

FA-GAN: Fused attentive generative adversarial networks for MRI image super-resolution



Mingfeng Jiang^{a,*}, Minghao Zhi^a, Liying Wei^a, Xiaocheng Yang^a, Jucheng Zhang^b, Yongming Li^c, Pin Wang^c, Jiahao Huang^{d,e}, Guang Yang^{d,e,*}

^a School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, 310018, China

^b Department of Clinical Engineering, the Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, 310019, China

^c College of Communication Engineering, Chongqing University, Chongqing, China

^d Cardiovascular Research Centre, Royal Brompton Hospital, London, SW3 6NP, UK

^e National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK

ARTICLE INFO

Keywords:

Super-resolution
Generative adversarial networks
Attention
Mechanism
MRI

ABSTRACT

High-resolution magnetic resonance images can provide fine-grained anatomical information, but acquiring such data requires a long scanning time. In this paper, a framework called the Fused Attentive Generative Adversarial Networks(FA-GAN) is proposed to generate the super-resolution MR image from low-resolution magnetic resonance images, which can reduce the scanning time effectively but with high resolution MR images. In the framework of the FA-GAN, the local fusion feature block, consisting of different three-pass networks by using different convolution kernels, is proposed to extract image features at different scales. And the global feature fusion module, including the channel attention module, the self-attention module, and the fusion operation, is designed to enhance the important features of the MR image. Moreover, the spectral normalization process is introduced to make the discriminator network stable. 40 sets of 3D magnetic resonance images (each set of images contains 256 slices) are used to train the network, and 10 sets of images are used to test the proposed method. The experimental results show that the PSNR and SSIM values of the super-resolution magnetic resonance image generated by the proposed FA-GAN method are higher than the state-of-the-art reconstruction methods.

1. Introduction

Image super-resolution refers to the reconstruction of high-resolution images from low-resolution images (Dong et al., 2016). High resolution means that the pixels in the image are denser and can display more flexible details (Li et al., 2019; Gholipour et al., 2010). These details are very useful in practical applications, such as satellite imaging, medical imaging, etc, which can better identify targets and find important features in high-resolution images (Zeng et al., 2019; Wang et al., 2020, 2019).

High-resolution (HR) MRI images can provide fine anatomical information, which is helpful for clinical diagnosis and accurate decision-making (Wang et al., 2018; Manjon et al., 2010). However, it not only requires expensive equipment but also requires a long scanning time, which brings challenges to image data acquisition. Therefore, further

applications are limited by slow data acquiring and imaging speed (Jafari-Khouzani, 2014; Rueda et al., 2013).

The super-resolution (SR) is a technique to generate a high-resolution (HR) image from a single or a group of low-resolution (LR) images, which can improve the visibility of image details or restore image details (Tourbier et al., 2015; Dong et al., 2016; Shi et al., 2018a). Without changing hardware or scanning components, SR methods can significantly improve the spatial resolution of MRI (Mahmoudzadeh and Kashou, 2014; Luo et al., 2017). Generally, there are three methods to implement image SR in MRI: interpolation-based, construction-based, and machine learning-based (Shi et al., 2018b; Jog et al., 2014).

The interpolation-based SR techniques assume that the area in the LR image can be extended to the corresponding area by using a polynomial or an interpolation function with a priori smoothness (Shi et al., 2016; Seeliger et al., 2018). The advantages of the interpolation-based

* Corresponding author at: National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK.

** Corresponding author at: School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, 310018, China.

E-mail addresses: m.jiang@zstu.edu.cn (M. Jiang), g.yang@imperial.ac.uk (G. Yang).

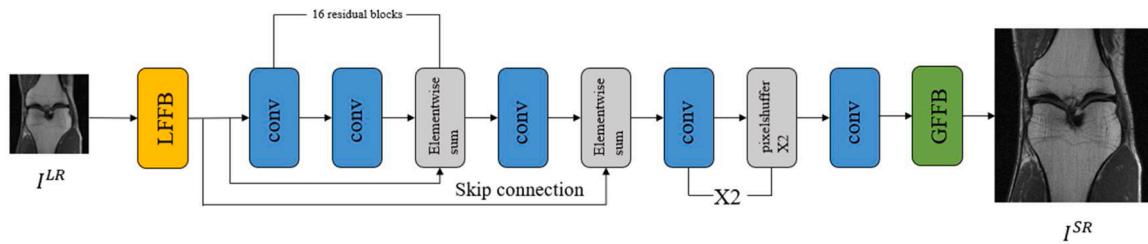


Fig. 1. The framework of the proposed FA-GAN network. LFFB denotes local feature fusion block and GFFB denotes global feature fusion block.

super-resolution reconstruction algorithm are simplicity and high real-time performance; the disadvantage is that it is too simple to make full use of the prior information of MR images. In particular, the super-resolution reconstruction algorithm based on a single MR image has obvious shortcomings, which in a blurred version of the corresponding HR reference image Huang et al. (2017); Armanious et al. (2019).

The reconstruction-based SR methods are introduced to solve an optimization problem incorporating two terms: the fidelity term, which penalizes the difference between a degraded SR image and an observed LR image, and the regularization term, which promotes sparsity and inherent characteristics of recovering the SR signal (Goodfellow et al., 2014; Luo et al., 2017). The performance of these techniques becomes suboptimal especially in the high-frequency region when the input data becomes too sparse or the model becomes even slightly inaccurate (Quan et al., 2018; Ledig et al., 2017a; Huang et al., 2017). These shortcomings reduce the effect of reconstruction-based SR methods to large magnifications, which may work well for small magnifications less than 4.

Machine learning techniques, particularly deep learning (DL)-based SR approaches, have recently attracted considerable attention because of their state-of-the-art performance in SR for natural images. Most recent algorithms rely on data-driven deep learning models to reconstruct the required details for accurate super-resolution (Liu et al., 2018; Latif et al., 2018). Deep learning-based methods aim to automatically learn the relationship between input and output directly from the training samples (Dong et al., 2014a; Zhang et al., 2018). At the same time, deep learning has also played a vital role in CT/PET image reconstruction, such as PET Image Reconstruction from Sinogram Domain (Liu et al., 2019; Hu et al., 2020; Häggström et al., 2019).

With the development of deep learning, the Generative Adversarial Network (GAN) proposed by Goodfellow et al., has recently been demonstrated that it has good performance in image transformation and super-resolution imaging. Sanchez et al. proposed the standard super-resolution GAN (SRGAN) framework for generating brain super-resolution images (Ramachandran et al., 2017). Most GAN-based image generation models are constructed using convolutional layers. Convolutions process information in local neighborhoods, however, using only convolutional layers is inefficient in establishing remote dependencies in images (Miyato et al., 2018; Iandola et al., 2016).

It is difficult to learn the dependencies between images using a small convolution kernel. However, the size of the convolution kernel is too large, which will reduce the model's performance. Besides, increasing the size of the convolution kernel can also expand the receptive field, but it inevitably increases the complexity of the model (Quan et al., 2018; Ledig et al., 2017b). Zhang et al. propose the Self-Attention Generative Adversarial Network (SAGAN) with attention-driven, long-range dependency modeling for image generation tasks (Liu et al., 2018).

In the previous work on reconstruction problems, deep learning based methods have two major issues (Latif et al., 2018). Firstly, they treat each channel-wise feature equally, but contributions to the reconstruction task vary from different feature maps. Secondly, the receptive field in a convolutional layer may cause to lose contextual information from original images, especially those high-frequency

components that contain valuable detailed information such as edges and texture. Therefore, the Channel-Attention module is designed to filter the useless features and to enhance the informative ones. Therefore, model parameters in shallower layers are to be updated mostly that are relevant to a given task. To the best of our knowledge, this is the first work to employ channel-wise attention to the MRI reconstruction problem (Dong et al., 2014b; Ting and Xiao, 2019). Combining the idea of MR reconstruction and image super-resolution, some researchers work on recovering HR images from low-resolution under-sampled K-space data directly (Shi et al., 2018c; Xu et al., 2018; Huang et al., 2021; Hu et al., 2021).

In this paper, a fused attentive generative adversarial network (FA-GAN) is proposed for generating super-resolution MR images from low-resolution MR ones. The novelty of this work can be concluded as following: 1)The local fusion feature block, consisting of different three-pass networks by using different convolution kernels, was proposed to extract image features at different scales, so as to improve the reconstruction performances of SR images; 2) The global feature fusion module, including the channel attention module, the self-attention module, and the fusion operation, was designed to enhance the important features of the MRI image, so that the super-resolution image is more realistic and closer to the original image; 3)The spectral normalization (SN) is introduced to the discriminator network, which can not only smooth and accelerate the training process of the deep neural network but also improve the model generalization performance.

2. Methodology

The proposed neural network model is designed to learn the image firstly, and then inversely map the LR image to the reference HR image Zhu et al. (2019); Bello et al. (2019). This model only takes LR images as input to generate SR images. The operation can be defined as

$$I^{LR} = f(I^{HR}) \quad (1)$$

$$I^{HR} = g(I^{LR}) = f^{-1}(I^{LR}) + R \quad (2)$$

where $I^{LR}, I^{HR} \in \mathbb{R}^{m \times n}$ are respectively LR and HR MRI images of size $m \times n$ and $f: I^{HR} \in \mathbb{R}^{m \times n} \rightarrow I^{LR} \in \mathbb{R}^{m \times n}$ denotes the down-sampling process that creates a LR counterpart from an HR image.

2.1. SR network with GAN

The network output is passed through a series of upsampling stages, where each stage doubles the input image size. The output is passed through a convolution stage to get the resolved image. Depending upon the desired scaling, the number of upsampling stages can be changed. The adversarial min-max problem is defined by

$$\min_{G} \max_{D} E_{I^{HR} \sim P_{train}(I^{HR})} [\log D(I^{HR})] + E_{I^{LR} \sim P_G(I^{LR})} [\log (1 - D(G(I^{LR})))] \quad (3)$$

The framework of the proposed FA-GAN network is shown in Fig. 1. The whole model takes the down-sampled low-resolution magnetic resonance image as input, extracts the features through the LFFB module, and generates the enlarged image through convolution and up-

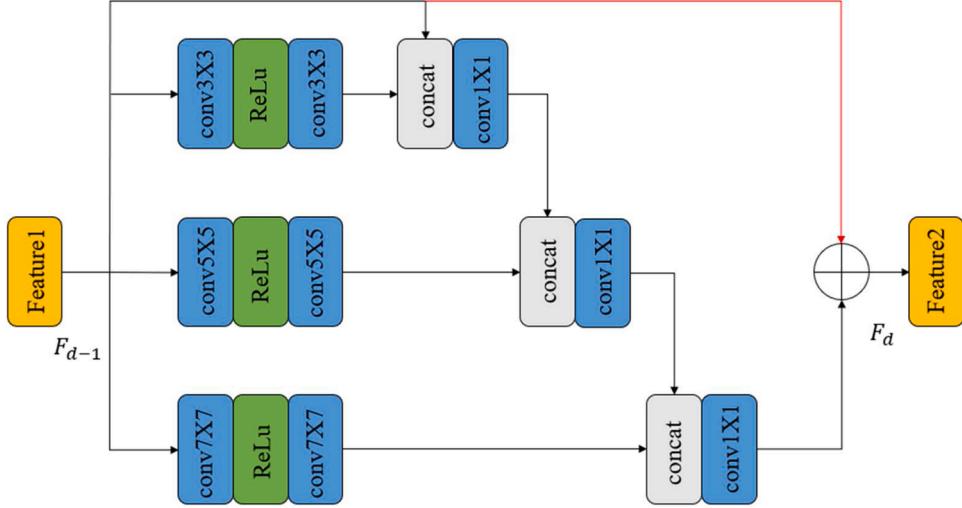


Fig. 2. Local feature fusion block.

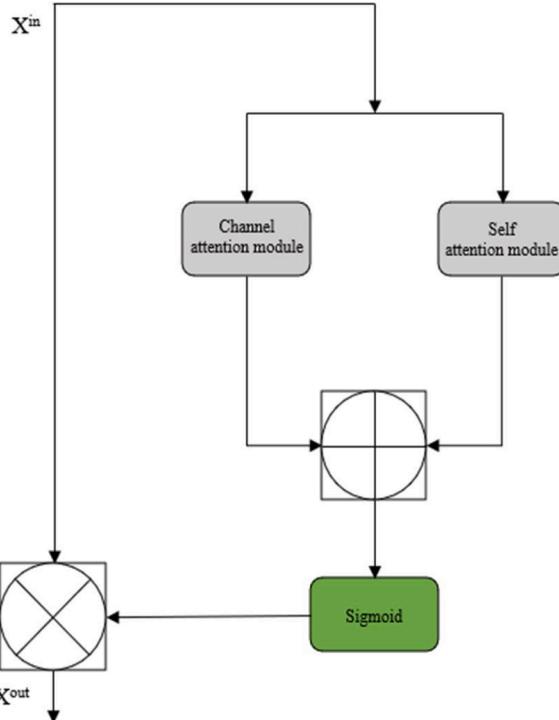


Fig. 3. Global feature fusion block. GAP denotes the global average pooling operation.

sampling. Finally, the GFFB module fuses the detailed features to generate a super-resolution magnetic resonance image. During the training process, HR references will be used to guide the optimization of model parameters. Moreover, the spectral normalization (SN) is introduced to the discriminator network to stabilize the training of GAN.

2.2. Local fusion feature block (LFFB)

Different from those previous experiments (Fu et al., 2019), the local fusion feature block consists of different three-pass networks by using different convolution kernels, as shown in Fig. 2. In this way, the information flows between those bypasses can be shared with each other, which allow our network to extract image features at different scales. The operation can be defined as

$$F_{t,1} = [C_{3 \times 3}(F_{t-1}), F_{t,1}] \quad (4)$$

$$F_{t,2} = [C_{5 \times 5}(F_{t-1}), F_{t,2}] \quad (5)$$

$$F_{t,3} = [C_{7 \times 7}(F_{t-1}), F_{t,3}] \quad (6)$$

$$F_d = F_{d,3} + F_{d-1} \quad (7)$$

where $C_{s \times s}$ means the S scale feature extractor. Ours proposed S scale feature extractor consist of three convolution layers with $s \times s$ kernel size and one ReLU intermediate activation layer. The operation of $F [\cdot]$ means the concatenation and 1×1 convolution, which is mainly designed to quickly fuse features and reduce the computational burden.

2.3. Global feature fusion block (GFFB)

The global feature fusion module includes three parts, namely the channel attention module, the self-attention module, and the fusion operation. Through these modules, the important features of the MRI image can be enhanced, so that the super-resolution image is more realistic and closer to the original image (Fig. 3).

(1) Channel-Attention Module

In this paper, a lightweight channel attention mechanism is introduced, which allows to selectively emphasize informative features and restrain less useful ones via a one-dimensional vector from global information. As illustrated in Fig. 4, a global average pooling is used to extract the global information across spatial dimensions $H \times W$ firstly. Then, it is followed by a dimension reduction layer with a reduction ratio of r , a ReLU activation, a dimension increase layer, and a sigmoid activation to generate SR image. The two dimension computable layers are implemented by fully connected layers. The final output of the recalibration is acquired by rescaling the input features.

(2) Self-Attention Module

The role of the self-attention module is to replace the traditional convolutional feature map with a self-attention feature map.

After convolution operation, the convolutional feature maps pass three branches $f(x), g(x), h(x)$ of the 1×1 convolution structure, and the size of the feature map is unchanged. $g(x)$ changes the number of channels, and the output of $h(x)$ keeps the number of channels unchanged. H and W represent the length and width of the feature map,

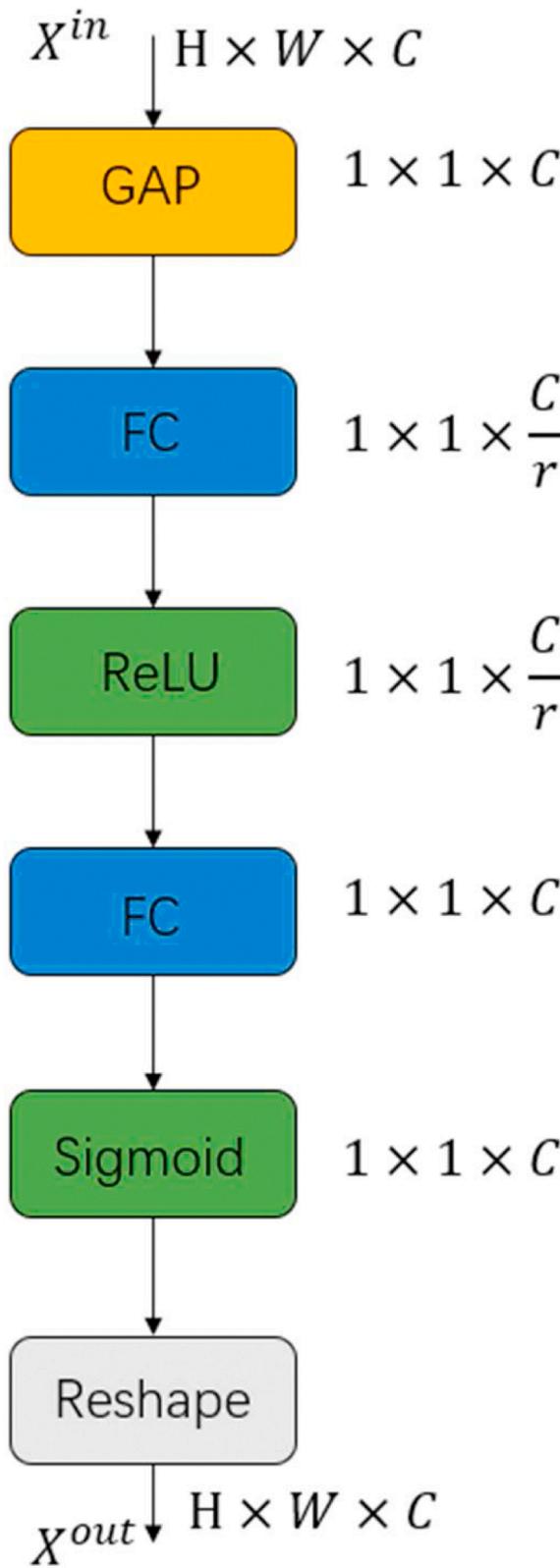


Fig. 4. Channel-Attention module. GAP denotes the global average pooling operation and Reshape denotes enlarge operation.

and C represents the number of channels. After transposing the output of $f(x)$, and multiplying the output matrix of $g(x)$, through normalizeing by softmax to get an $[H^*W, H^*W]$ attention map. By multiplying the attention map with the output of $h(x)$ to get a $[H^*W, C]$ feature map,

and using a 1×1 convolutions to reshape the output to $[H, W, C]$ to get the feature map at this time. The structure of the self-attention module is shown in Fig. 5.

$$s_{ij} = f(x_i)^T g(x_j) \quad (8)$$

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^N \exp(s_{ij})} \quad (9)$$

where $\beta_{j,i}$ indicates the extent to which the model attends to the i^{th} location when synthesizing the j^{th} region. Here, C is the number of channels and N is the number of feature locations of features from the previous hidden layer.

The output of the attention layer is o and can be expressed as:

$$o_j = v(\sum_{i=1}^N \beta_{j,i} h(x_i)), h(x_i) = W_h x_i, v(x_i) = W_v x_i \quad (10)$$

In the above formulation, w_g , w_f , w_h , and w_v are the learned weight matrices, which are implemented as 1×1 convolutions.

Besides, we further multiply the output of the attention layer by a scale parameter and add back the input feature map. Therefore, the final output is given by,

$$y_{SA} = \gamma o_i + x_i \quad (11)$$

where γ is a learnable scalar and initialized to 0. Introducing learnable γ can make the network first rely on the information of the local neighborhood, and then gradually learn to assign more weight to non-local information.

(3) Fusion Operation

A. Direct Connection.

The direct connection function can be implemented by adding the two terms directly as following:

$$f(\alpha R_i^0, \beta Y_i) = \alpha R_i^0 + \beta Y_i \quad (12)$$

where i is the index of a feature. R represents the output of Channel Attention, and Y represents the output of Self-attention. Both α and β are set to 0.5 as the preset value.

B. Weighted Connection.

Compared to the direct connection, the weighted connection introduces the competition between R and Y . Besides, it can be easily extended to a softmax form, which is more robust and less sensitive to trivial features. Both α and β are set to 0.5 as the preset value. To avoid introducing extra parameters, we calculate weights using R and Y . The weighted connection function is represented as

$$f(\alpha R_i^0, \beta Y_i) = \frac{|\alpha R_i^0|^2}{\alpha R_i^0 + \beta Y_i} + \frac{|\beta Y_i|^2}{\alpha R_i^0 + \beta Y_i} \quad (13)$$

(4) Loss Function

The loss function is used to estimate the difference between the value generated or fitted by the model and the real value, that is, the difference between the reconstructed MRI and the original MRI. The smaller the loss function, the stronger the model is. In order to improve the quality of model reconstruction, we propose to use perceptual loss, pixel loss, and adversarial loss as the combined loss function of the generator. Perceptual loss mimics human visual differences, and pixel loss is the difference between pixels in the image domain.

$$l_x^{SR} = l_x^{SR} + 10^{-3} l_{Gen}^{SR} \quad (14)$$

In the following we describe possible choices for the content loss l_x^{SR} and the adversarial loss l_{Gen}^{SR} .

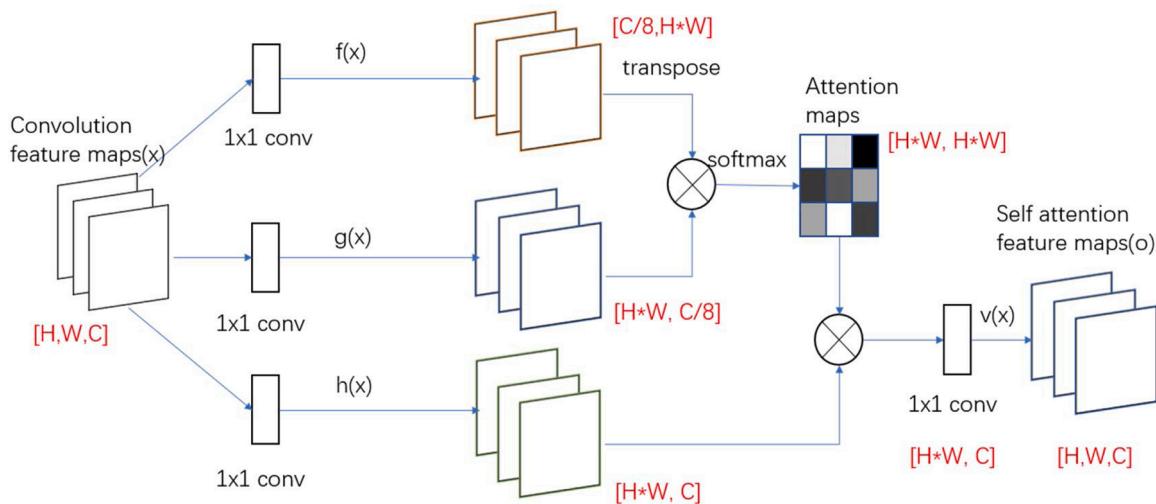


Fig. 5. Self-Attention module.

This paper uses the Euclidean distance between VGG features, which is more relevant to human perception, as the content loss, as shown below:

$$l_x^{SR} = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} (\phi_{ij}(I^{HR})_{x,y} - \phi_{ij}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (15)$$

ϕ_{ij} indicates that the extracted feature is the j -th convolutional layer before the i -th largest pooling layer. W_{ij}, H_{ij} represents the dimension of the feature layer.

The adversarial loss function is the average discriminator probability value of the samples generated by the generator. The formula is as follows:

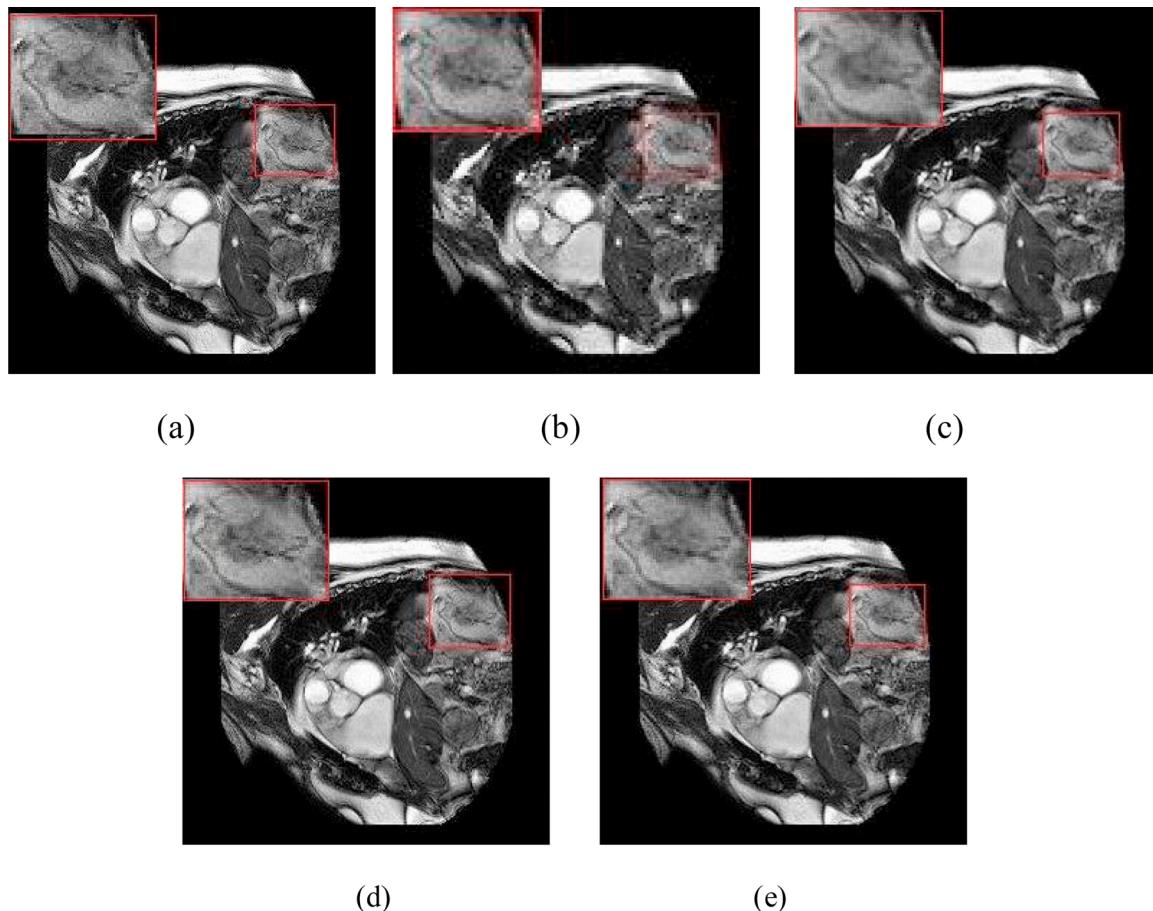


Fig. 6. The reconstructed super-resolution cardiac MR images by using different GAN based methods(4×). (a) the real high resolution MR image, and the reconstructed super-resolution MRI using (b)SR-GAN(PSNR 32.37,SSIM 0.9948), (c)SA-SR-GAN(PSNR 34.07,SSIM 0.9959), (d)CA-SR-GAN(PSNR 34.12,SSIM 0.9963),(e)FA-GAN(PSNR 34.28,SSIM 0.9966).

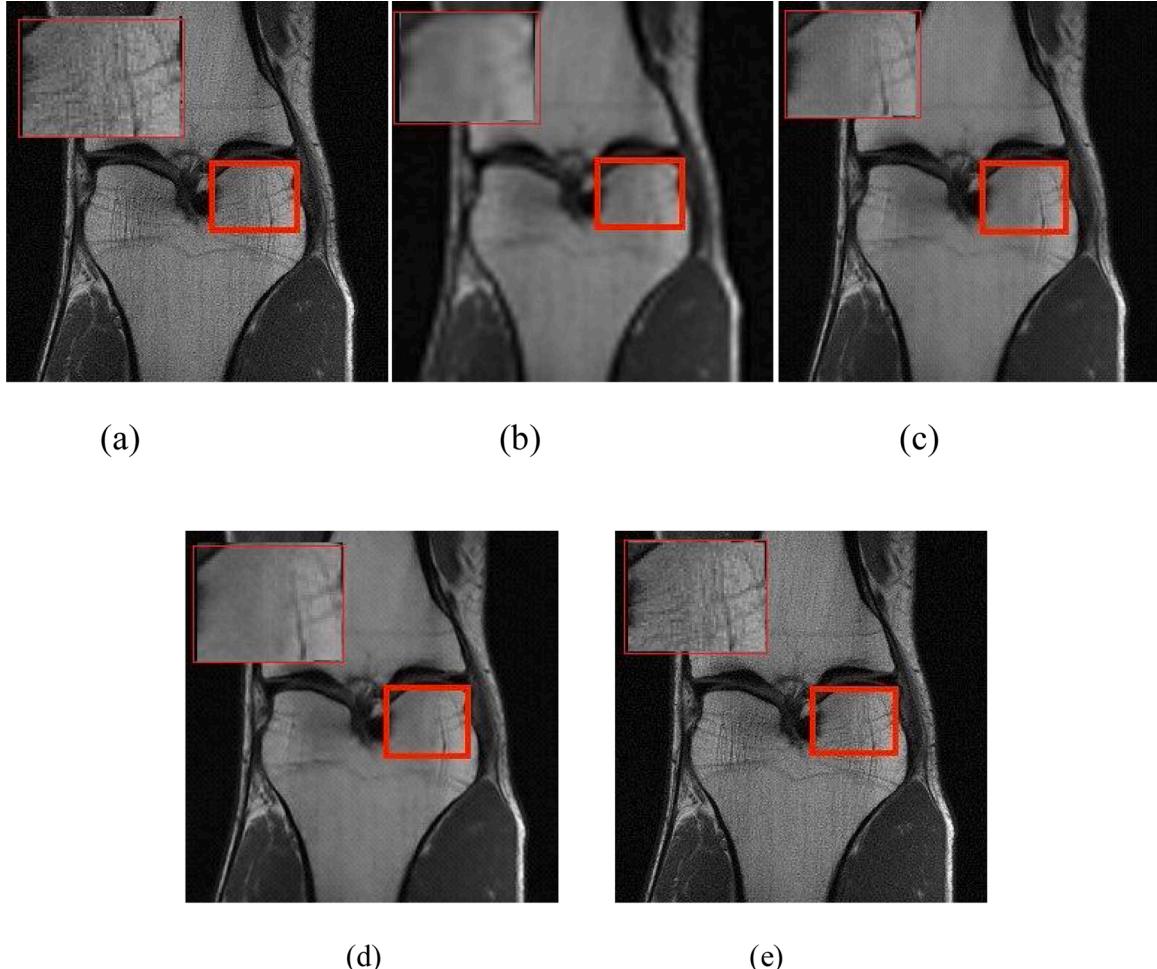


Fig. 7. The reconstructed super-resolution knee MR images by using different GAN-based methods(4×). (a) the real high-resolution MR image, and the reconstructed super-resolution MRI using (b)SR-GAN(PSNR 28.90 SSIM 0.9905), (c)SA-SR-GAN(PSNR 30.10 SSIM 0.9912), (d)CA-SR-GAN(PSNR 30.05 SSIM 0.9910). (e)FA-GAN (PSNR 30.28 SSIM 0.9926).

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (16)$$

$D_{\theta_D}(G_{\theta_G}(I^{LR}))$ represents the probability that the discriminator judges the image generated by the generator as the original magnetic resonance image.

3. Experimental results

3.1. Datasets and metrics

All the experiment use TeslaV100-SXM2 GPU and four different MRI data sets to train and test the model. Randomly select 50 samples with 3D formats for training, of which 40 samples are used as the training set (3200 2D MRI) and 10 samples are used as validation set (960 2D MRI). All low-resolution images in the experiment are obtained by bicubic interpolation. At the same time, to ensure the fairness of the test, we conducted two independent tests to verify the performance of the proposed FA-GAN model. The first test experiment was to randomly select 10 samples for the test set (960 two-dimensional MRI images) to test and calculate the average quantitative index, and the second one was to select a two-dimensional MRI with obvious features from the testing sets. The optimizing procedure is implemented by using Adam optimization algorithm with 0.9. The FA-GAN networks were trained with a learning rate of 0.0001. The model training takes 10 h at a time.

The experiment uses three evaluation criteria to evaluate the reconstructed image: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and freshet Inception Distance score(FID). The definition of PSNR is

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{M \times N \sum_{i=1}^M \sum_{j=1}^N (y(i,j) - x(i,j))^2}\right) \quad (17)$$

where x represents the original image, y represents the super-resolution reconstructed image, and i, j respectively represent the coordinate position of the pixel, and M, N represent the size of the image. The SSIM can be defined by

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (18)$$

where μ_x and μ_y represent the mean of the image x and y respectively, σ_x and σ_y represent the variance of the image x and y respectively, the σ_{xy} covariance of the image x and y , C_1 and C_2 the constant value used to maintain stability. The expression of FID is

$$FID(x_r, g) = \left\| \mu_x - \mu_g \right\|_2^2 + Tr(\sum x_r + \sum g - 2(\sum x_r \sum g)^{\frac{1}{2}}) \quad (19)$$

In the formula, Tr represents the sum of the elements on the diagonal

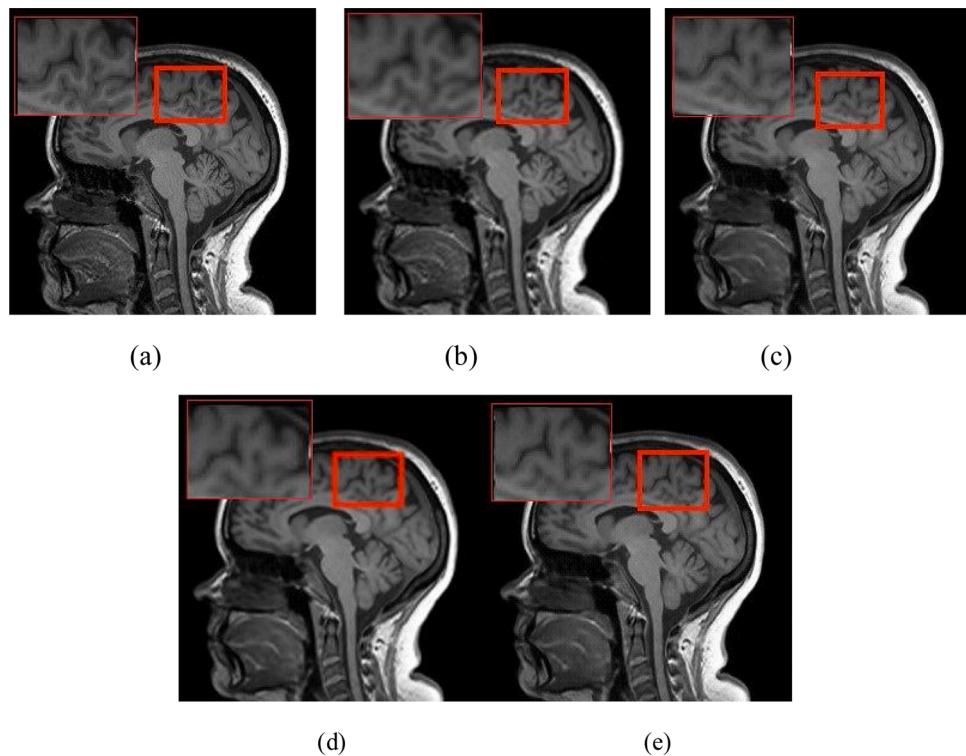


Fig. 8. The reconstructed super-resolution brain MR images by using different GAN-based methods(4×). (a) the real high resolution MR image, and the reconstructed super-resolution MRI using (b)SR-GAN(PSNR 39.90 SSIM 0.9957), (c)SA-SR-GAN(PSNR 41.78 SSIM 0.9969), (d)CA-SR-GAN(PSNR 40.68 SSIM 0.9960), (e)FA-GAN (PSNR 42.07 SSIM 0.9974).

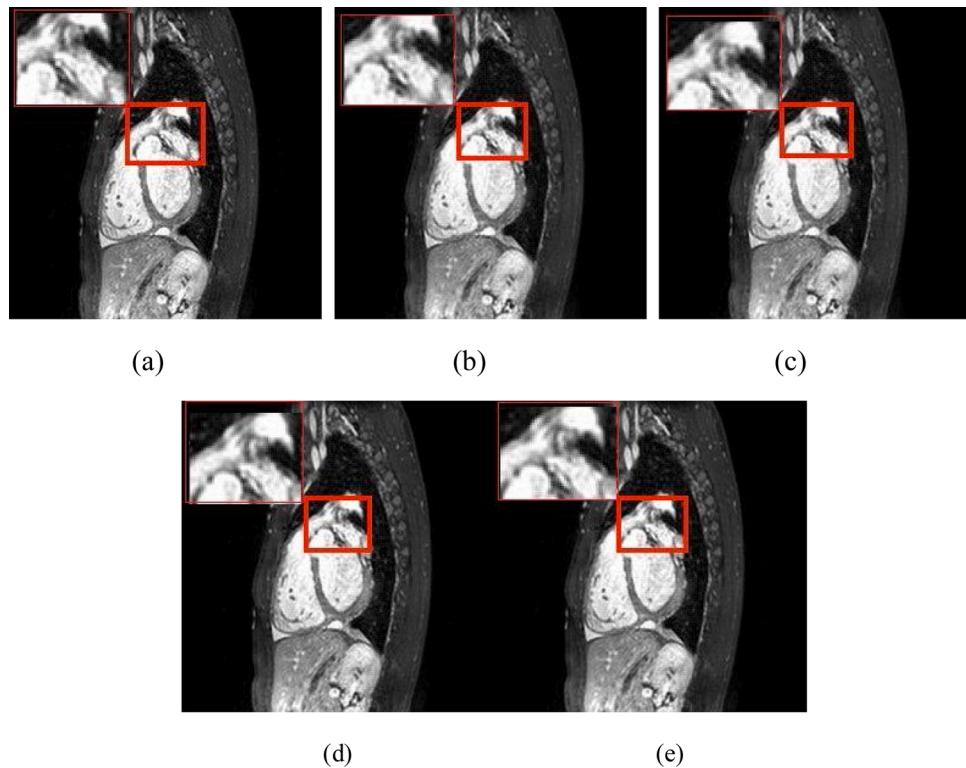


Fig. 9. The reconstructed super-resolution MR images by using different GAN based methods on MM-WHS(4×). (a) the real high resolution MR image, and the reconstructed super-resolution MRI using (b)SR-GAN(PSNR 34.42 SSIM 0.9937), (c)SA-SR-GAN(PSNR 35.93 SSIM 0.9966), (d)CA-SR-GAN(PSNR 36.59 SSIM 0.9974), (e)FA-GAN(PSNR 36.77 SSIM 0.9980).

Table 1

Average PSNR, SSIM by using different GAN-based methods (cardiac).

method	SRGAN		CA-SR-GAN		SA-SR-GAN		FA-GAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
X2	33.73 ± 3.75	0.9957 ± 0.0032	34.22 ± 3.48	0.9959 ± 0.0028	35.66 ± 3.23	0.9962 ± 0.0026	38.58 ± 3.15	0.9979 ± 0.0023
X4	32.30 ± 4.26	0.9946 ± 0.0067	33.75 ± 4.79	0.9953 ± 0.0073	34.09 ± 4.42	0.9959 ± 0.0052	34.26 ± 4.35	0.9965 ± 0.0058
X8	31.35 ± 4.31	0.9951 ± 0.0058	32.89 ± 4.73	0.9952 ± 0.0078	33.45 ± 4.13	0.9956 ± 0.0031	33.70 ± 4.20	0.9960 ± 0.0023

Table 2

Average PSNR, SSIM by using different GAN-based methods(brain).

method	SRGAN		CA-SR-GAN		SA-SR-GAN		FA-GAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	43.35 ± 5.46	0.9989 ± 0.0046	43.56 ± 5.28	0.9991 ± 0.0042	43.97 ± 5.79	0.9991 ± 0.0037	44.11 ± 5.11	0.9992 ± 0.0039
×4	39.88 ± 5.86	0.9955 ± 0.0068	40.56 ± 5.12	0.9963 ± 0.0082	41.76 ± 4.36	0.9970 ± 0.0055	42.11 ± 4.12	0.9974 ± 0.0062
×8	30.83 ± 5.41	0.8891 ± 0.0071	30.96 ± 5.67	0.8902 ± 0.0093	31.02 ± 5.32	0.8907 ± 0.0084	31.27 ± 5.36	0.8935 ± 0.0059

Table 3

Average PSNR, SSIM by using different GAN-based methods (knee).

method	SRGAN		CA-SR-GAN		SA-SR-GAN		FA-GAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	34.21 ± 6.51	0.9963 ± 0.0037	35.28 ± 5.76	0.9964 ± 0.0031	36.28 ± 5.69	0.9966 ± 0.0042	36.58 ± 5.17	0.9969 ± 0.0033
×4	28.89 ± 5.81	0.9903 ± 0.0057	29.75 ± 5.72	0.9908 ± 0.0066	30.13 ± 5.42	0.9913 ± 0.0072	30.27 ± 5.61	0.9926 ± 0.0056
×8	26.31 ± 5.29	0.9722 ± 0.0089	27.46 ± 5.10	0.9813 ± 0.0073	28.22 ± 4.52	0.9842 ± 0.0075	28.83 ± 4.36	0.9877 ± 0.0063

Table 4

Average PSNR, SSIM by using different GAN-based methods (MM-WHS).

method	SRGAN		CA-SR-GAN		SA-SR-GAN		FA-GAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	36.27 ± 3.27	0.9967 ± 0.0032	38.93 ± 4.16	0.9981 ± 0.0041	39.68 ± 4.09	0.9989 ± 0.0037	39.78 ± 3.89	0.9992 ± 0.0034
×4	34.39 ± 5.23	0.9933 ± 0.0031	35.94 ± 5.49	0.9968 ± 0.0052	36.63 ± 5.43	0.9973 ± 0.0047	36.81 ± 5.06	0.9981 ± 0.0027
×8	27.44 ± 4.28	0.9722 ± 0.0046	27.89 ± 4.36	0.9842 ± 0.0039	28.43 ± 4.49	0.9863 ± 0.0053	28.92 ± 4.21	0.9879 ± 0.0042

of the matrix, μ is the mean, Σ is the covariance, x_r represents a real image, and g is a generated image.

3.2. Experimental results

In the experiment, we set the parameters of the comparison experiment with the optimal parameters in order to compare the best reconstruction performance. Figs. 6–8 show the reconstructed two-dimensional super resolution MR image by using different GAN-based algorithms. Due to the small difference in visual observation, we choose a two-dimensional MRI with more prominent features and zoom in on a specific area. From Figs. 6–9, when compared with other three GAN-based methods, it can be found that the FA-GAN based reconstruction algorithm has a clearer texture structure in detail than the other methods. Due to the combination of the self-attention module and channel attention module, with a richer high-frequency texture detail under a large-scale factor, the proposed FA-GAN method preserves more detailed structural information and the fine outline of the MR image. The reconstructed image has clear texture details and most of the aliasing artifacts are effectively suppressed even with $4\times$ super resolution.

Tables 1–4 show the average values of two quantified indicators of PSNR and SSIM of 3200 two-dimensional MRI reconstructed by using different algorithms. Table 1 presents cardiac super-resolution MRI reconstruction performances by using the differences methods, Table 2 shows the brain super-resolution MRI reconstruction performances of different methods, Table 3 provides the knee super-resolution MRI reconstruction results of different methods, and Table 4 shows the MM-WHS by using different GAN-based methods. The reconstruction

Table 5Average FID under different methods ($4\times$).

method	SRGAN	CA-SR-GAN	SA-SR-GAN	FA-GAN
FID				
cardiac	35.78 ± 2.58	26.65 ± 2.49	24.23 ± 2.32	18.97 ± 2.13
brain	20.38 ± 3.06	17.22 ± 3.19	16.57 ± 2.86	12.43 ± 2.36
knee	43.87 ± 4.21	39.89 ± 4.34	38.43 ± 3.71	33.59 ± 3.79
MMWHS	35.32 ± 3.47	34.77 ± 3.59	32.55 ± 2.96	28.15 ± 2.98

performances of these GAN-based methods are listed in terms of the PSNR and SSIM values at three magnifications of 2, 4, and 8 times. As shown in Tables 1–4, it can be found that the proposed FA-GAN method achieve the highest PSNR and SSIM among these four GAN-based reconstruction methods, and the following are SA-SR-GAN, CA-SR-GAN and SRGAN. From the tables, the FA-GAN method can improve the average PSNR of the reconstructed images by about 0.44–4.85 and SSIM by about 0.0003–0.0044, especially in the cardiac MR image with $2\times$ super-resolution reconstruction. The proposed FA-GAN can improve the reconstruction performances obviously in terms of the PSNR and SSIM values.

Table 5 illustrates the super-resolution MRI reconstruction performances by using different GAN-based methods in terms of FID. As shown in Table 5, it can be found that the FA-GAN can effectively reduce the FID. A lower FID means that the reconstructed SR MR images are closer to the real high resolution MR images, which means that the quality of the reconstructed SR MR images is higher.

Table 6
Ablation studies on MICCAI 2013 grand challenge public data set.

Model		×2	×4	×8
FA-GAN	PSNR	38.58	34.26	33.70
	SSIM	0.9979	0.9965	0.9960
-SA	PSNR	37.99	34.29	33.59
	SSIM	0.9972	0.9966	0.9954
-CA	PSNR	37.84	34.02	33.42
	SSIM	0.9970	0.9962	0.9951
-LFFB	PSNR	38.04	34.13	33.65
	SSIM	0.9973	0.9959	0.9957
-SA-CA	PSNR	37.45	33.70	33.47
	SSIM	0.9968	0.9957	0.9952
-SA-LFFB	PSNR	37.27	33.65	32.89
	SSIM	0.9963	0.9956	0.9943
-CA-LFFB	PSNR	37.03	33.44	32.12
	SSIM	0.9960	0.9954	0.9938

Table 7
Average PSNR,SSIM under different connections(4×).

method	Direct Connection		Weighted Connection	
	PSNR	SSIM	PSNR	SSIM
cardiac	33.98 ± 4.54	0.9962 ± 0.0074	34.26 ± 4.35	0.9965 ± 0.0058
	41.70 ± 4.89	0.9969 ± 0.0059	42.11 ± 4.12	0.9974 ± 0.0062
knee	29.89 ± 5.45	0.9924 ± 0.0063	30.27 ± 5.61	0.9926 ± 0.0056
	35.67 ± 5.99	0.9975 ± 0.0032	36.81 ± 5.06	0.9981 ± 0.0027

Table 8
Average PSNR,SSIM under different values of the parameters(4×).

	$\alpha = 0.4, \beta = 0.6$		$\alpha = 0.5, \beta = 0.5$		$\alpha = 0.6, \beta = 0.4$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
cardiac	33.92 ± 4.37	0.9961 ± 0.0088	34.26 ± 4.35	0.9965 ± 0.0058	33.10 ± 4.91	0.9961 ± 0.0079
	41.88 ± 4.91	0.9966 ± 0.0067	42.11 ± 4.12	0.9974 ± 0.0062	42.02 ± 4.96	0.9970 ± 0.0082
knee	29.73 ± 6.25	0.9922 ± 0.0089	30.27 ± 5.61	0.9926 ± 0.0056	30.13 ± 5.84	0.9920 ± 0.0077

4. Discussion

To demonstrate the effect of each component, we carry out seven ablation experiments of local feature fusion block (LFFB), channel

attention (CA), and self-attention (SA). By removing the local features fusion block, our model falls back to a network similar to SRGAN but with the attention block. The results confirm that making full use of local features fusion block will significantly improve performance. One possible reason is that fusing hierarchical features improves the information flow and eases the difficulty of training. We can conclude from Table 6 that the proposed FA-GAN model with all components achieves the best performance. The integration of local feature fusion block and global feature fusion block not only improves 1–2 dB on PSNR, but also gets much better visual effects in image details than the other methods with part components, as shown in Figs. 6–9.

According to the results of the ablation experiment, as shown in Table 6, it can be seen that the CA and LFFB modules together play the most important role in the super resolution MR image reconstruction, which affect the reconstruction performances obviously. However, the affection of the SA module is relative small, and the reconstruction quality drops slightly. Table 7 illustrates the reconstruction effect under different connection modes. It can be clearly seen that the weighted connection has achieved better results. Thus, we use a weighted connection in our method.

For the selection of parameters α and β , we have done the following three sets of comparative experiments. As shown in Table 8, the experimental results show that the parameters are the optimize values when $\alpha = 0.5$ and $\beta = 0.5$.

In this paper, the spectral normalization (SN) is introduced to the discriminator network, so as to stabilize the training of GAN and limit the Lipschitz constant of the discriminator. Compared with other normalization techniques, spectral normalization does not require additional hyper parameter adjustments (set the spectral norm of all weight layers to 1). Fig. 10 shows the effect of SN on FA-GAN, which makes the loss value steadily drop and makes the whole training process more stable.

Fig. 11 illustrates the loss value of the training process of SR image reconstruction by using four different GAN-based methods with $\times 4$ times. It can be found that the loss value by using FA-GAN method decreases monotonously with iteration increasing, while the other methods decrease in waves, which indicates that the proposed FA-GAN method combined with spectral normalization makes the training more stable.

5. Conclusion

This paper proposed a new method for super-resolution magnetic resonance images reconstruction by using fusion attention based generative adversarial networks (FA-GAN). Two different attention mechanisms are integrated into the SRGAN framework to obtain

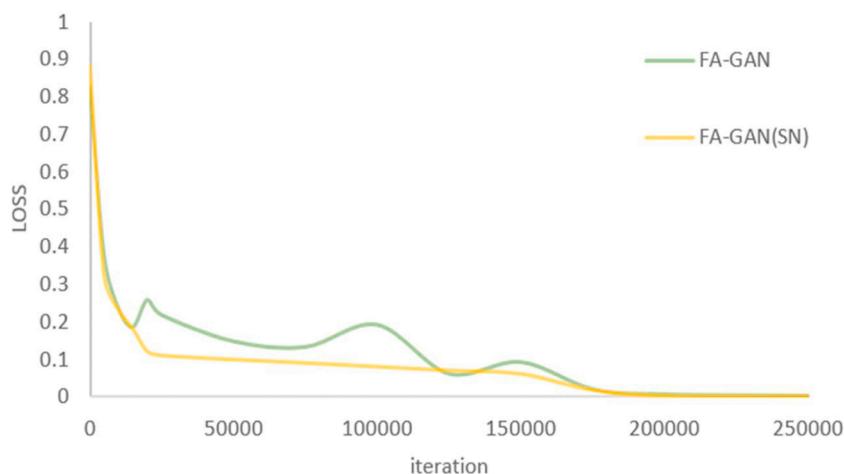


Fig. 10. Comparison of loss value of FA-GAN with or without SN.

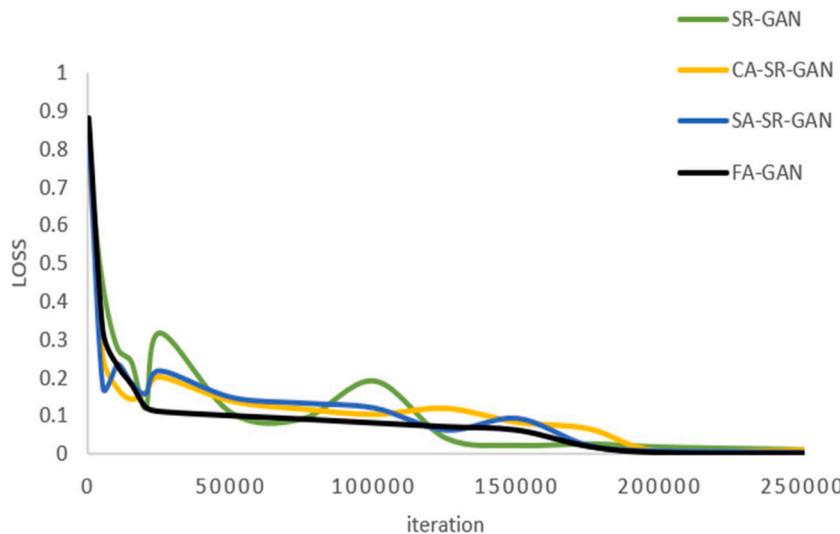


Fig. 11. Comparison of loss under factor 4 and different methods.

important features. Compared with the SRGAN framework, the proposed FA-GAN method can reconstruct super-resolution images with higher PSNR, SSIM and lower FID, and the reconstructed SR images preserve much closer image details to the real high-resolution image. In the future work, the proposed FA-GAN method can be used to reconstruct the super resolution MR images like 7 T resolution from 3 T MR equipment, which can improve the resolution of the MR image without change the hardware.

Author contributions

MJ, MZ, LW, XY, JZ, YL, PW, JH and GY conceived and designed the study, contributed to data analysis, contributed to data interpretation, and contributed to the writing of the report. MJ, JH and GY contributed to the literature search. MJ, MZ, LW, XY, JZ, YL, and PW contributed to data collection. MJ, MZ, LW, XY, JZ, YL, and PW performed data curation and contributed to the tables and figures. All authors contributed to the article and approved the submitted version.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (61672466, 62011530130 and 61771080), in part by Joint Fund of Zhejiang Provincial Natural Science Foundation (LSZ19F010001), in part by the Key Research and Development Program of Zhejiang Province (2020C03060), in part by the 521 Talents project of Zhejiang Sci-Tech University, in part by the British Heart Foundation (Project Number: TG/18/5/34111, PG/16/78/32402), in part by the Hangzhou Economic and Technological Development Area Strategical Grant (Imperial Institute of Advanced Technology), in part by the European Research Council Innovative Medicines Initiative on Development of Therapeutics and Diagnostics Combatting Coronavirus Infections Award DRAGON: rapiD and secuRe AI imaging based diaGnosis, stratification, follow-up, and preparedness for coronavirus pandemics [H2020-JTI-IMI2 101005122], in part by the AI for Health Imaging Award CHAMELEON: Accelerating the Lab to Market Transition of AI Tools for Cancer Management [H2020-SC1-FA-DTS-2019-1 952172], and in part by the UK Research and Innovation (MR/V023799/1).

References

- Armanious, K., Gatidis, S., Nikolaou, K., et al., 2019. Retrospective correction of rigid and Non-rigid MR motion artifacts using GANs. In: Proceedings of the IEEE 16th International Symposium on Biomedical Imaging. Venezia, pp. 1550–1554.
- Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q.V., 2019. Attention augmented convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, pp. 3286–3295.
- Dong, C., Loy, C.C., He, K., et al., 2014a. Learning a deep convolutional network for image super-resolution. In: Proceedings of the European Conference on Computer Vision. Switzerland, pp. 184–199.
- Dong, C., Loy, C., He, K., et al., 2014b. Learning a deep convolutional network for image super-resolution. In: European Conference on Computer Vision. Cham, pp. 184–199.
- Dong, C., Loy, C.C., He, K., Tang, X., 2016. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 295–307.
- Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, pp. 3146–3154.
- Gholipour, A., Estroff, J.A., Warfield, S.K., 2010. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. *IEEE Trans. Med. Imaging* 29, 1739–1758.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al., 2014. Generative adversarial nets. In: Proceedings of the Neural Information Processing Systems. Montreal, pp. 2672–2680.
- Häggström, I., Schmidlein, C.R., Campanella, G., 2019. Fuchs DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. *Med. Image Anal.* 54, 253–262.
- Hu, Z., Xue, H., Zhang, Q., Gao, J., et al., 2020. DPIR-Net: direct PET image reconstruction based on the Wasserstein generative adversarial network. *IEEE Trans. Radiat. Plasma Med. Sci.* 5 (1), 35–43.
- Hu, X., Wang, H., Luo, Y., et al., 2021. Single MR image super-resolution via mixed self-similarity attention network. *Proc. SPIE* 11596, Medical Imaging 2021: Image Processing, 1159605. <https://doi.org/10.1117/12.2580860>.
- Huang, G., Liu, Z., Van Der Maaten, L., et al., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, pp. 4700–4708.
- Huang, W., Jia, S., Ke, Z., et al., 2021. SRR-Net: A Super-Resolution-Involved Reconstruction Method for High Resolution MR Imaging. *arXiv preprint arXiv: 2104.05901*.
- Iandola, F.N., Moskewicz, M.W., Ashraf, K., et al., 2016. Firecaffe: near-linear acceleration of deep neural network training on compute clusters. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, pp. 2592–2600.
- Jafari-Khouzani, K., 2014. MRI upsampling using feature-based nonlocal means approach. *IEEE Trans. Med. Imaging* 33, 1969–1985.
- Jog, A., Carass, A., Prince, J.L., 2014. Improving magnetic resonance resolution with supervised learning. In: Proceedings of the 11th International Symposium on Biomedical Imaging. Beijing, pp. 987–990.
- Latif, S., Asim, M., Usman, M., et al., 2018. Automating Motion Correction in Multishot MRI Using Generative Adversarial Networks. *ArXiv: 1811.09750*.
- Ledig, C., Theis, L., Huszár, F., et al., 2017a. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 4681–4690.
- Ledig, C., Theis, L., Huszár, F., et al., 2017b. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, pp. 4681–4690.

- Li, L., Xu, M., Liu, H., et al., 2019. A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Trans. Med. Imaging* 39 (2), 413–424.
- Liu, H., Xu, J., Wu, Y., et al., 2018. Learning deconvolutional deep neural network for high resolution medical image reconstruction. *Infor. Sci.* 468, 142–154.
- Liu, Z., Chen, H., Liu, H., 2019. Deep learning based framework for direct reconstruction of PET images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham.
- Luo, J., Mou, Z., Qin, B., et al., 2017. Fast single image super-resolution using estimated low-frequency k-space data in MRI. *Magn. Reson. Imaging* 40, 1–11.
- Mahmoudzadeh, A.P., Kashou, N.H., 2014. Interpolation-based super-resolution reconstruction: effects of slice thickness. *Med Imaging* 1, 034007.
- Manjon, J.V., Coupe, P., Buades, A., Fonov, V., Collins, D.L., Robles, M., 2010. Nonlocal MRI upsampling. *Med. Image Anal.* 14, 784–792.
- Miyato, T., Kataoka, T., Koyama, M., et al., 2018. Spectral Normalization for Generative Adversarial Networks. ArXiv: 1802.05957.
- Quan, T.M., Nguyen-Duc, T., Jeong, W.K., 2018. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans. Med. Imaging* 37 (6), 1488–1497.
- Ramachandran, P., Zoph, B., Le, Q.V., 2017. Searching for Activation Functions. ArXiv: 1710.05941.
- Rueda, A., Malpica, N., Romero, E., 2013. Single-image super-resolution of brain MR images using overcomplete dictionaries. *Med. Image Anal.* 17, 113–132.
- Seeliger, K., Güclü, U., Ambrogioni, L., et al., 2018. Generative adversarial networks for reconstructing natural images from brain activity. *Neuro Image* 181, 775–785.
- Shi, W., Caballero, J., Huszár, F., et al., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, pp. 1874–1883.
- Shi, J., Li, Z., Ying, S., et al., 2018a. Image super-resolution via wide residual networks with fixed skip connection. *IEEE EMBS. Int. Conf. Biomed. Health Inform.* 23, 1129–1140.
- Shi, J., Liu, Q., Wang, C., Zhang, Q., Ying, S., Xu, H., 2018b. Super-resolution reconstruction of MR imagewith a novel residual learning network algorithm. *Phys. Med. Biol.* 63, 085011.
- Shi, J., Liu, Q., Wang, C., et al., 2018c. Super-resolution reconstruction of MR image with a novel residual learning network algorithm. *Phys. Med. Biol.* 63 (8), 085011.
- Ting, Z., Xiao, Q.W., 2019. Pyramid feature attention network for saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, pp. 3085–3094.
- Tourbier, S., Bresson, X., Hagmann, P., Thiran, J.-P., Meuli, R., Cuadra, M.B., 2015. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *NeuroImage* 118, 584–597.
- Wang, S., Tan, S., Gao, Y., et al., 2018. Learning joint-sparse codes for Calibration-Free Parallel MR imaging. *IEEE Trans. Med. Imaging* 37 (1), 251–261.
- Wang, S., Ke, Z., Cheng, H., et al., 2019. DIMENSION: dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. *NMR Biomed.*, e4131
- Wang, S., Cheng, H., Ying, L., et al., 2020. DeepcomplexMRI: exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magn. Reson. Imaging* 68, 136–147.
- Xu, C., Xu, L., Gao, Z., et al., 2018. Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. *Med. Image Anal.* 50, 82–94.
- Zeng, W., Peng, J., Wang, S., et al., 2019. A comparative study of CNN-based super-Resolution methods in MRI reconstruction. In: Proceedings of the 16th International Symposium on Biomedical Imaging. Venezia, pp. 1678–1682.
- Zhang, H., Goodfellow, I., Metaxas, D., et al., 2018. Self-attention Generative Adversarial Networks. ArXiv: 1805.08318.
- Zhu, X., Cheng, D., Zhang, Z., Lin, S., Dai, J., 2019. An empirical study of spatial attention mechanisms in deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, pp. 6688–6697.