method has the best metric scores in all evaluation indictors. Fig. 6 demonstrates the reconstruction images from the cardiac datasets of different algorithms. It can be seen that the visual results of the CNN-based algorithm and unrolled methods gradually become better in turn, which can be convincingly verified by the difference maps in Fig. 6(c1)-(g3). Particularly since the residual-domain-based regularization constraint underlines the high-frequency components, and the transformer module enhances the exploitation of global features. Therefore, our TRANS-Net can generate the best accurate images (as observed by the difference maps in Fig. 6(g1)-(g3)) with minor

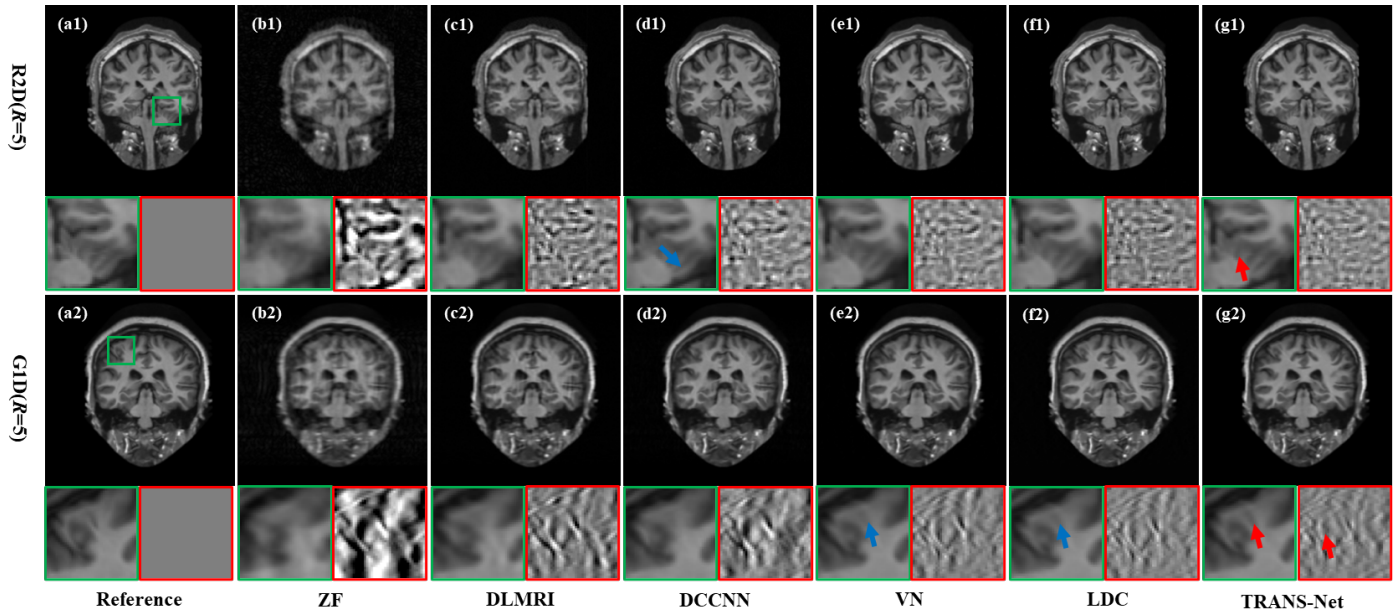

Fig. 5. Reconstruction results from the brain dataset for different methods on the R2D and G1D undersampling patterns with the acceleration factor R=5, including reconstructed images, magnified ROIs and the corresponding difference maps. The display windows of the reconstructed images and ROIs are [0, 1], the display windows of the difference maps are [-0.06, 0.06].

Table II
QUANTITATIVE EVALUATIONS OF DIFFERENT METHODS ON THE CARDIAC DATASET WITH DIFFERENT UNDERSAMPLING PATTERNS AND ACCELERATION FACTORS

Method	Mask	G2D(R=10)	G2D(R=5)	G2D(R=3.3)	R2D(R=10)	R2D(R=5)	R2D(R=3.3)	G1D(R=10)	G1D(R=5)	G1D(R=3.3)
ZF	RMSE	0.0427	0.0327	0.0258	0.0520	0.0318	0.0214	0.0660	0.0361	0.0322
	PSNR	27.42	29.76	31.82	25.75	30.01	33.46	23.63	28.88	29.87
	SSIM	0.7890	0.8318	0.8740	0.7080	0.8190	0.8971	0.7094	0.8453	0.8553
TGV	RMSE	0.0266	0.0191	0.0134	0.0465	0.0261	0.0163	0.0590	0.0266	0.0232
	PSNR	31.53	34.41	37.50	26.73	31.77	35.88	24.61	31.55	32.73
	SSIM	0.8720	0.9168	0.9505	0.7502	0.8753	0.9429	0.7156	0.8840	0.8898
DLMRI	RMSE	0.0182	0.0118	0.0078	0.0440	0.0190	0.0091	0.0574	0.0195	0.0167
	PSNR	34.87	38.67	42.53	27.23	34.59	41.04	24.86	34.30	35.60
	SSIM	0.9281	0.9599	0.9797	0.7751	0.9215	0.9744	0.7270	0.9261	0.9395
DAGAN	RMSE	0.0196	0.0135	0.0100	0.0415	0.0184	0.0099	0.0472	0.0198	0.0192
	PSNR	34.19	37.71	40.02	27.66	34.74	40.11	26.54	34.10	34.38
	SSIM	0.9129	0.9530	0.9683	0.7690	0.9183	0.9678	0.7437	0.9119	0.9145
DCCNN	RMSE	0.0143	0.0094	0.0073	0.0317	0.0144	0.0077	0.0465	0.0155	0.0136
	PSNR	36.93	40.56	42.81	30.01	36.90	42.29	26.66	36.23	37.37
	SSIM	0.9453	0.9703	0.9805	0.8391	0.9421	0.9776	0.7553	0.9438	0.9525
VN	RMSE	0.0122	0.0085	0.0060	0.0288	0.0126	0.0072	0.0427	0.0138	0.0120
	PSNR	38.33	41.45	44.53	30.84	38.07	42.92	27.40	37.25	37.46
	SSIM	0.9570	0.9738	0.9852	0.8580	0.9537	0.9800	0.7923	0.9534	0.9602
LDC	RMSE	0.0108	0.0078	0.0055	0.0264	0.0112	0.0066	0.0392	0.0120	0.0103
	PSNR	39.40	42.20	45.31	31.61	39.08	43.77	28.15	38.49	39.76
	SSIM	0.9646	0.9734	0.9873	0.8749	0.9605	0.9827	0.8123	0.9619	0.9682
TRANS-Net	RMSE	**0.0100**	**0.0075**	**0.0049**	**0.0250**	**0.0107**	**0.0063**	**0.0367**	**0.0115**	**0.0102**
	PSNR	**40.02**	**42.57**	**46.33**	**32.09**	**39.45**	**44.14**	**28.73**	**38.95**	**39.89**
	SSIM	**0.9682**	**0.9790**	**0.9894**	**0.8850**	**0.9620**	**0.9837**	**0.8313**	**0.9635**	**0.9689**

details and clear edges (as marked by the red arrows and red circles). Both brain and cardiac datasets validate the outstanding performance of the proposed method qualitatively and quantitively.

D. Analysis for Iteration Number

The iteration number in the unrolled reconstruction method is an important component. Hence, experiments from the brain dataset on the G2D undersampling pattern with the acceleration

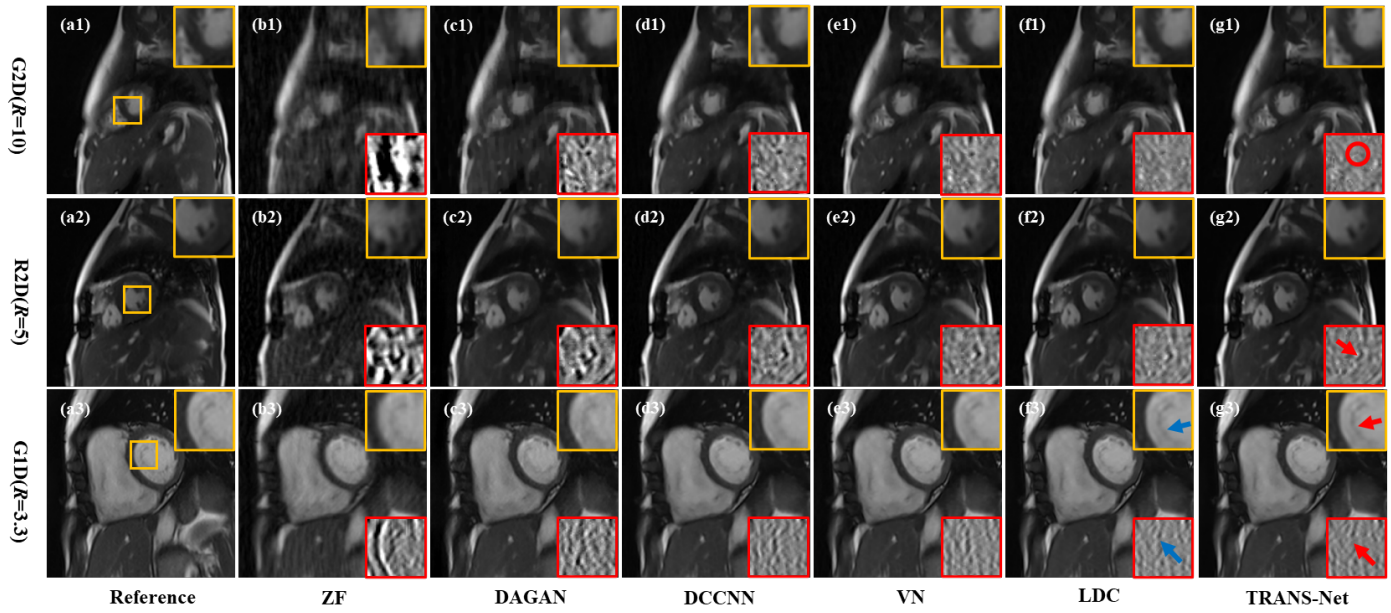

Fig. 6. Reconstruction results from the cardiac dataset for different methods on various undersampling patterns with different acceleration factors, including reconstructed images, magnified ROIs and the corresponding difference maps. The display windows of the reconstructed images and ROIs are [0, 1], the display windows of the difference maps with the acceleration factors R=10, R=5, and R=3.3 are [-0.06, 0.06], [-0.04, 0.04], and [-0.07, 0.07], respectively.

factor R=10 were performed to explore the influence of iteration number on the final results. Fig. 7 illustrates how the iteration numbers affect the performance of the proposed method that the more iterations, the better the performance. Meanwhile, TRANS-Net can achieve competitive results with only a few iterations, suggesting the effectiveness of the residual-domain-based regularization term and transformer module. However, the larger iteration number will greatly increase the computational cost but gain little improvement. Hence, to balance the trade-off between performance and hardware consumption, the iteration number in this work is set to 10.

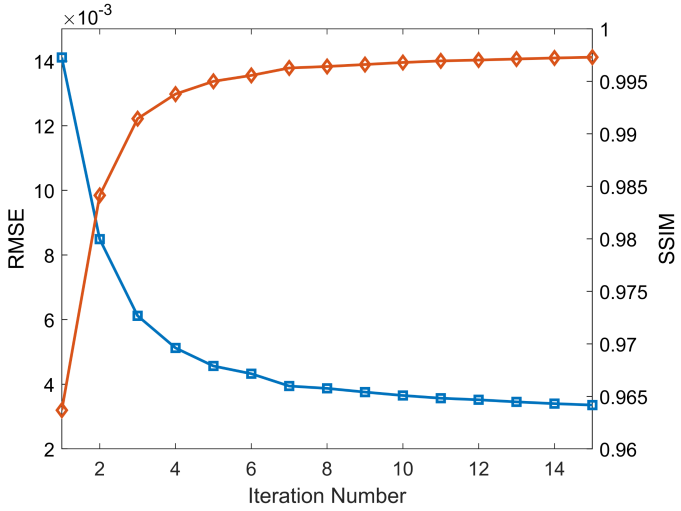

Fig. 7. Quantitative evaluations for different iteration numbers of TRANS-Net from the brain dataset on the G2D undersampling pattern with the acceleration factor R=10.

E. Ablation Study

In this section, an ablation study was performed to expose the effects of different modules in the proposed TRANS-Net on the reconstruction performance. Experiments were based on the brain dataset with G2D undersampling pattern and acceleration factors R=10, R=5, R=3.3.

The proposed method has two deep priors, i.e., residual-domain-based regularization term (RRT) and image-domain-based regularization term (IRT). First, only the RRT was adopted as the baseline model. Then, only the IRT was introduced to establish the first comparison model. Next, IRT, but without the transformer module, was integrated into the baseline model to build the second comparison model. Last, the transformer module was added to the second model to construct the third comparison model, i.e., TRANS-Net.

Table III
QUANTITATIVE EVALUATIONS FOR THE ABLATION STUDY

Method	Mask	G2D(R=10)	G2D(R=5)	G2D(R=3.3)
+RRT	RMSE	0.0095	0.0067	0.0046
	PSNR	40.44	43.43	46.72
	SSIM	0.9113	0.9553	0.8740
+IRT	RMSE	0.0059	0.0037	0.0026
	PSNR	44.99	48.78	51.78
	SSIM	0.9823	0.9925	0.9935
+RRT+IRT	RMSE	0.0047	0.0031	0.0021
	PSNR	46.07	50.07	53.84
	SSIM	0.9884	0.9930	0.9975
TRANS-Net	RMSE	**0.0042**	**0.0027**	**0.0018**
	PSNR	**47.44**	**51.24**	**55.05**
	SSIM	**0.9918**	**0.9933**	**0.9985**

Table III presents the quantitative evaluations for the ablation study. Compared to Table I, RRT can boost the reconstruction images to some degree. Certainly, IRT has better performance in all evaluations over RRT because the IRT directly regularizes final images rather than only constraints the intermediate results like RRT. Further, incorporating the RRT

and IRT, the second comparison model outperforms all the competitive algorithms in Table I, which strongly claims the effectiveness of RRT. It is worth noting that TRANS-Net gets impressive results in all cases, affirming that the transformer module positively promotes the global feature extraction for IRT.

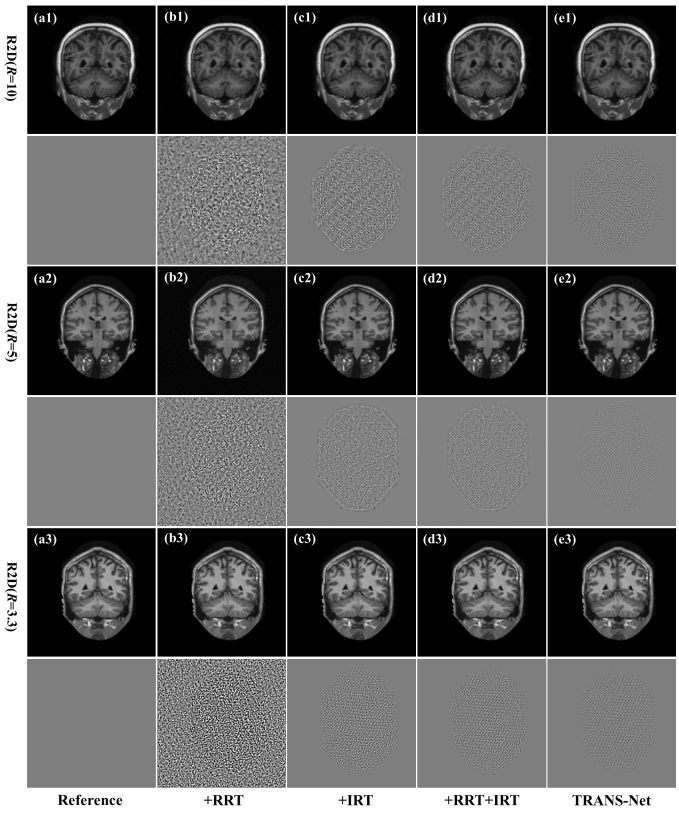

Fig. 8. Reconstruction results from the brain dataset for different modules on the G2D undersampling pattern with different acceleration factors, including reconstructed images and the corresponding difference maps. The display windows of the reconstructed images are [0, 1], the display windows of the difference maps with the acceleration factors R=10, R=5, and R=3.3 are [-0.05, 0.05], [-0.03, 0.03], and [-0.01, 0.01], respectively.

Besides, to visually probe the effects of different modules, the reconstruction results of various comparison models were evaluated. It can be seen that only a single RRT can reduce the artifacts and restore most of the tissues. Moreover, from the difference maps in Fig. 8(b1)-(b3), we can observe that most of the high-frequency information is also preserved, which confirms the effectiveness of RRT. Meanwhile, only one IRT, which can be considered as the classical unrolled method, has better accuracy than RRT (as indicated by the difference maps in Fig. 8(c1)-(c3)). This demonstrates the drawback of RRT that it cannot guarantee the quality of final images. Consequentially, the combination of RRT and IRT gets the best results among the first three models, that it only losses some subtle edges. Last, to further explore the global features, the transformer module is integrated into the second comparison model to produce the TRANS-Net, resulting in greatly promising images with very few errors compared to the reference images (as noticed by the difference maps in Fig. 8(e1)-(e3)).

F. Generalization of TRANS-Net

In order to inspect the generalization of the proposed method, we conducted experiments to explore the performance of TRANS-Net on domain variance, undersampling pattern variance and acceleration factor variance, respectively.

1) Domain Variance

This part mainly explores the performance of the proposed method against domain variance. Three models were trained on the G1D undersampling pattern with acceleration factors R=5. Meanwhile, four different test schemes were performed. The training and testing information is presented in Table IV.

Table IV
ABBREVIATED OF DIFFERENT TRAINING AND TESTING SCHEMES

Scheme	**AB**	**BA**	**AC**	**BC**
Test dataset	Brain	Cardiac	Brain	Cardiac
Train dataset	Cardiac	Brain	Brain&Cardiac	Brain&Cardiac

Table V lists the quantitative evaluations of four different test schemes. Comparing Tables I and II, we can notice that the performance of TRANS-Net will decrease severely when the training data and testing data are different. However, the proposed method performs better when the TRANS-Net is trained on all the datasets.

Table V
QUANTITATIVE EVALUATIONS OF DIFFERENT TEST SCHEMES

Scheme	**AB**	**BA**	**AC**	**BC**
RMSE	0.0157	0.0857	**0.0107**	**0.0304**
PSNR	36.26	21.37	**39.48**	**30.37**
SSIM	0.9367	0.4114	**0.9770**	**0.9016**

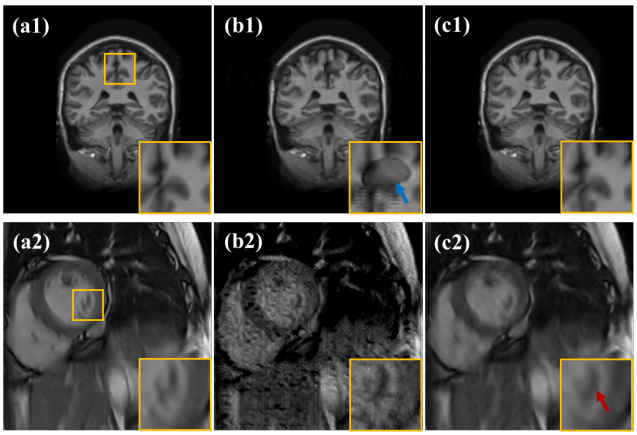

Fig. 9. Reconstruction results of different test schemes. (a1)-(a2) Reference result, (b1) **AB** result, (c1) **AC** result, (b2) **BA** result, (c2) **BC** result. All the display winodws are [0,1].

Fig. 9 shows the reconstruction results of different test schemes. From Fig. 9(b1)(b2), it can be seen that TRANS-Net introduces fake structures (as pointed by the blue arrow in Fig. 9(b1) and artifacts. Again, the proposed method could restore most tissues if the training datasets are diverse.

2) Undersampling Pattern Variance

This section focuses on the effects of undersampling pattern variance on the proposed method. Three models were trained on the cardiac dataset with acceleration factors R=5. Also, four

different test schemes were performed. The training and testing information is presented in Table VI.

Table VI
ABBREVIATED OF DIFFERENT TRAINING AND TESTING SCHEMES

Scheme	**AB**	**BA**	**AC**	**BC**
Test pattern	G1D	G2D	G1D	G2D
Train pattern	G2D	G1D	G1D&G2D	G1D&G2D

It can be noted that the evaluations of **AB** are similar to the results in Table II, which means that TRANS-Net is more robust to undersampling pattern variance than domain variance. Again, when the training datasets increase, the proposed method performs better.

Table VII
QUANTITATIVE EVALUATIONS OF DIFFERENT TEST SCHEMES

Scheme	**AB**	**BA**	**AC**	**BC**
RMSE	0.0134	0.0363	**0.0121**	**0.0080**
PSNR	37.52	29.07	**38.36**	**42.00**
SSIM	0.9564	0.8301	**0.9610**	**0.9768**

3) Acceleration Factor Variance

This content aims to investigate how the acceleration factor variance affects the proposed method. Three models were trained on the cardiac dataset with G2D undersampling pattern. Also, four different test schemes were performed. The training and testing information is presented in Table VIII.

Table VIII
ABBREVIATED OF DIFFERENT TRAINING AND TESTING SCHEMES

Scheme	**AB**	**BA**	**AC**	**BC**
Test factor	R=10	R=5	R=10	R=5
Train factor	R=5	R=10	R=5&R=10	R=5&R=10

From Table IX, we can observe that the proposed method has the most generalization ability with regard to acceleration factor variance compared to Table V and Table VII. The reason may be that the same undersampling pattern presents similar artifacts even though with different acceleration factors.

Table IX
QUANTITATIVE EVALUATIONS OF DIFFERENT TEST SCHEMES

Scheme	**AB**	**BA**	**AC**	**BC**
RMSE	0.0125	0.0078	**0.108**	**0.0076**
PSNR	38.08	42.29	**39.47**	**42.47**
SSIM	0.9620	0.9781	**0.9658**	**0.9785**

To sum up, one trained TRANS-Net model can reconstruct multiply MRI data with appropriate performance when the training datasets cover all the imaging conditions.

G. Removal of Noise and Motion Artifact

1) Noise Removal

To further probe the performance of the proposed method in noise removal, experiments from the brain dataset on the G1D undersampling pattern with the acceleration factor R=5 were performed. Specifically, the training and test images were generated by additionally adding Gaussian noise in the frequency domain with three different standard deviations, which were 10, 20, and 30 respectively.

Table X provides the evaluation of the proposed method with different noise levels. It can be observed that TRANS-Net performs decreases little when the noise level increases, which indicates the proposed method is robust to the noise.

Table X
QUANTITATIVE EVALUATIONS OF DIFFERENT NOISE LEVELS

Method	Noise Level	10	20	30
ZF	RMSE	0.0241	0.0560	0.0694
	PSNR	32.45	25.05	23.18
	SSIM	0.7401	0.4569	0.3627
TRANS-Net	RMSE	**0.011**	**0.0124**	**0.0152**
	PSNR	**39.02**	**38.14**	**36.38**
	SSIM	**0.9612**	**0.9585**	**0.9520**

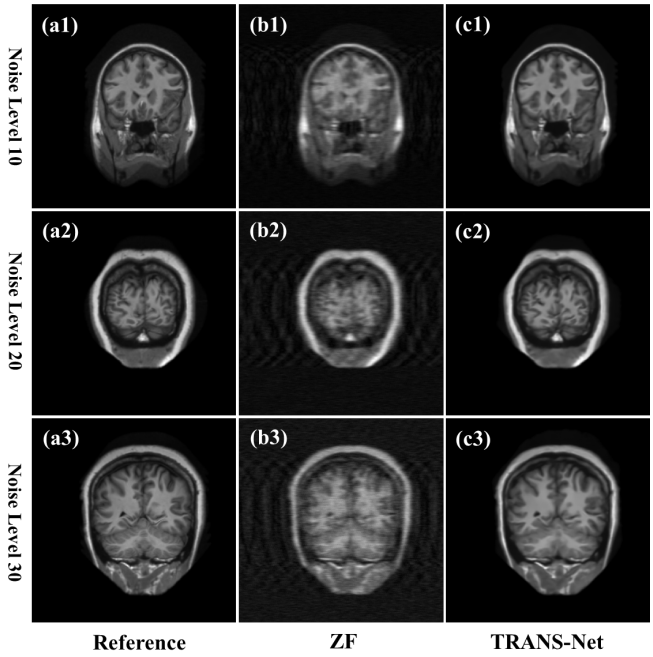

Fig. 10. Reconstruction results from the brain dataset for different noise levels on the G1D undersampling pattern with acceleration factor R=5. The display windows of the reconstructed images are [0, 1].

Fig. 10 illustrates the visual results of the proposed method in noise removal. Compared to ZF, TRANS-Net could achieve the balance between noise-artifact removal and tissue detail preservation.

2) Motion Artifact Removal

This section pays attention to the motion artifact removal of the proposed method. Experiments from the cardiac dataset on the G2D undersampling pattern with the acceleration factor R=5 were performed. The training and test images with motion artifacts were generated according to [54].

Fig. 11 shows the qualitative results of the proposed method in motion artifact removal. Again, TRANS-Net can provide competitive results in removing motion blurring and retaining tissue edges.

H. Computational Cost And Memory Usage

Table XI presents the average computational cost and GPU memory usage of different methods for processing a single brain image. TGV and DLMRI consume much longer time than

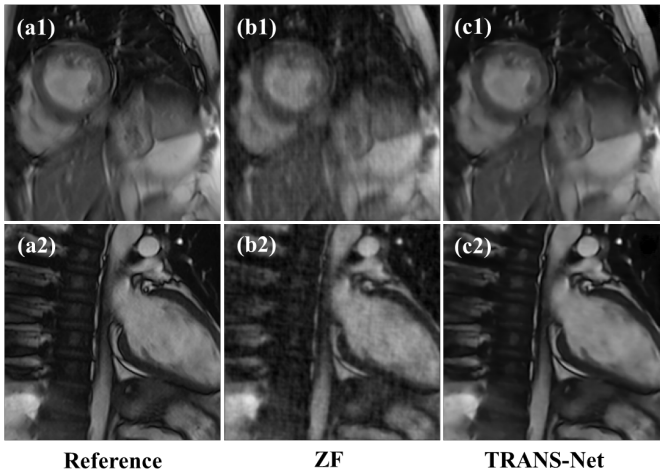

Fig. 11. Reconstruction results from the cardiac dataset for motion artifact removal on the G2D undersampling pattern with acceleration factor R=5. The display windows of the reconstructed images are [0, 1].

DL-based methods because they involve sophisticated optimization of regularization terms. All the DL-based methods can reconstruct the MRI result within a very fast time. Because of the lightweight U-shape network, the proposed method requires the smallest memory among all DL-based methods, which is more feasible to be deployed on various devices.

Table XI
COMPUTATIONAL COST (UNIT: SECOND) AND GPU MEMORY USAGE (UNIT: GB) OF DIFFERENT METHODS

Method	TGV	DL MRI	DA GAN	DC CNN	VN	LDC	TRANS -Net
Time	67	354	0.02	0.04	0.11	0.03	0.08
Memory	/	/	11.8	3.6	11.6	2.3	1.4

V. CONCLUSION

The unrolled methods have been widely applied in image restoration [3, 15, 31, 35, 37] due to the superior performance profited from the combination of CS theory and deep neural networks. They can overcome the defects of CS-MRI in the hyper-parameter selection, regularization term design and improve the interpretability for deep priors. However, one popular scheme of those methods usually conducts the regularization term on the reconstruction results in the image domain but ignores the importance of the error images in the residual space, which may limit their further improvement. Inspired by [55], we developed a transformer-enhanced residual-error alternative suppression network for MR imaging. Differing from the existing unrolled methods, the TRANS-Net additionally employs a regularization term on the error images reconstructed from the undersampled k-space to obtain accurate difference maps between the intermediate result and the reference image, which significantly emphasizes the high-frequency information from the reconstructed image and accelerates the convergences for the proposed method. Then, another traditional deep regularization prior deployed on the image domain aims to refine the final image. Nevertheless, this deep prior on the image domain only pays attention to the local features. Therefore, we adopt the transformer module to enlarge the receptive field and explore the latent relationship of the global image. Experiments based on the brain and cine cardiac datasets with different undersampling patterns and acceleration factors were performed to validate the performance of the proposed TRANS-Net method. Compared to other reconstruction algorithms, TRANS-Net shows remarkable performance in tissue restoration and subtle detail preservation.

Although the TRANS-Net framework gives impressive results for fast MR imaging, it still has some issues that should be noticed in the future. First, in this paper, we did not consider the k-space based processing methods like [22, 23], which are also commonly used for MR imaging. Therefore, how to effectively combine the TRANS-Net and k-space based methods is still an interesting problem. Second, although the transformer module brings promising improvement, it is a preliminary attempt for unrolled MR imaging. Hence, a novel transformer architecture for unrolled methods should be developed with small learnable parameters and few computational resources.

REFERENCES

[1] G. Yang, S. Yu, H. Dong *et al.*, "DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310-1321, 2018.

[2] K. G. Hollingsworth, "Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction," *Physics in Medicine Biology*, vol. 60, no. 21, pp. R297, 2015.

[3] J. Schlemper, J. Caballero, J. V. Hajnal *et al.*, "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491-503, 2018.

[4] W. Wu, D. Hu, K. An *et al.*, "A High-Quality Photon-Counting CT Technique Based on Weight Adaptive Total-Variation and Image-Spectral Tensor Factorization for Small Animals Imaging," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-14, 2021.

[5] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086-1098, 2007.

[6] F. Knoll, K. Bredies, T. Pock *et al.*, "Second order total generalized variation (TGV) for MRI," *Magnetic Resonance in Medicine*, vol. 65, no. 2, pp. 480-491, 2011.

[7] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in International conference on curves and surfaces, 2010, pp. 711-730.

[8] S. Ravishankar, and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028-1041, 2010.

[9] Y. Huang, J. Paisley, Q. Lin *et al.*, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5007-5019, 2014.

[10] S. G. Lingala, Y. Hu, E. DiBella *et al.*, "Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: k-t SLR," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1042-1054, 2011.

[11] R. Otazo, E. Candes, and D. K. Sodickson, "Low‐rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125-1136, 2015.

[12] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull *et al.*, "MRI denoising using non-local means," *Medical Image Analysis*, vol. 12, no. 4, pp. 514-523, 2008.

[13] S. Wang, H. Yu, Y. Xi *et al.*, "Spectral-Image Decomposition With Energy-Fusion Sensing for Spectral CT Reconstruction," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021.

[14] K. Zhang, W. Zuo, Y. Chen *et al.*, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, 2017.

[15] J. Zhang, and B. Ghanem, "ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing," in CVPR, 2018, pp. 1828-1837.

[16] S. A. H. Hosseini, B. Yaman, S. Moeller *et al.*, "Dense Recurrent Neural Networks for Accelerated MRI: History-Cognizant Unrolling of Optimization Algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1280-1291, 2020.

[17] Z. Huang, J. Zhang, Y. Zhang *et al.*, "DU-GAN: Generative Adversarial Networks With Dual-Domain U-Net-Based Discriminators for Low-Dose CT Denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-12, 2022.

[18] S. Wang, Z. Su, L. Ying *et al.*, "Accelerating magnetic resonance imaging via deep learning," in ISBI, 2016, pp. 514-517.

[19] D. Lee, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing MRI," in ISBI, 2017, pp. 15-18.

[20] T. M. Quan, T. Nguyen-Duc, and W. K. Jeong, "Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1488-1497, 2018.

[21] F. Liu, A. Samsonov, L. Chen *et al.*, "SANTIS: sampling‐augmented neural network with incoherent structure for MR image reconstruction," *Magnetic Resonance in Medicine*, vol. 82, no. 5, pp. 1890-1904, 2019.

[22] T. Eo, Y. Jun, T. Kim *et al.*, "KIKI‐net: cross‐domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magnetic Resonance in Medicine*, vol. 80, no. 5, pp. 2188-2201, 2018.

[23] H. El‐Rewaidy, A. S. Fahmy, F. Pashakhanloo *et al.*, "Multi-domain convolutional neural network (MD‐CNN) for radial reconstruction of dynamic cardiac MRI," *Magnetic Resonance in Medicine*, vol. 85, no. 3, pp. 1195-1208, 2021.

[24] Q. Ma, J. Jiang, X. Liu *et al.*, "Deep Unfolding Network for Spatiospectral Image Super-Resolution," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 28-40, 2022.

[25] H. Zhang, B. Liu, H. Yu *et al.*, "MetaInv-Net: Meta Inversion Network for Sparse View CT Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 621-634, 2021.

[26] W. Xia, Z. Lu, Y. Huang *et al.*, "MAGIC: Manifold and Graph Integrative Convolutional Network for Low-Dose CT Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3459-3472, 2021.

[27] K. Hammernik, T. Klatzer, E. Kobler *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055-3071, 2018.

[28] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394-405, 2019.

[29] J. Cheng, Z. Cui, W. Huang *et al.*, "Learning data consistency and its application to dynamic MR imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3140-3153, 2021.

[30] Z. Ke, W. Huang, Z. X. Cui *et al.*, "Learned Low-rank Priors in Dynamic MR Imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3698-3710, 2021.

[31] D. Liang, J. Cheng, Z. Ke *et al.*, "Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 141-151, 2020.

[32] F. Knoll, K. Hammernik, C. Zhang *et al.*, "Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 128-140, 2020.

[33] Y. Hu, Y. Xu, Q. Tian *et al.*, "RUN‐UP: Accelerated multishot diffusion‐weighted MRI reconstruction using an unrolled network with U‐Net as priors," *Magnetic Resonance in Medicine*, vol. 85, no. 2, pp. 709-720, 2021.

[34] V. Vishnevskiy, J. Walheim, and S. Kozerke, "Deep variational network for rapid 4D flow MRI reconstruction," *Nature Machine Intelligence*, vol. 2, no. 4, pp. 228-235, 2020/04/01, 2020.

[35] J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in NIPS, 2016.

[36] W. Wu, D. Hu, W. Cong *et al.*, "Stabilizing deep tomographic reconstruction: Part B. Convergence analysis and adversarial attacks," *Patterns*, pp. 100475, 2022/04/06/, 2022.

[37] H. Chen, Y. Zhang, Y. Chen *et al.*, "LEARN: Learned Experts' Assessment-Based Reconstruction Network for Sparse-Data CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1333-1347, 2018.

[38] B. Zhu, J. Z. Liu, S. F. Cauley *et al.*, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487-492, 2018/03/01, 2018.

[39] Y. Han, and J. C. Ye, "Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418-1429, 2018.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015, pp. 234-241.

[41] D. Hu, J. Liu, T. Lv *et al.*, "Hybrid-Domain Neural Network Processing for Sparse-View CT Reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 88-98, 2021.

[42] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in ICML, 2015, pp. 448-456.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in NIPS, 2012, pp. 1097-1105.

[44] A. Dosovitskiy, L. Beyer, A. Kolesnikov *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations, 2020.

[45] H. Chen, Y. Wang, T. Guo *et al.*, "Pre-Trained Image Processing Transformer," *arXiv e-prints*, pp. 2012.00364, 2020.

[46] Z. Zhang, L. Yu, X. Liang *et al.*, "TransCT: Dual-path Transformer for Low Dose Computed Tomography," *arXiv e-prints*, pp. 2103.00634, 2021.

[47] A. Hatamizadeh, D. Yang, H. Roth *et al.*, "UNETR: Transformers for 3D Medical Image Segmentation," *arXiv e-prints*, pp. 2103.10504, 2021.

[48] J. Chen, Y. Lu, Q. Yu *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint*, vol. arXiv:.04306, 2021.

[49] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," in NIPS, 2017, pp. 5998-6008.

[50] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, pp. 1412.6980, December 01, 2014, 2014.

[51] W. Zhou, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[52] M. Claus, and J. v. Gemert, "ViDeNN: Deep Blind Video Denoising," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1843-1852.

[53] H. Chen, Y. Zhang, M. K. Kalra *et al.*, "Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524-2535, 2017.

[54] Q. Lyu, H. Shan, Y. Xie *et al.*, "Cine Cardiac MRI Motion Artifact Reduction Using a Recurrent Neural Network," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 2170-2181, 2021.

[55] W. Wu, D. Hu, W. Cong *et al.*, "Stabilizing deep tomographic reconstruction: Part A. Hybrid framework and experimental results," *Patterns*, pp. 100474, 2022/04/06/, 2022.