IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 1, 2021, accepted May 31, 2021, date of publication June 7, 2021, date of current version June 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3086839

A Modified Generative Adversarial Network Using Spatial and Channel-Wise Attention for CS-MRI Reconstruction

GUANGYUAN LI[1], JUN LV[1], AND CHENGYAN WANG[2]
[1]School of Computer and Control Engineering, Yantai University, Yantai 264005, China
[2]Human Phenome Institute, Fudan University, Shanghai 200433, China

Corresponding author: Jun Lv (ljdream0710@pku.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61902338.

ABSTRACT Compressed sensing (CS) can speed up the magnetic resonance imaging (MRI) process and reconstruct high-quality images from under-sampled *k*-space data. However, traditional CS-MRI suffers from slow iterations and artifacts caused by noise when the acceleration factor is high. Currently, deep learning has been introduced to address these issues. Although some improvements have been achieved, the reconstruction problem under high under-sampling rates has not been solved. Thus, our study proposed a novel CS-MRI reconstruction method called RSCA-GAN. The generator of RSCA-GAN is a residual U-net consisting of both spatial and channel-wise attention. The proposed RSCA-GAN was compared to the zero-filling, DAGAN, RefineGAN, and RCA-GAN using both cartesian and non-cartesian under-sampling masks on brain and knee datasets. The sampling rates of cartesian masks are 25%, 16.7%, and 12.5% and the sampling rates of spiral and radial masks are 30%, 20%, and 10%. Peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and normalized mean square error (NMSE) were used to evaluate the reconstructed image quality, and the rank-sum test was adopted to evaluate the significant difference among different approaches. $P < 0.05$ indicated statistical significance. The results demonstrated that RSCA-GAN outperforms the other approaches for all the quantitative metrics. Thus, RSCA-GAN exhibits excellent reconstruction performance at high under-sampling rates and is suitable for clinical applications.

INDEX TERMS Compressed sensing, magnetic resonance imaging, generative adversarial network, spatial and channel-wise attention, cyclic consistency loss, deep residual block.

I. INTRODUCTION

Magnetic resonance imaging (MRI) is a medical imaging technique with the characteristics of non-invasiveness, non-radiation, and high contrast. However, the acquisition speed of MRI data is limited by two factors: a) physical aspects, (gradient strength and magnetic field strength) and b) physiological aspects such as nerve stimulation [1]. The prolonged scan makes the patient feel uncomfortable, and will cause motion artifacts. Hence, accelerated MRI is clinically significant. The main reason for the long scanning time is that the data is collected line-by-line in *k*-space (frequency domain and Fourier space of the image) [2]. Although some hardware improvements have been made to reduce the MRI scan time such as echo-planar imaging sequence [3] and parallel imaging techniques [4], the acquisition time is still long [2]. Moreover, the theory of compressed sensing [5]–[10] has also been introduced to MRI reconstruction. However, CS-MRI has not been applied for clinical measurements [11]. The main reasons [12] are as follows: 1) CS-MRI under-sampling has the assumption of signal incoherence; 2) The transform domain is extremely shallow, and the image details related to the biological tissue could not be collected; 3) Non-linear optimization algorithms require a long time for iterative calculations; 4) Hyper-parameters need to be selected for over-regularization, making the reconstructed image blurry or causing obvious artifacts.

Recently, with the development of generative adversarial network (GAN) [13], several methods [14]–[16] are proposed to improve the image quality with modified

The associate editor coordinating the review of this manuscript and approving it for publication was Yunjie Yang.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

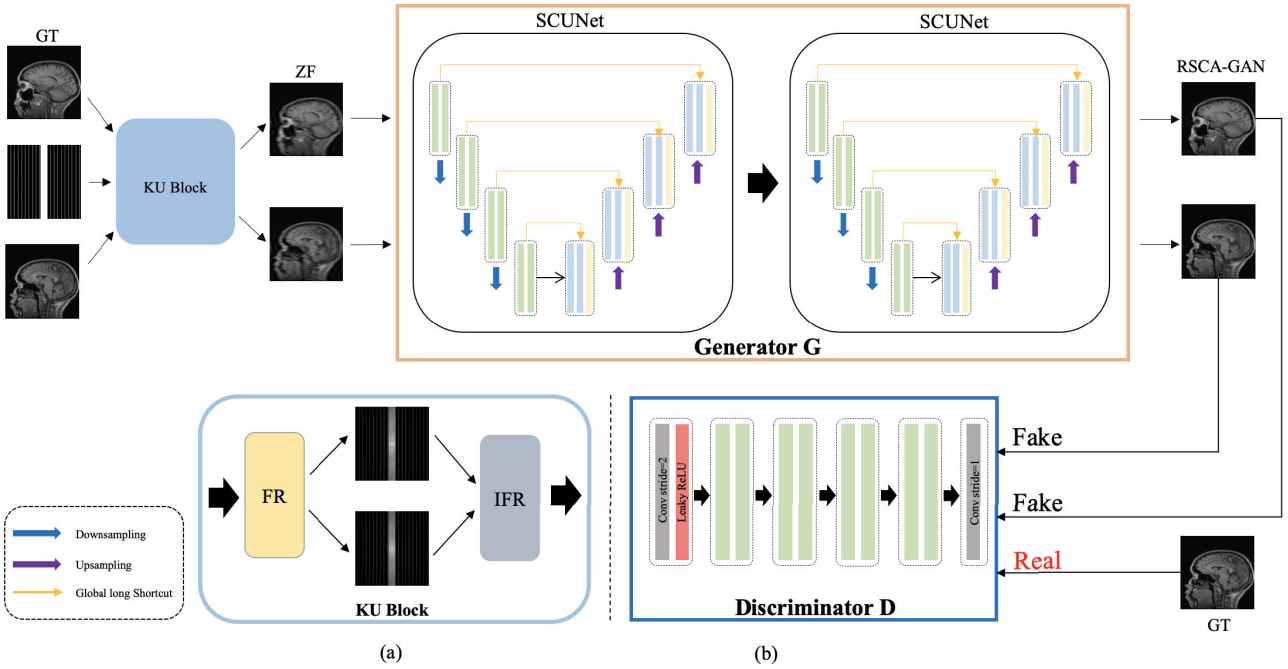

FIGURE 1. (a) Framework of KU Block. KU: k-space under-sampled. FR: Fourier transform block. IFR: inverse Fourier transform block. (b) Framework of the proposed method. The generator of the network is connected with two residual autoencoders, SCUNet. Each SCUNet is composed of spatial and channel-wise attention, and residuals were used to increase the depth of the network. The discriminator is composed of six layers that were used to determine the true or false of the image.

CS-MRI reconstruction. DCGAN [17] replaced the structure of the generator and discriminator with CNN to better extract the image features. WGAN [18] uses Wasserstein distance as the optimization method during training. Then, cycleGAN [19] is used to solve the problem under conditions that two images are not in pairs. Yang *et al.* [20] proposed a novel deep de-aliasing generative adversarial network (DAGAN) for fast CS-MRI reconstruction. Chen *et al.* [21] proposed a new model based on GAN (PTGAN) to convert T2-weighted MRI images into PD-weighted MRI images without disrupting the structure and texture of the image. RefineGAN [22] is a combination of cyclic consistency, U-net [23] and deep residual CNN [24]. Two residual U-nets were used as a generator to perform end-to-end MRI reconstruction, which deepens the network of the generator. Han *et al.* [25] proposed a two-step GAN for the generation and improvement of tumor brain magnetic resonance images. Moreover, many studies have applied attention mechanism to the image reconstruction field. Spatial and channel-wise attention based CNN (SCA-CNN) [26] was applied for natural image processing. The attention mechanism extracts the details of the image and assigns different weights to the channel or spatial feature maps. Huang *et al.* [27] used channel-wise attention in the up-sampling of U-net for MR image reconstruction.

Based on the above studies, we proposed a GAN based framework called RSCA-GAN, which combines residual U-net with spatial and channel-wise attention, to accelerate MRI reconstruction and eliminate the artifacts caused by highly under-sampling rates.

II. METHOD

A. K-SPACE UNDER-SAMPLING BLOCK

The overall framework of our method is illustrated in Figure 1. We designed a k-space under-sampled (KU) block as shown Figure 1(a), which input the GT image S_{GT} and under-sampling mask R of the module, and the output was zero-filling image S_{ZF}. There are two sub-modules in the KU block, namely Fourier transform R (FR) block and inverse Fourier transform R (IFR) block. The FR block transforms the GT image from the time domain to frequency domain by Fourier transform and uses mask R to obtain the under-sampled k-space samples. Subsequently, the under-sampled image in the frequency domain is transformed into the time domain by the inverse Fourier transform of the IFR block to obtain the zero-filling image S_{ZF}. The IFR block is represented as follows:

$$IF_R = F^* (F_R) = F^* (F (S_{GT}) \cdot R) \tag{1}$$

where F_R represents the formula of the FR block, F stands for Fourier transform, and F^* represents the inverse Fourier transform.

Figure 1(b) shows that the generator architecture of our proposed method is composed of two residual SCUNets with attention mechanism. Each coding block has two layers, and each decoding block has three layers. The first two layers are

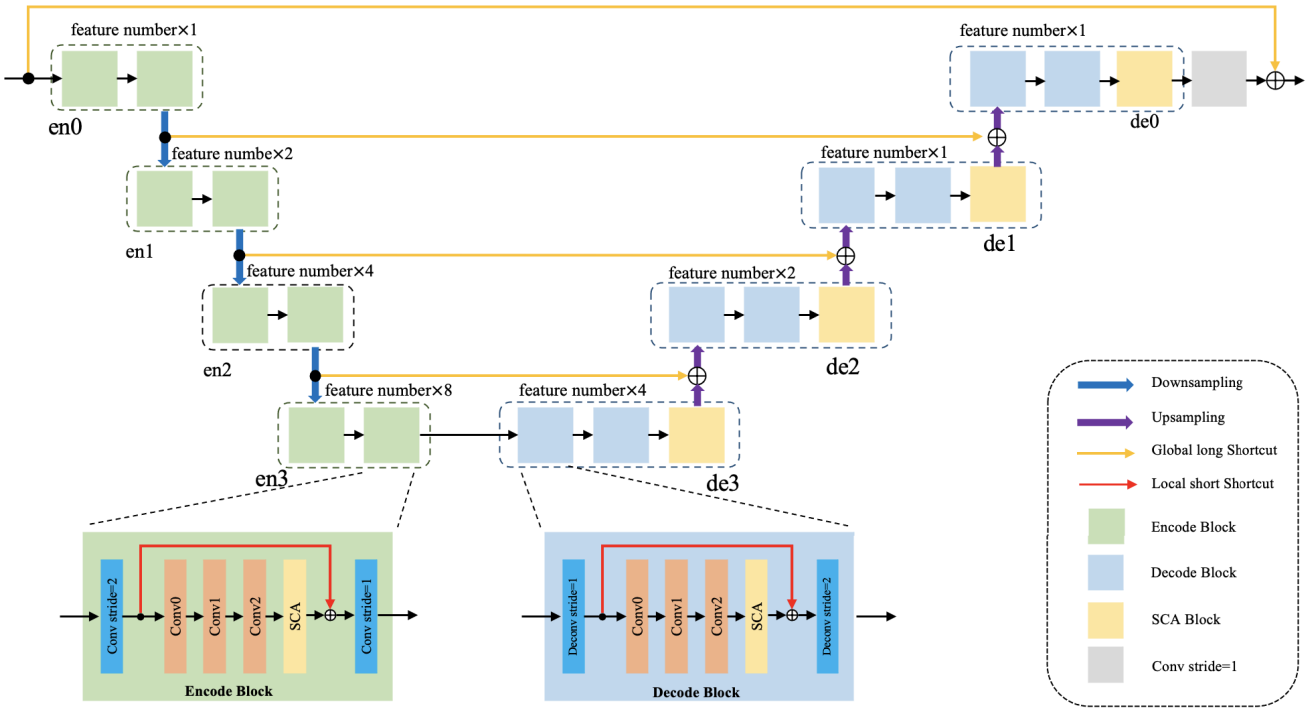

FIGURE 2. Architecture of residual SCUNet. The encode block is indicated by green, and the decode block is indicated by blue. The 4D tensor is used as input, using the 2D convolution with filter size of 3 × 3 and stride of 2. The number of feature maps is defined as feature number = 64. The residual block is represented by orange, which is used to increase the depth of the network. SCA block is indicated in yellow composed of spatial and channel attention.

composed of residual blocks with short-skip connections, and the third layer of the decoded block is the SCA block. The generator network is depicted in Figure 2.

B. GENERATIVE ADVERSARIAL NETWORK

The proposed method is based on GAN. The main structure of GAN is composed of a generator G and a discriminator D. The generator generates a fake image by inputting information in real image, which is used to fool the discriminator. The work of the discriminator is to judge the real/fake images generated by the generator. For the image generated by the generator, the result of the discriminator is *fake*. For the real image, the result of the discriminator is *real*. The generator is constantly learning to generate images infinitely close to the real image, and the discriminator is constantly optimized to judge the image generated by the generator. Thus, in the process of continuous generation and discrimination, the adversarial loss function is generated, and the network is continuously optimized by generating the adversarial loss of function, which is expressed as follows:

$$L_{adv} = \underset{\hat{y}\in P_G}{E}\left[1 - \log D\left(G\left(S_{ZF}\right)\right)\right] + \underset{y\in P_{\text{data}}}{E}\left[\log D(S_{GT})\right] \quad (2)$$

where P_G represents the data distribution learned by the generator, P_{data} represents the data distribution of MR image training set. $\hat{y} = G\left(S_{ZF}\right)$ represents the MR image generated by the generator. $E_y\left[\log D\left(S_{GT}\right)\right]$ represents the expected value of the discriminator for real image discrimination, and we expect that the discrimination result is closer to 1. $E_{\hat{y}}\left[1 - \log D\left(G\left(S_{ZF}\right)\right)\right]$ denotes the expected value of the discriminator for the generated image discrimination. We expect the discrimination result to be close to 0, such that the discriminator and the generator constitute antagonistic training.

C. CYCLIC DATA CONSISTENCY LOSS

Ideally, if there are sufficient datasets for training, any existing image S_{RE} reconstructed by the generator can be mapped to the zero-filling reconstruction S_{ZF} through the network. Therefore, under the condition of the limited training dataset, if we want to correctly map zero-filling data $S_{ZF}[m]$ and completely reconstruct $S_{RE}[m]$, m indicates mapping and strengthens the bridge between the datasets; an additional constraint was used: cyclic data consistency loss L_{cc}. It was composed of two losses and appears in a cyclic manner: the first term is the under-sampling frequency domain loss L_f, which is used to ensure that the difference between $r[i]$ and $r_{RE}[i]$ is the smallest. L_f can be expressed as:

$$L_f(G) = \text{MAE}\left(r[i], r_{RE}[i]\right) \quad (3)$$

where $r[i]$ is the acquired samples in the frequency domain, then $r[i]$ obtains $S_{ZF}[i]$ through the IFR block, and finally, the generator outputs the reconstructed image $S_{RE}[i]$. $r_{RE}[i]$ is the under-sampled reconstructed image obtained by $S_{RE}[i]$ through the FR block. MAE means mean-absolute error. The second term L_i is the MAE loss in image space, which

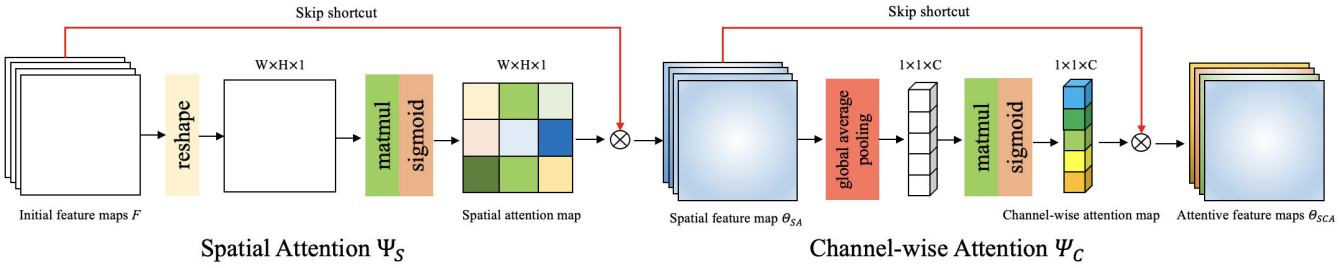

FIGURE 3. Architecture of spatial attention and channel-wise attention. The white feature maps represent the initial feature maps, and the colored feature maps represent the activated feature maps. The feature maps firstly activate the spatial attention and pass through the channel-wise attention module, finally obtaining fully activated feature maps.

is used to ensure that $S_{GT}[j]$ and $S_{RE}[j]$ are as similar as possible, L_i can be expressed as:

$$L_i(G) = \text{MAE}\left(S_{GT}[j], S_{RE}[j]\right) \tag{4}$$

where $S_{GT}[j]$ is the GT image, $S_{RE}[j]$ is the reconstructed image by the generator, and the input of the generator is $S_{ZF}[j]$, obtained by FR and IFR blocks. Therefore, L_{cc} can be expressed as:

$$L_{cc}(G) = L_f(G) + L_i(G) \tag{5}$$

According to the generative adversarial loss (2) and cyclic data consistency loss (5), the total loss function of the proposed method is expressed as:

$$L_{RSCA} = L_{adv}(G, D) + \lambda L_f(G) + \eta L_i(G) \tag{6}$$

where λ and η are super-parameters that stabilize make $L_f(G)$ and $L_i(G)$. In the experiment, we fixed the values of λ and η, as $\lambda = 1.0$ and $\eta = 10.0$.

D. GENERATOR ARCHITECTURE

As shown in Figure 1(b), the generator of our method is composed of two SCUNet modules. Each module consists of long skip connected residual blocks, encoding blocks, decoding blocks, and SCA. We used SCA after the decoding block in the up-sampling process. As shown in Figure 2, each encoding block and decoding block is composed of two cascaded short-skip connection residual blocks. SCUNet consists of four encoding blocks, four decoding blocks, and a 2D convolution layer. Among them, the encoding block is green, and the decoding block is blue. The 4D tensor is used as input, with 2D convolution layer, filter size of 3×3, and stride of 2. The corresponding encoder and decoder are connected by a global long skip connection. Each local short skip connection residual block is orange, which can be used to increase the depth of the network. It is composed of three convolutional layers (conv0, conv1, and conv2) and one SCA block. Among them, conv0 is a convolution with stride of 1, filter size of 3×3. Conv1 and conv0 have the same stride and filter size, but the number of feature maps is feature number/2. The stride, filter size, and feature number of conv2 are the same as those of conv0. The output of conv2 is used as the input of the SCA block. The structure of the SCA block will be introduced in the following section. Each encoding block adopts a down-sampling method, while each decoding block adopts an up-sampling method. After each layer of decoding blocks, a SCA block is used to extract the details of the image. The input of each SCA block comes from the decoding block of the same layer, and the output is used as the input of the next block. After the last SCA block, a 2D convolution is used, and stride is 1. In order to obtain the output image with the same size of the input image, the same mode is used in the convolution operation. The dual cascade generator is applied, although one generator can also complete end-to-end operations, i.e., from zero-filling MRI image to the final reconstructed image. However, in some iterative methods, additional operations are required to substantiate these findings. Therefore, similar to the recurrent neural network, the double cascade generator can also be regarded as a recurrent network.

E. SPATIAL AND CHANNEL-WISE ATTENTION

The method based on self-attention [28] mechanism has demonstrated that attention module can make the network to extract the detailed information from the MR image and reconstruct an image close to the GT image. Meanwhile, the attention mechanism in GAN facilitates GAN to generate images that fulfill the clinical medical standards. Previous studies have shown [29]–[33] that channel-wise attention in GAN can be used to reconstruct more details in MR images than other methods without channel-wise attention. However, the spatial attention has been neglected by these approaches. The spatial attention can facilitate the accurate location of effective information and enhances the learning efficiency of the neural network, thereby ensuring that the details in the MR image are not missed. Thus, according to the SCA-CNN [26] and UCA [27], we added spatial attention and channel-wise attention to GAN and used SCA in multiple modules. The structure diagram of SCA is illustrated in Figure 3.

1) SPATIAL ATTENTION

First, all input feature maps are calculated the mean value at the same position, which decreases the number of channels to one. Thus, the shape of initial feature maps is transformed from W×H×C to W×H×1, where W is width, H is

height. F_{Si} is the spatial visual feature of i-th position, where i=[1,2, . . . ,m] and m = W×H. Then, the feature maps are passed to the attention module and the weights of spatial visual features are randomly assigned at different positions as W_S. The sigmoid activation function is applied to acquire the spatial attention map. The final feature maps are obtained by a skip connection. Thus, the spatial attention can be expressed as follows:

$$\Theta_{SA} = sig\left(mat\left(W_S, F_{Si}\right) + B_S\right) \otimes F \tag{7}$$

where *mat* represents the multiplication of matrix W_S and matrix F_{Si}, $\otimes$ represents the multiplication of the corresponding elements in the two matrices, *sig* is the sigmoid activation function, and B_S is the bias term.

2) CHANNEL-WISE ATTENTION

A global average pooling is first applied to the feature maps obtained from Θ_{SA}, and the results can be expressed as F_{Ci} (i=[1,2,. . . ,c], c means the number of channels). Also, random weights W_C are assigned to different channels and sigmoid activation function are adopted to obtain the channel-wise attention map (1 × 1×C). Thus, the channel-wise attention can be expressed as:

$$\Theta_{SCA} = sig\left(mat\left(W_C, F_{Ci}\right) + B_C\right) \otimes \Theta_{SA} \tag{8}$$

where B_C is the bias term.

3) SPATIAL-CHANNEL ATTENTION

$\mathcal{F}_S$ and $\mathcal{F}_C$ are both linear functions, Ψ_S and Ψ_C can be expressed as follows:

$$\Psi_S = \mathcal{F}_S\left(\Theta_{SA}, F\right) \tag{9}$$

$$\Psi_C = \mathcal{F}_C\left(\Theta_{SCA}, \Psi_S\right) \tag{10}$$

Therefore, S-C is expressed as:

$$\mathcal{A} = \mathcal{T}\left(F, \Psi_S, \Psi_C\right) \tag{11}$$

where $\mathcal{A}$ represents the final feature maps, and $\mathcal{T}$ represents the process function, indicating that the feature map F obtains the attention weight with respect to spatial attention and channel-wise attention.

F. DISCRIMINATOR ARCHITECTURE

As shown in Figure 1(b), the discriminator D consists of six layers. The middle four layers share the same structure of the encoding block of generator G. The first layer is composed of 2D convolution and Leaky ReLU layer, and the last layer is 2D convolution layer. There are two different results of the discriminator. One is $D\left(S_{GT}\right) = 1$ and the other is $D\left(S_{RE}\right) = D\left(G\left(S_{ZF}\right)\right) = 0$.

III. EXPERIMENT

A. MODEL TRAINING

We used tensorpack [34] for training, which is a visualization toolkit for neural network training based on tensorflow [35]. The graphics processing unit (GPU) used for training and testing was NVIDIA Tesla V100 (4 cores, each with 16 GB memory). The optimizer used the ADAM algorithm [36], [37] for 300 epochs with learning rate $1e^{-4}$. The batch size is 100 and the total training time was about 6 h.

B. MRI DATASET

The brain datasets were obtained from Calgary Campinas brain MR raw data repository [38], containing 35 participants. 2500/3500 MRI brain slices were used for training, and the remaining were used for testing. The knee datasets were obtained from Stanford Fully Sampled 3D FSE Knees repository [39], containing 20 subjects. Among them, 1800/2000 slices were used for training, and the remaining for testing. The matrix size of all the image were reshaped to 256 × 256. All fully-sampled k-space data were retrospectively under-sampled using cartesian sampling with sampling rate of 25%, 16.7%, and 12.5%. We also used another two non-cartesian trajectory, spiral and radial undersampling masks. Each mask includes three sampling rates, 30%, 20%, and 10%.

C. EVALUATION METHOD

We compared the proposed method RSCA-GAN with DAGAN, RefineGAN, and RCA-GAN under different sampling rates. The RCA-GAN is modified on UCA [27], using only channel-wise attention. The hyperparameters selection of DAGAN [20] and RefineGAN [22] is based on the previous works. The PSNR, SSIM and NMSE values were used to evaluate the performance of the reconstruction results. Meanwhile, the rank-sum test [40] method was used to test the significant difference among each method. P<0.05 indicates statistical significance.

IV. RESULTS

Figure 4 shows the brain reconstruction results of different methods using the Cartesian under-sampling mask with AF of 4×, 6×, and 8×. When the AF is 4 and 6, ZF has obvious aliasing artifacts. Although DAGAN eliminates most of the artifacts and displays clear brain structure, the error map still shows blurring. The RefineGAN, RCA-GAN, and the proposed RSCA-GAN can reconstruct brain images well. However, both the error map and the quantitative value indicates that the RSCA-GAN shows the best performance. When the AF is 8, ZF is extremely blurry. Meanwhile, the details of the brain cannot be distinguished well in DAGAN. However, RefineGAN removes the aliasing artifacts well and shows acceptable reconstruction result. And RCA-GAN shows better image quality than RefineGAN since it adopts channel-wise attention. The error map indicates that the proposed RSCA-GAN shows the best quantitative values and sharper edge of grey matter in brain.

Figure 5 shows the knee reconstruction results of different methods using the Cartesian under-sampling mask with AF of 4×, 6×, and 8×. When the AF is 4 and 6, ZF presented obvious artifacts and could not distinguish the detail structures of knee. The error map shows that there

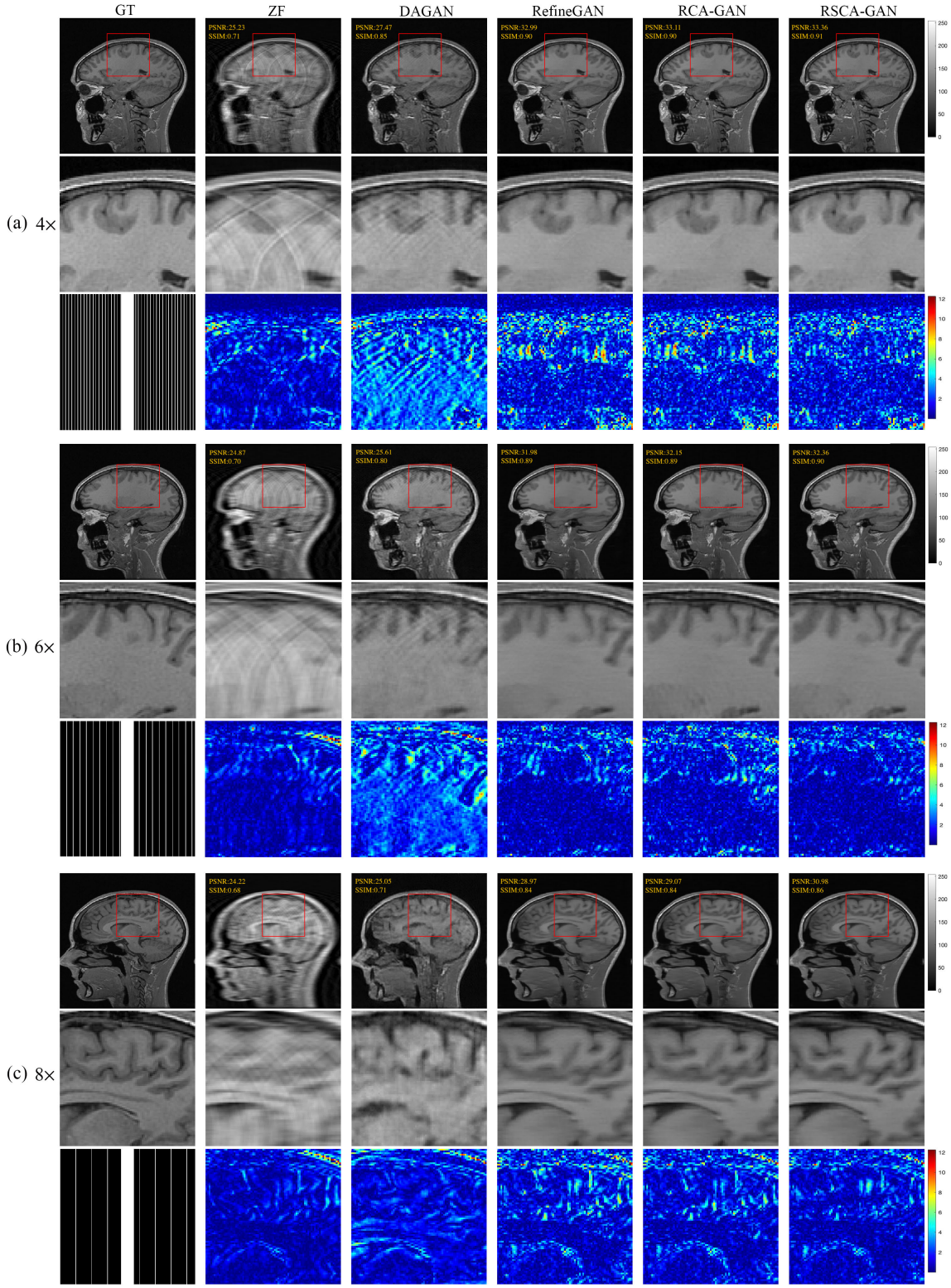

FIGURE 4. The brain reconstruction results of different methods using the Cartesian under-sampling mask with AF of (a) 4×, (b) 6×, and (c) 8×. From left to right, the images are GT, ZF, DAGAN, RefineGAN, RCA-GAN, and RSCA-GAN, respectively. The 1^{st}, 4^{th}, and 7^{th} rows show the reconstruction results, the 2^{nd}, 5^{th}, and 8^{th} rows show the zoomed-in area and the 3^{rd}, 6^{th}, and 9^{th} rows show the error maps for various reconstruction images and the corresponding under-sampling masks.

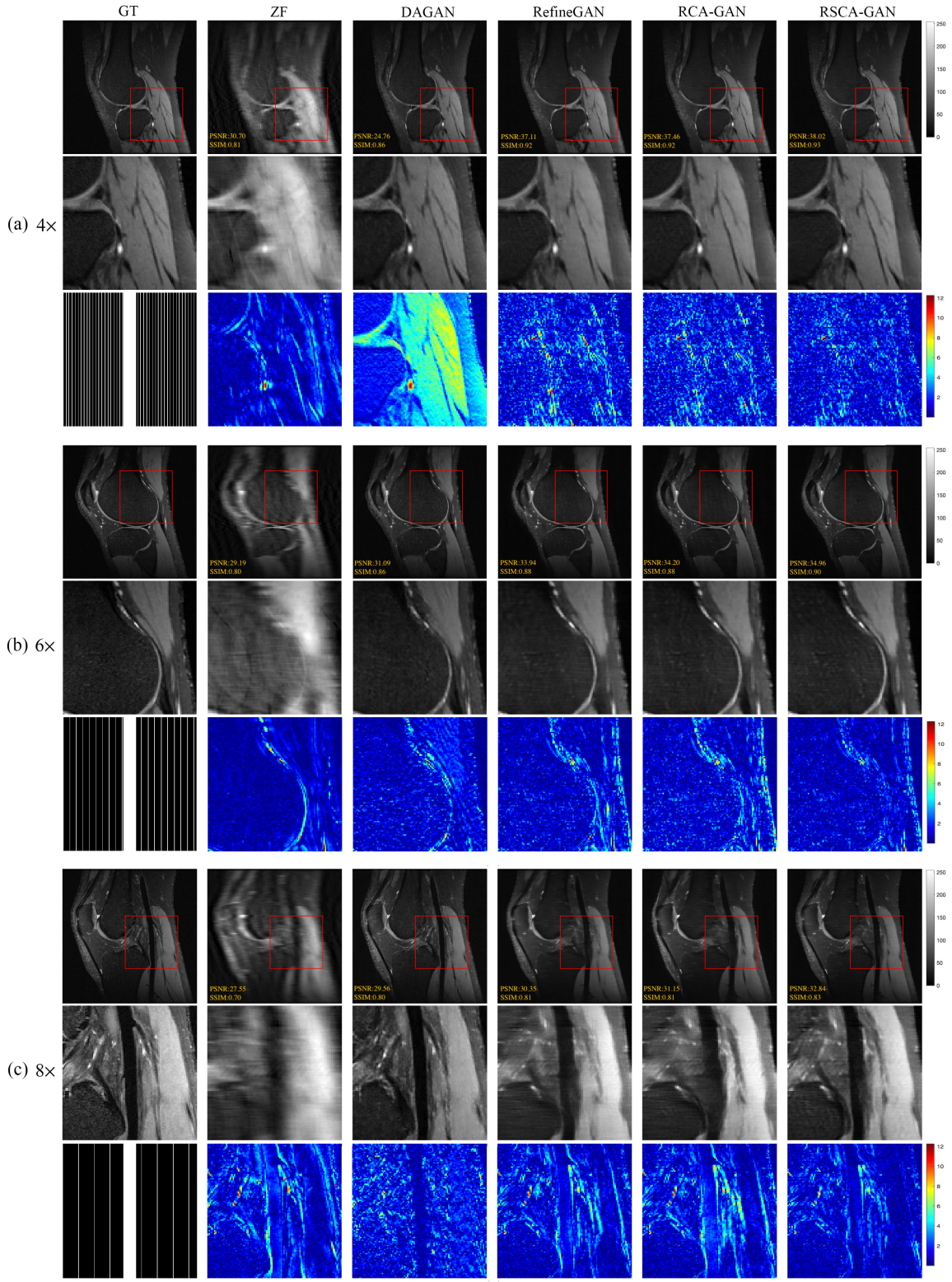

FIGURE 5. The knee reconstruction results of different methods using the Cartesian under-sampling mask with AF of (a) 4×, (b) 6×, and (c) 8×. From left to right, the images are GT, ZF, DAGAN, RefineGAN, RCA-GAN, and RSCA-GAN, respectively. The 1^{st}, 4^{th}, and 7^{th} rows show the reconstruction results, the 2^{nd}, 5^{th}, and 8^{th} rows show the zoomed-in area and the 3^{rd}, 6^{th}, and 9^{th} rows show the error maps for various reconstruction images and the corresponding under-sampling masks.

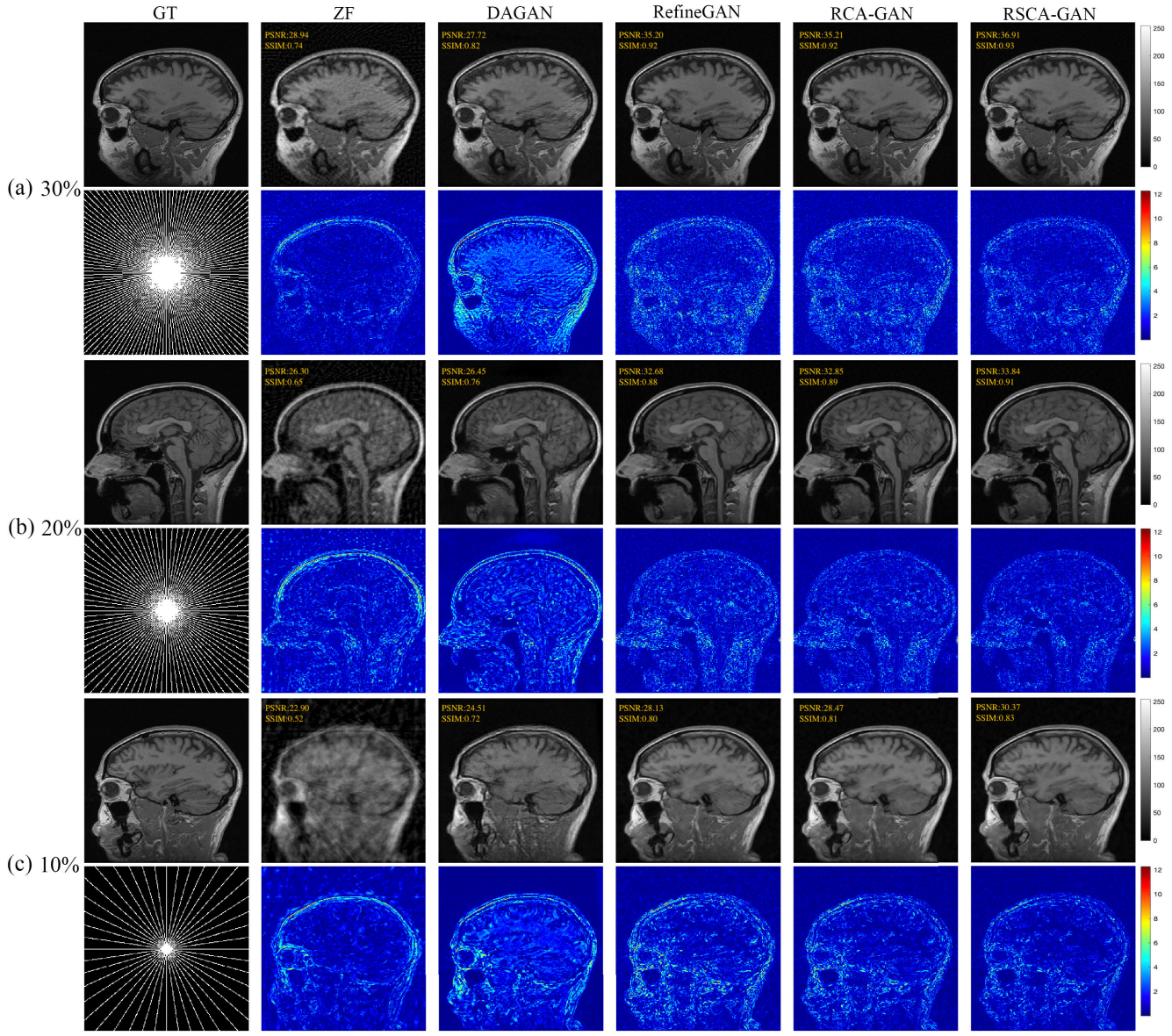

FIGURE 6. The brain reconstruction results of different methods using the radial under-sampling mask with sampling rates of (a) 30%, (b) 20%, and (c) 10%. From left to right, the images are GT, ZF, DAGAN, RefineGAN, RCA-GAN, and RSCA-GAN, respectively. The 1^{st}, 3^{rd}, and 5^{th} rows show the reconstruction results, the 2^{nd}, 4^{th}, and 6^{th} rows show the error maps for various reconstruction images and the corresponding under-sampling masks.

still existed significant differences between DAGAN and GT. The RefineGAN and RCA-GAN shows better results than DAGAN. Moreover, as shown in the zoomed-in area, our proposed RSCA-GAN shows the best results with the least error, especially in the edge area of the knee. When the AF is 8, the RSCA-GAN outperforms the other methods with the highest PSNR and SSIM. Besides, as we can see from the error maps, the RSCA-GAN restores more details such as blood vessels.

Figure 6 and Figure 7 shows the brain reconstruction results of different methods using the radial and spiral under-sampling mask with sampling rates of 30%, 20%, and 10%, respectively. As the sampling rate decreases, ZF displays increasingly obvious artifacts. Although the overall brain structure is recovered by DAGAN, the edge information is still blurred. RefineGAN and RCA-GAN can reconstruct the images well, but the reconstructed images are still very noisy and some details are lost. When the sampling rate is 10%, RSCA-GAN shows the best reconstruction results. It can be seen from the error map that edge details of brain are well recovered. Meanwhile, the PSNR and SSIM values are significantly improved relative to the RCA-GAN.

Figure 8 and Figure 9 shows the knee reconstruction results of different methods using the radial and spiral under-sampling mask with sampling rates of 30%, 20%, and 10%, respectively. As the sampling rates decreases, aliasing

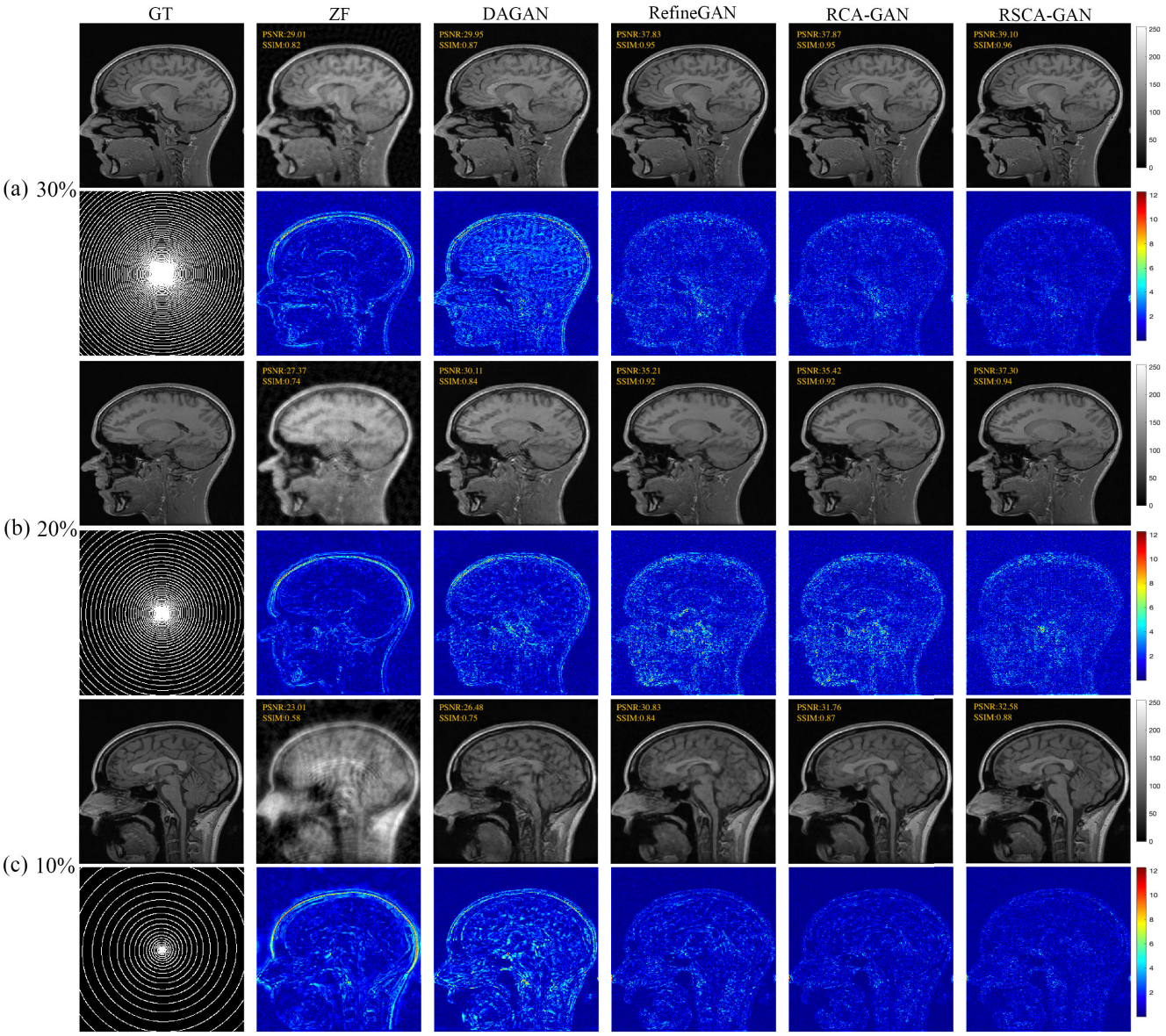

FIGURE 7. The brain reconstruction results of different methods using the spiral under-sampling mask with sampling rates of (a) 30%, (b) 20%, and (c) 10%. From left to right, the images are GT, ZF, DAGAN, RefineGAN, RCA-GAN, and RSCA-GAN, respectively. The 1^{st}, 3^{rd}, and 5^{th} rows show the reconstruction results, the 2^{nd}, 4^{th}, and 6^{th} rows show the error maps for various reconstruction images and the corresponding under-sampling masks.

artifacts become increasingly apparent in the ZF. DAGAN is able to recover most parts of the knee structure, but the errormaps indicate that the reconstruction results still contain residual artifacts. RefineGAN, RCA-GAN, and RSCA-GAN all can obtain acceptable reconstruction results. However, the reconstruction of RSCA-GAN is better in terms of details and edges of the image. In addition, the lower the sampling rate, the larger the difference among them.

Table 1 and Table 2 show the quantitative assessment of PSNR, SSIM, and NMSE values (mean and standard deviation) of brain and knee test images obtained with different reconstruction methods under different sampling rates with different masks, respectively. The quantitative parameters show that RSCA-GAN is significantly better than the other methods ($P<0.05$). Meanwhile, we found that the gap between RSCA-GAN and other methods shows more obvious as the sampling rate decreases.

V. DISCUSSION

The current study proved the feasibility of RSCA-GAN to reconstruct highly under-sampled MRI using both Cartesian and non-Cartesian trajectories. The RSCA-GAN combines the spatial attention, the channel-wise attention and the residual U-net. In terms of all the quantitative metrics, our proposed RSCA-GAN outperforms the DAGAN, RefineGAN, and RCA-GAN.

Previous studies [41], [42] have demonstrated that residual U-net could remove the aliasing artifacts and preserves

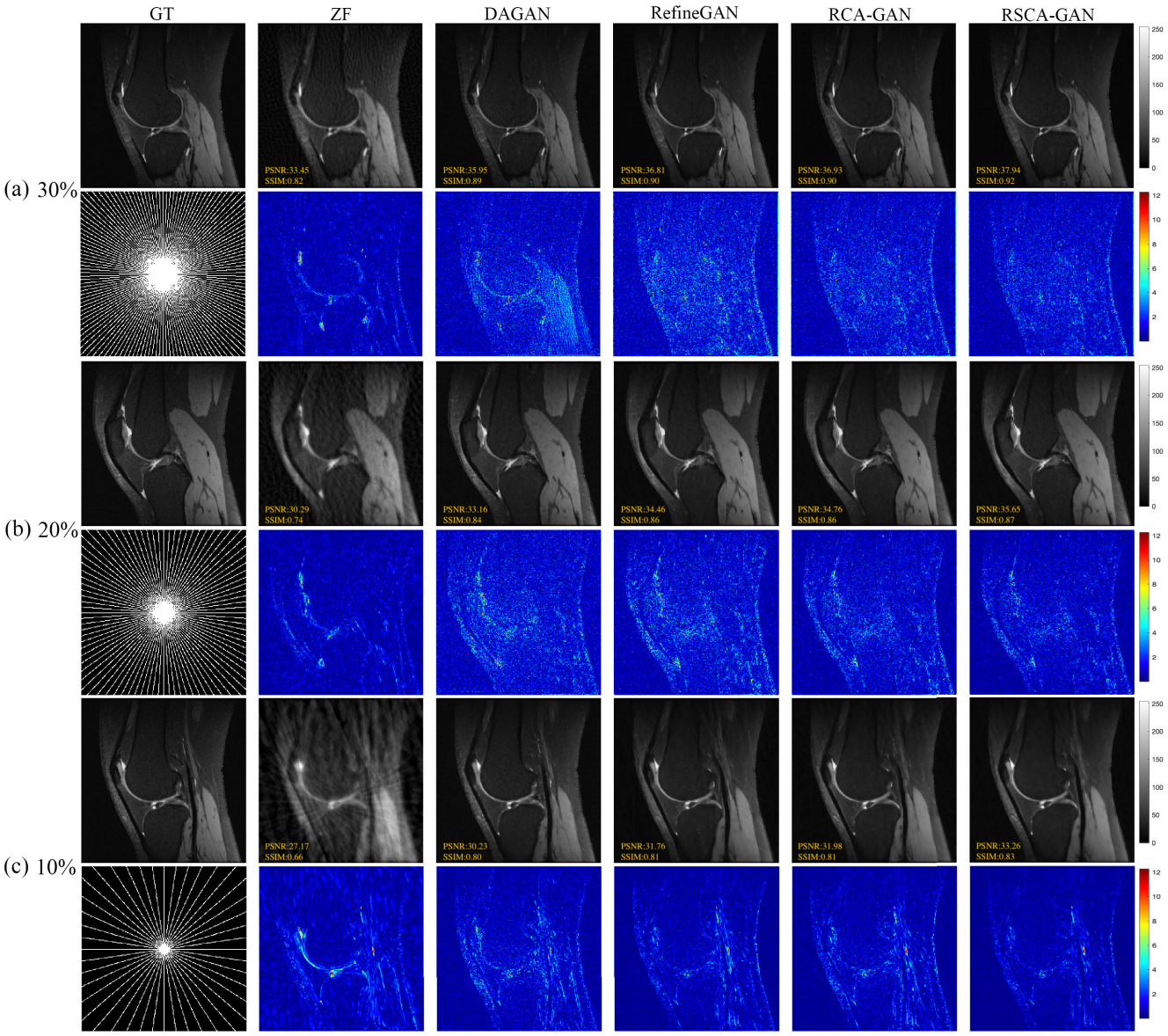

FIGURE 8. The knee reconstruction results of different methods using the radial under-sampling mask with sampling rates of (a) 30%, (b) 20%, and (c) 10%. From left to right, the images are GT, ZF, DAGAN, RefineGAN, RCA-GAN, and RSCA-GAN, respectively. The 1^{st}, 3^{rd}, and 5^{th} rows show the reconstruction results, the 2^{nd}, 4^{th}, and 6^{th} rows show the error maps for various reconstruction images and the corresponding under-sampling masks.

the image details well. The generator of our proposed RSCA-GAN is a variant of residual U-net. The U-net [23] model was first proposed to solve the problem of image segmentation. The long skip connection transfers the useful information of the input image to the up-sampling layers, which restore the loss of image details caused by the down-sampling process. The spatial and channel-wise attention used in our method also utilized long skip connections to fine-tune the output at each stage. This long skip connection is similar to the residual network, which directly passes the input image information to the output through a short skip connection, preventing the loss of image details. The network only needs to learn the difference between the output and the input image, thereby simplifying the difficulty of network learning. In addition, the residual network also could increase the depth of the network.

Our main contribution is to add spatial and channel-wise attention to the residual U-net and refine the framework of the generator. Each encoding and decoding block consist of two short skip connection residual blocks with spatial and channel-wise attention. The attention mechanism is to make the network with human characteristics, so that it can focus on the useful information and ignore other useless information. And the results proved that the RSCA-GAN can better restore the image details. The UCA framework [27] only applied channel-wise attention in the residual U-net

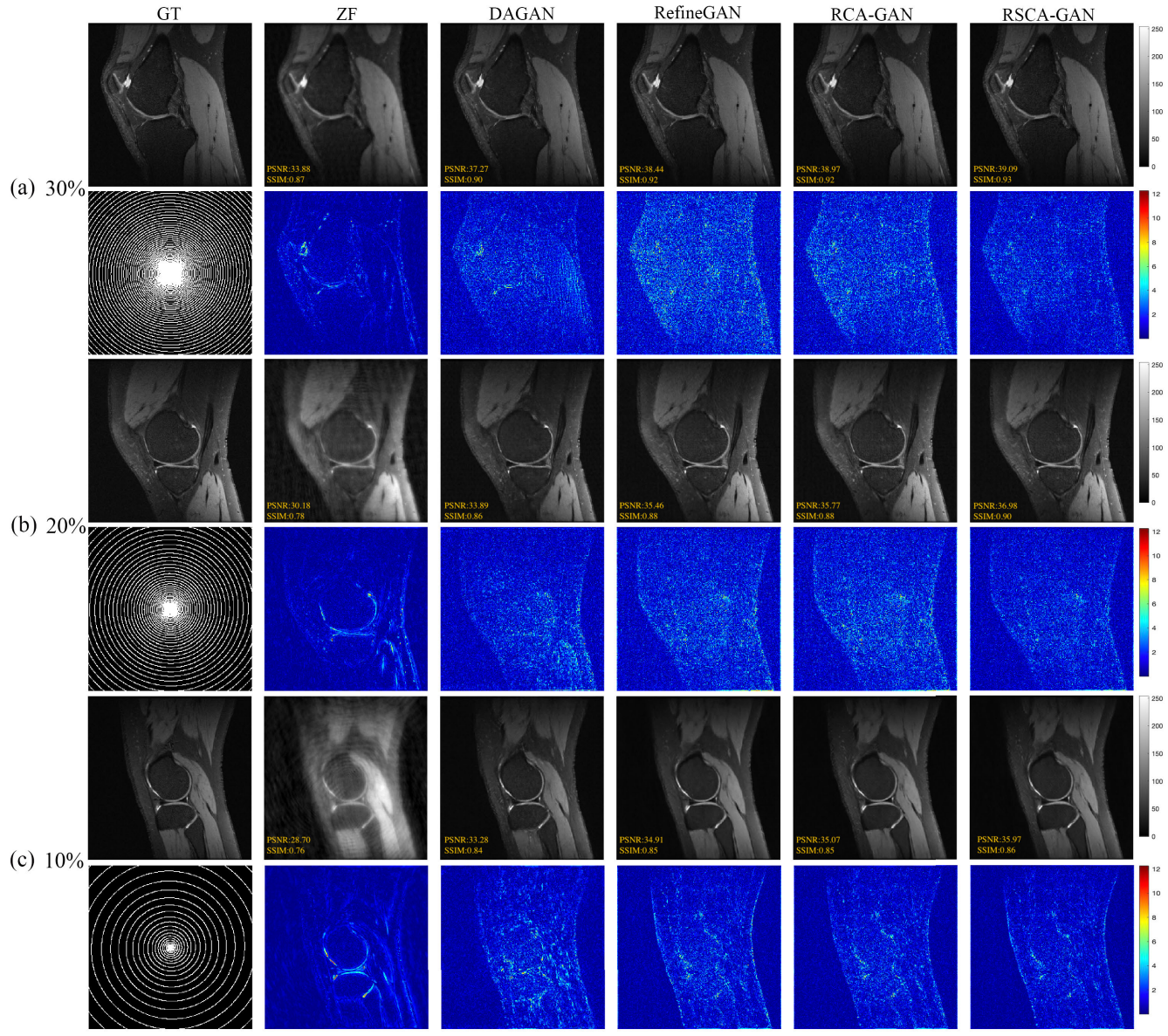

FIGURE 9. The knee reconstruction results of different methods using the spiral under-sampling mask with sampling rates of (a) 30%, (b) 20%, and (c) 10%. From left to right, the images are GT, ZF, DAGAN, RefineGAN, RCA-GAN, and RSCA-GAN, respectively. The 1^{st}, 3^{rd}, and 5^{th} rows show the reconstruction results, the 2^{nd}, 4^{th}, and 6^{th} rows show the error maps for various reconstruction images and the corresponding under-sampling masks.

for MRI reconstruction. Zhu *et al.* [43] showed that spatial attention can be applied to GAN and reconstruct desired MR images. However, the study did not use channel-wise attention. Previous study [26] has demonstrated the feasibility of applying spatial attention and channel-wise attention to CNN for natural image captioning. Thus, inspired by the previous study [44], our proposed RSCA-GAN also adds spatial attention before channel-wise attention in the residual U-net. The spatial attention mechanism [45] enable the network to learn how to pay attention to useful information in image space. Severe aliasing artifacts will appear when MR images suffer from highly under-sampling, but some clear structure is still retained in the images. Spatial attention can make the network to focus on the useful information, so that the network can restore the image details more accurate. The role of the channel attention mechanism is to assign different weights to each channel, allowing the network to focus on the important features and suppress the unimportant ones.

We applied the channel-wise attention after spatial attention in our proposed RSCA-GAN. Thus, it can be deduced that the neural network first obtains the information of the position in the feature maps through spatial attention module. And then, the effective information of that position is passed to channel attention module. However, if there is no spatial attention module before the channel-wise attention module, the network cannot accurately locate the position of

TABLE 1. Quantitative assessment of PSNR, SSIM, and NMSE ($\times 10^{-7}$) values (mean and standard deviation) of brain test images obtained with different reconstruction methods under different sampling rates with different masks. SR: sampling rate.

Mask	SR	Metric	ZF	DAGAN	RefineGAN	RCA-GAN	RSCA-GAN
cartesian	25%	PSNR	25.18(0.68)	26.87(2.30)	32.73(0.94)	33.06(1.01)	**33.14(1.02)**
		SSIM	0.72(0.02)	0.85(0.01)	0.90(0.01)	0.91(0.01)	**0.91(0.01)**
		NMSE	7.33(0.95)	5.44(2.51)	1.31(0.33)	1.22(0.34)	**1.20(0.32)**
	16.7%	PSNR	24.47(0.67)	25.16(1.62)	30.77(0.84)	31.02(0.87)	**31.11(0.83)**
		SSIM	0.70(0.02)	0.77(0.01)	0.87(0.02)	0.88(0.01)	**0.89(0.01)**
		NMSE	8.63(1.11)	7.62(2.34)	1.97(0.44)	1.91(0.43)	**1.89(0.40)**
	12.5%	PSNR	23.09(0.68)	24.02(1.36)	28.32(0.80)	29.47(0.83)	**30.49(0.83)**
		SSIM	0.69(0.02)	0.73(0.02)	0.84(0.01)	0.85(0.01)	**0.86(0.01)**
		NMSE	12.13(9.42)	9.76(2.38)	5.24(2.83)	4.51(2.75)	**3.40(2.41)**
radial	30%	PSNR	28.83(0.72)	28.99(2.33)	35.21(1.05)	35.22(1.09)	**36.98(1.06)**
		SSIM	0.73(0.02)	0.82(0.02)	0.92(0.01)	0.92(0.01)	**0.93(0.01)**
		NMSE	3.16(0.41)	3.31(2.49)	0.74(0.18)	0.74(0.19)	**0.67(0.18)**
	20%	PSNR	26.47(0.64)	26.96(1.90)	32.83(0.94)	33.07(0.95)	**33.95(0.93)**
		SSIM	0.65(0.02)	0.78(0.02)	0.89(0.02)	0.90(0.01)	**0.91(0.01)**
		NMSE	5.43(0.66)	5.47(2.41)	1.27(0.25)	1.20(0.49)	**1.13(0.26)**
	10%	PSNR	23.14(0.57)	24.15(1.32)	28.23(0.87)	29.22(0.85)	**30.19(0.84)**
		SSIM	0.52(0.02)	0.71(0.03)	0.82(0.02)	0.83(0.02)	**0.84(0.02)**
		NMSE	11.61(1.19)	9.45(2.23)	3.13(5.72)	2.94(5.45)	**2.76(5.31)**
spiral	30%	PSNR	29.55(0.71)	29.87(2.79)	38.61(0.82)	38.76(0.90)	**39.84(0.86)**
		SSIM	0.83(0.01)	0.87(0.01)	0.95(0.00)	0.95(0.00)	**0.96(0.00)**
		NMSE	2.68(0.35)	2.55(2.55)	0.34(0.06)	0.32(0.06)	**0.21(0.07)**
	20%	PSNR	26.62(0.63)	28.87(2.29)	35.11(0.85)	35.66(0.87)	**36.68(0.84)**
		SSIM	0.73(0.01)	0.80(0.04)	0.92(0.01)	0.92(0.01)	**0.93(0.01)**
		NMSE	5.26(0.61)	4.34(2.51)	0.69(0.14)	0.66(0.13)	**0.59(0.13)**
	10%	PSNR	22.96(0.57)	25.09(1.59)	30.91(0.84)	31.95(0.92)	**32.09(0.89)**
		SSIM	0.60(0.02)	0.76(0.03)	0.86(0.02)	0.88(0.01)	**0.89(0.01)**
		NMSE	12.16(1.14)	7.73(2.33)	1.97(0.36)	1.56(0.36)	**1.50(0.35)**

Bold, the best results.

TABLE 2. Quantitative assessment of PSNR, SSIM, and NMSE ($\times 10^{-7}$) values (mean and standard deviation) of knee test images obtained with different reconstruction methods under different sampling rates with different masks. SR: sampling rate.

Mask	SR	Metric	ZF	DAGAN	RefineGAN	RCA-GAN	RSCA-GAN
cartesian	25%	PSNR	27.88(1.70)	27.31(5.45)	36.83(1.33)	37.00(1.12)	**37.07(1.12)**
		SSIM	0.76(0.03)	0.86(0.06)	0.91(0.01)	0.91(0.01)	**0.92(0.01)**
		NMSE	8.21(2.80)	8.61(4.18)	1.25(0.03)	1.21(0.03)	**1.19(0.02)**
	16.7%	PSNR	27.36(1.63)	26.85(5.28)	34.73(1.25)	34.95(1.16)	**35.28(1.27)**
		SSIM	0.74(0.03)	0.80(0.07)	0.88(0.02)	0.88(0.01)	**0.89(0.02)**
		NMSE	9.19(2.89)	9.68(4.29)	2.06(0.06)	1.96(0.06)	**1.80(0.04)**
	12.5%	PSNR	28.11(1.29)	29.85(1.57)	30.81(1.16)	31.65(1.27)	**32.08(1.12)**
		SSIM	0.74(0.04)	0.81(0.02)	0.81(0.02)	0.82(0.03)	**0.83(0.02)**
		NMSE	8.10(1.82)	5.53(1.41)	4.58(1.84)	3.68(1.80)	**3.25(1.28)**
radial	30%	PSNR	32.69(1.13)	35.02(1.39)	36.21(1.29)	36.57(1.24)	**37.39(1.20)**
		SSIM	0.80(0.03)	0.87(0.02)	0.89(0.02)	0.89(0.02)	**0.90(0.02)**
		NMSE	2.81(0.65)	2.06(0.44)	1.24(0.25)	1.19(0.24)	**1.09(0.23)**
	20%	PSNR	30.77(1.15)	33.69(1.51)	34.72(1.30)	34.91(1.14)	**35.83(1.20)**
		SSIM	0.75(0.03)	0.84(0.02)	0.86(0.02)	0.86(0.01)	**0.87(0.01)**
		NMSE	4.38(0.98)	2.84(0.86)	1.77(0.41)	1.69(0.40)	**1.62(0.41)**
	10%	PSNR	27.95(1.24)	30.38(1.61)	31.89(1.04)	31.95(1.08)	**32.82(1.04)**
		SSIM	0.68(0.04)	0.79(0.03)	0.81(0.02)	0.81(0.02)	**0.82(0.02)**
		NMSE	8.36(1.74)	4.13(2.70)	3.42(1.00)	3.35(1.07)	**3.17(0.92)**
spiral	30%	PSNR	33.48(1.12)	36.19(1.35)	38.49(1.33)	38.79(1.41)	**39.07(1.30)**
		SSIM	0.86(0.02)	0.90(0.01)	0.92(0.01)	0.92(0.01)	**0.93(0.01)**
		NMSE	2.34(0.52)	1.07(0.31)	0.74(0.15)	0.72(0.15)	**0.68(0.17)**
	20%	PSNR	30.97(1.14)	34.88(1.39)	36.75(1.27)	36.83(1.39)	**37.81(1.30)**
		SSIM	0.81(0.03)	0.88(0.02)	0.90(0.02)	0.90(0.02)	**0.91(0.02)**
		NMSE	4.16(0.88)	1.78(0.68)	1.10(0.22)	1.04(0.24)	**0.98(0.22)**
	10%	PSNR	27.84(1.20)	32.12(1.46)	34.42(1.22)	34.64(1.44)	**35.14(1.33)**
		SSIM	0.73(0.03)	0.82(0.02)	0.85(0.02)	0.85(0.03)	**0.86(0.02)**
		NMSE	8.55(1.64)	3.45(2.23)	1.88(0.42)	1.86(0.43)	**1.81(0.47)**

Bold, the best results.

the useful information. Although it is also possible to assign channel attention weights by learning all spatial information, the learning efficiency of neural networks is low, and some image details will be missed. Therefore, we used spatial attention before channel-wise to resolve the problem of partial image detail loss caused by allocating weights in all spaces.

Therefore, consistent with the previous findings [26], [45], the RSCA-GAN suggested better reconstruction results than RCA-GAN, which means that using channel-wise attention after spatial attention can obtain higher image quality than only using channel-wise attention.

RefineGAN has demonstrated that better reconstructed images will be obtained using two cascaded generators compared to using only one generator. Therefore, our method also adopts two residual U-nets with attention modules as the generator of the network, which deepens the depth of the generator to extract more image details. Besides, the results show that RSCA-GAN can obtain the optimal reconstruction regardless of whether the sampling mask is Cartesian or non-Cartesian. This indicates that the reconstruction process is more similar to image denoising and is independent of the image contents.

Nevertheless, our study has some limitations. First, the input of the network is a single-channel image, while the images obtained in clinical scanners are multi-channel. Thus, it is necessary to synthesize the multi-channel MR image into a single-channel, which increases the workload. In the future study, we aim to change the input of the network from single-channel to multi-channel, which is consistent with clinical routine requirement. Second, the network uses supervised learning during network training. Thus, we should use an unsupervised training method in the network when it is difficult to obtain fully-sampled data.

VI. CONCLUSION

In this study, we proposed a modified generative adversarial network called RSCA-GAN by combining spatial and channel-wise attention for CS-MRI reconstruction. This method can reconstruct MR images under high acceleration factors with both Cartesian and non-Cartesian under-sampling masks. The results demonstrated that RSCA-GAN outperforms the other approaches for all the quantitative metrics. Thus, the application of RSCA-GAN in clinical practice is promising.

REFERENCES

[1] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[2] T. Eo, Y. Jun, T. Kim, J. Jang, H. Lee, and D. Hwang, "KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2188–2201, Nov. 2018.

[3] E. T. Tan, S.-K. Lee, P. T. Weavers, D. Graziani, J. E. Piel, Y. Shu, J. Huston, M. A. Bernstein, and T. K. F. Foo, "High slew-rate head-only gradient for improving distortion in echo planar imaging: Preliminary experience," *J. Magn. Reson. Imag.*, vol. 44, no. 3, pp. 653–664, Sep. 2016.

[4] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akcakaya, "Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 128–140, Jan. 2020.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] Y. Chen, Q. Zhao, X. Hu, and B. Hu, "Multi-resolution parallel magnetic resonance image reconstruction in mobile computing-based IoT," *IEEE Access*, vol. 7, pp. 15623–15633, 2019.

[7] C. Qing-Huan, S. Guan-Qun, and N. Sheng-Dong, "Compressive sensing low-field mri reconstruction with dual-tree wavelet transform and wavelet tree sparsity," *Chin. J. Magn. Reson.*, vol. 35, no. 4, pp. 486–497, 2018.

[8] D. Liang, H. Wang, Y. Chang, and L. Ying, "Sensitivity encoding reconstruction with nonlocal total variation regularization," *Magn. Reson. Med.*, vol. 65, no. 5, pp. 1384–1392, May 2011.

[9] Q. Liu, S. Wang, L. Ying, X. Peng, Y. Zhu, and D. Liang, "Adaptive dictionary learning in sparse gradient domain for image recovery," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4652–4663, Dec. 2013.

[10] S. Yang, X. Hai-bin, and Y. Guang, "Dictionary learning with segmentation for compressed-sensing magnetic resonance imaging," *Chin. J. Magn. Reson.*, vol. 33, no. 4, pp. 559–569, 2016.

[11] O. N. Jaspan, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: A review of the clinical literature," *Brit. J. Radiol.*, vol. 88, no. 1056, Dec. 2015, Art. no. 20150487.

[12] K. G. Hollingsworth, "Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction," *Phys. Med. Biol.*, vol. 60, no. 21, pp. R297–R322, Nov. 2015.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 2672–2680.

[14] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[15] T. M. Quan, T. Nguyen-Duc, and W. K. Jeong, "Compressed sensing MRI reconstruction with cyclic loss in generative adversarial networks," 2017, *arXiv:1709.00753*. [Online]. Available: http://arxiv.org/abs/1709.00753

[16] Y. Li, J. Li, F. Ma, S. Du, and Y. Liu, "High quality and fast compressed sensing MRI reconstruction via edge-enhanced dual discriminator generative adversarial network," *Magn. Reson. Imag.*, vol. 77, pp. 124–136, Apr. 2021.

[17] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[18] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[20] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.

[21] T. Chen, X. Song, and C. Wang, "Preserving-texture generative adversarial networks for fast multi-weighted MRI," *IEEE Access*, vol. 6, pp. 71048–71059, 2018.

[22] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] C. Han, L. Rundo, R. Araki, Y. Nagano, Y. Furukawa, G. Mauri, H. Nakayama, and H. Hayashi, "Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection," *IEEE Access*, vol. 7, pp. 156966–156977, 2019.

[26] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5659–5667.

[27] Q. Huang, D. Yang, P. Wu, H. Qu, J. Yi, and D. Metaxas, "MRI reconstruction via cascaded channel-wise attention network," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1622–1626.

[28] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," *CoRR*, vol. abs/1803.02155, pp. 464–468, Mar. 2018.
[29] Y. Guo, C. Wang, H. Zhang, and G. Yang, "Deep attentive wasserstein generative adversarial networks for mri reconstruction with recurrent context-awareness," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 167–177.
[30] Z. Zhang and L. Schomaker, "DTGAN: Dual attention generative adversarial networks for text-to-image generation," 2020, *arXiv:2011.02709*. [Online]. Available: http://arxiv.org/abs/2011.02709
[31] Z. Yuan, M. Jiang, Y. Wang, B. Wei, Y. Li, P. Wang, W. Menpes-Smith, Z. Niu, and G. Yang, "SARA-GAN: Self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing MRI reconstruction," *Frontiers Neuroinform.*, vol. 14, Nov. 2020, Art. no. 611666.
[32] Z. Geng, J. Tao, and J. Xu, "A modified GAN for compressed sensing MRI," *J. Phys., Conf. Ser.*, vol. 1642, Sep. 2020, Art. no. 012001.
[33] Z. Huang, Z. Chen, Q. Zhang, G. Quan, M. Ji, C. Zhang, Y. Yang, X. Liu, D. Liang, H. Zheng, and Z. Hu, "CaGAN: A cycle-consistent generative adversarial network with attention for low-dose CT imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1203–1218, 2020.
[34] (2016). *Tensorpack*. [Online]. Available: https://githubcom/tensorpack/
[35] *Tensorflow*. Accessed: 2017. [Online]. Available: http://www.tensorflow.org/
[36] J. Sun, J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 10–18.
[37] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-Net: A deep learning approach for compressive sensing MRI," 2017, *arXiv:1705.06869*. [Online]. Available: http://arxiv.org/abs/1705.06869
[38] R. Souza, O. Lucena, J. Garrafa, D. Gobbi, M. Saluzzi, S. Appenzeller, L. Rittner, R. Frayne, and R. Lotufo, "An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement," *NeuroImage*, vol. 170, pp. 482–494, 2018. [Online]. Available: https://sites.google.com/view/calgary-campinas-dataset/home
[39] K. Epperson, A. M. Sawyer, M. Lustig, M. Alley, and M. Uecker, "Creation of fully sampled MR data repository for compressed sensing of the knee," in *Proc. 22nd Annu. Meeting Sect. Magn. Reson. Technol.*, Salt Lake City, UT, USA, 2013. [Online]. Available: http://mridata.org/
[40] L. Deng, J. Pei, J. Ma, and D. L. Lee, "A rank sum test method for informative gene discovery," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Seattle, WA, USA, Aug. 2004, pp. 410–419.
[41] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 109–119, Mar. 2019.
[42] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 4509–4522, Jun. 2016.
[43] W. Zhou, H. Du, W. Mei, and L. Fang, "Spatial orthogonal attention generative adversarial network for MRI reconstruction," *Med. Phys.*, vol. 48, no. 2, pp. 627–639, Feb. 2021.
[44] H. Xu and K. Saenko, "Ask, attend and answer: Exploring question-guided spatial attention for visual question answering," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 451–466.
[45] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6688–6697.

GUANGYUAN LI was born in Jining, Shandong, China, in 1997. He is currently pursuing the M.S. degree with the School of Computer and Control Engineering, Yantai University. His research interests include medical image processing and deep learning.

JUN LV was born in Yantai, Shandong, China, in 1990. She received the B.S. degree in intelligence science and technology from Xidian University, in 2013, and the Ph.D. degree in biomechanics and medical engineering from Peking University, in 2018. She is currently a Lecturer with the School of Computer and Control Engineering, Yantai University. Her research interests include medical image reconstruction, image registration, and image segmentation.

CHENGYAN WANG received the B.S. degree in biomedical engineering from the Beijing Institute of Technology, Beijing, China, in 2012, and the Ph.D. degree in biomechanics and medical engineering from Peking University, Beijing, in 2017.

He was a Postdoctoral Fellow with the Institute of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, from 2017 to 2018, and the Institute of Electronic Engineering, University of Illinois at Urbana–Champaign, IL, USA, from 2018 to 2019. Since 2019, he has been an Assistant Professor with the Human Phenome Institute, Fudan University, Shanghai. His research interests include image reconstruction, medical image analysis, and deep learning.