

Correction of Motion Artifact in MRI with Contextual and Cross-space Fusion

Guo Jiang[♣] Ziniu Zhang[♣] Xiaorong Pu^{♣,†} Junchi Liu[♣]
Yuting Wang[♣] Yazhou Ren[♣]

[♣] School of Computer Science and Engineering,
University of Electronic Science and Technology of China
[♣] Sichuan Provincial People's Hospital

guojiang.cn@outlook.com michaelzhangziniu@gmail.com
puxiaor@uestc.edu.cn

Abstract

Magnetic Resonance Imaging (MRI) is a powerful and effective technique that has been broadly used in the early diagnosis of many diseases. However, it is extremely sensitive to subject motions, which could produce artifacts in the reconstructed images and may lead to misdiagnosis. In this paper, we develop a generative adversarial network (GAN) based network called F(usion)GAN for the correction of motion artifact in MRI. Within FGAN, there are two approaches of information fusion to improve the correction performance: 1) generators for k-space and real-space data are in a cascaded structure and are jointly optimized to fuse cross-space information, and 2) stacked U-Nets with self-assisted priors in real-space are adapted to fuse contextual information between adjacent slices in MRI images. Multiple loss functions are utilized to keep the data consistency in the data flow of FGAN. The results of extensive experiments on multiple real MRI datasets prove the feasibility and effectiveness of FGAN in both denoising and texture restoration. On two publicly available datasets, it obtains substantial improvements over the state-of-the-art methods by an average of 10.68% and 9.69% in terms of PSNR, respectively. Source code is available for research purpose.

1. Introduction

Magnetic Resonance Imaging (MRI) is a broadly used medical imaging technique which serves as the main approach to the early diagnosis of many diseases. However, it is extremely sensitive to subject movements. Even slight motions of patients during data acquisition can cause significant degrading of the image quality. Motion artifacts typically manifest as blurring or ghosting. As shown in Figure 1, motion artifacts dramatically affect the recognition of de-

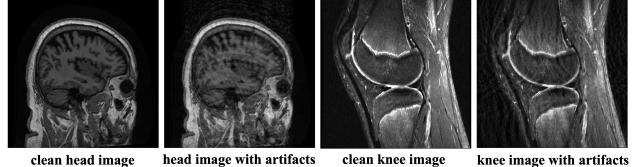


Figure 1. The clear contrast between clean image and image with motion artifacts

tails in organs based on human observation and may lead to misdiagnosis.

Many approaches have been proposed to correct motion artifacts. Generally, there are three main ways: restricting patient's motions, prior correction and backtracking correction. Restricting patient's motion is the most direct method to prevent motion artifacts. But the uncomfortable and even annoying feeling of restriction remains a problem. Moreover, autonomous motions such as heartbeats and breath can't be avoided. In prior correction of motion artifacts, artificial priors like specific scan adjustments for each organ would assist devices to improve the acquisition quality [12, 22]. The sampling space remains relatively static with the patient, so the motion artifacts can be prevented. But this approach requires longer data acquisition time and may instead lead to more motion artifacts. Backtracking correction of motion artifacts is about removing artifacts from images that are already damaged. Since the features of MRI images can be high-dimensional, traditional machine learning algorithms can not fully learn the characteristics of the entire image and may fail to achieve satisfying performance. Therefore, deep learning for artifacts correction has recently attracted an increasing interest.

Achieving outstanding results in image processing, deep learning based models have also outperformed traditional artifact correction methods. However, most of them consider MRI images as normal natural images and have ignored the special properties of MRI data. Such methods do not make full use of the inherent information of MRI im-

[†] Corresponding Author.

ages and have some hidden drawbacks like poor data consistency between k-space and real-space. The corrected image may generally look close to the artifact-free ground-truth, but the k-space data obtained from it may have clear difference with that obtained from the ground-truth image. Furthermore, MRI images are produced as ordered series. Adjacent slices are likely to have shared structural information. Hence, additional knowledge priors like undistorted texture details can be obtained from adjacent slices of the target image.

In this study, we introduce the k-space generator and stacked U-Nets into common GAN-based image restoration network. The cascaded generators perform artifact correction in both k-space and real-space. A pair of cross-space loss functions are developed to preserve the data consistency between both spaces. The stacked U-Nets exploit the contextual information extracted from adjacent images to fix the detail loss. Experiments on multiple real MRI datasets demonstrate the superiority of the proposed FGAN compared to the state-of-the-art methods in terms of PSNR, SSIM and visual analysis.

Our contributions can be summarized as follows:

- We develop a GAN-based network with cascaded generators to preserve the data consistency between k-space and real-space in the output image.
- FGAN combines cross-space fusion and contextual fusion to improve the performance in artifacts correction with a concise structure and clear data flows.
- Extensive experiments on multiple real MRI datasets demonstrate the feasibility and effectiveness of FGAN in denoising and detail restoration.

2. Related Works

2.1. Generative Adversarial Networks

Generative adversarial network (GAN) is a kind of generative model introduced by Goodfellow *et al.* [9]. Differing from other generative models, GANs don't rely on Markov chains and generate an implicit distribution that is approximate to the desired transformation or generation. There is a generator and a discriminator within GANs as the opposing sides of a ping-pong game. The generator generates fake data distribution that is approximate to the ground-truth and aims to make it impossible for the discriminator to distinguish the sampled generated fake data from real data. The discriminator is normally a binary classifier. The main purpose of discriminator is to identify whether the input is generated or real. The purpose function of GANs can be formulated as:

$$\min_G \max_D F(D, G) = \mathbb{E}_{x \sim \mathbb{P}_r} [\log D(x)] + \mathbb{E}_{y \sim \mathbb{P}_g} [1 - \log D(y)] \quad (1)$$

Where G and D stand for the generator and the discriminator, respectively. $\mathbb{P}_r(x)$ is the distribution of real data and $\mathbb{P}_g(y)$ is the distribution of data generated by the generator.

During training, the generator and the discriminator are alternatively updated. Original GAN is unstable in training and may generate nonsensical results. So Radford *et al.* [19] propose a family of architectures called DCGAN. They use convolutions with strides to replace spatial pooling functions in the discriminator. They also eliminate fully connected layers, which reduce the number of parameters. Batch normalization is applied to avoid gradient loss and mode collapse. However, Arjovsky *et al.* [2] think DCGAN is still unable to solve Mode Collapse thoroughly. They find that the main cause of Mode Collapse is that traditional GANs use Jensen-Shannon Divergence to approximate the distance between $\mathbb{P}_r(x)$ and $\mathbb{P}_g(y)$. Therefore, Earth-Mover Distance is instead adapted and leads to a much more stable and variational output.

2.2. Deep Learning for Image Restoration

Due to the complexity of visual data, many deep learning approaches have been proposed to perform image restoration like denoising, deblurring and artifact correction.

Isola *et al.* [14] introduce a conditional adversarial network named Pix2pix as a general-purpose solution to image-to-image translation problems like image restoration. With Pix2pix, it is no longer a need to hand-engineer the mapping and loss functions to achieve satisfying result. Zhu *et al.* [25] propose a GAN-based network named CycleGAN to perform Image-to-image translation. It introduces cycle consistency loss to push the optimization of mapping in the absence of paired examples. Kupyn *et al.* [17] introduce an end-to-end learned method named DeblurGAN for motion deblurring. Based on conditional GAN and content loss, it achieves state-of-the-art performance with much faster processing speed. Chen *et al.* [6] leverage a variational autoencoder (VAE) to learn the noise distribution in low-dose CT (LDCT) and then construct abundant pairs of induced-LDCT and normal-dose CT (NDCT) images with the learned noise distribution and open NDCT datasets. This opens a new gate for research that is hindered by the lack of paired data. Recently, cross domain adaptation has played an important role in medical image restoration tasks [10]. Huang *et al.* [13] alleviate the domain shift problem in cross domain adaptation by extracting the shared semantic information of all domains. This idea of utilizing shared information between domains also inspires the cross-space fusion in FGAN.

Duffy *et al.* [8] develop an approach to simulate how MRI images are damaged with motion artifacts. They then use a convolutional neural network to learn the mapping from damaged image to clean image with paired data they obtained through the artifacts simulation. Oksuz *et*

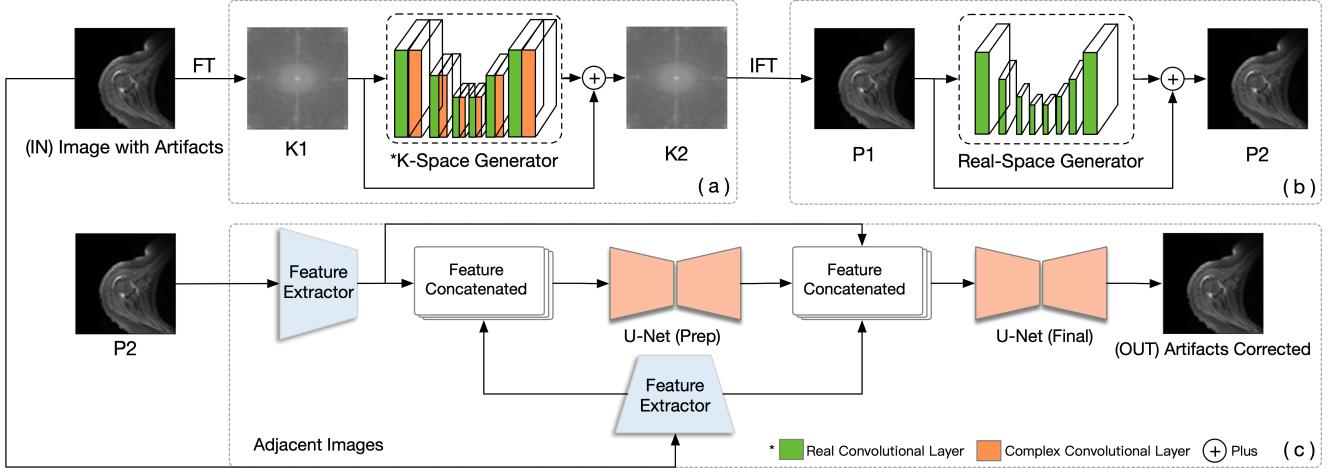


Figure 2. The overall architecture of the proposed model FGAN. It consists of three components: (a) and (b) the cascaded artifacts removal generators in k-space and real-space for cross-space fusion; (c) the stacked U-Nets with feature extraction for contextual fusion.

al. [18] utilize a GAN-based framework to directly correct the k-space data. The final corrected image then can be obtained through Inverse Fourier Transform. Al-Masni *et al.* [1] develop an efficient retrospective deep learning method for artifacts correction with self-assisted priors. Additional knowledge priors are learned from the corrupted images themselves without the need for additional contrast data. Armanious *et al.* [3] propose a GAN-based framework named MedGAN. Its generator consists of several fully convolutional encoder-decoder networks with skip connections. Some non-adversarial losses are leveraged to achieve better performance. Ji [15] considers both k-space and real-space information in MRI images with a GAN-based network. The idea of simultaneous optimization of both space sets a new pattern for artifacts correction. Wang *et al.* [24] investigate the real-valued and complex-valued convolutional neural networks (CNNs) to explore the multi-channel correlations of the k-space data in MRI images.

3. Method

The proposed F(fusion)GAN implements and combines the ideas of cross-space fusion and contextual fusion through three sub-networks in a GAN-based structure. The overall architecture of FGAN is illustrated in Figure 2. In this section, we detail the two types of information fusion and how the three components within FGAN collaborate to fulfill them.

3.1. Cross-space Fusion

The raw output of MRI exists as matrices of complex number that represents the spatial frequency and phase of all scanned areas. Through Inverse Fourier Transform and other pre-processing, the raw output (k-space data) can then

be reconstructed as images (real-space data) used for diagnosis. To preserve the data consistency between k-space and real-space, and make full use of the information of k-space (where the motion artifacts are produced), we introduce the k-space generator into common GAN-based artifacts correction network.

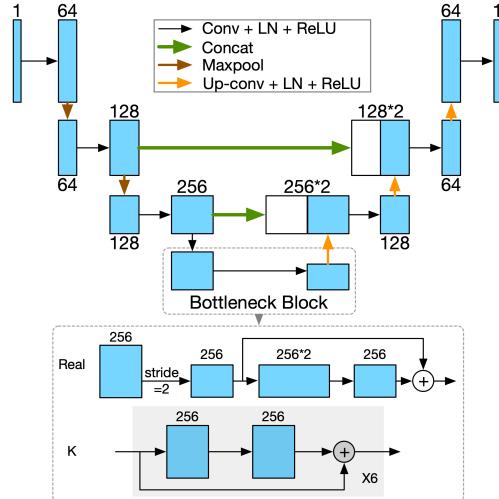


Figure 3. The shared U-Net structure for both generators. The bottleneck block is space-specified. LN: layer norm.

Generators for both spaces are designed as similar U-Net [20] structures. Due to the special format of k-space data, we adapt the complex convolutional layer from DeepcomplexMRI [24] to the k-space generator. Normal complex layers consider the real and imaginary part of a complex-valued matrix as individual layers and perform convolutions separately. This ignores the physical meaning of the complex values and damages the interpretation of mod-

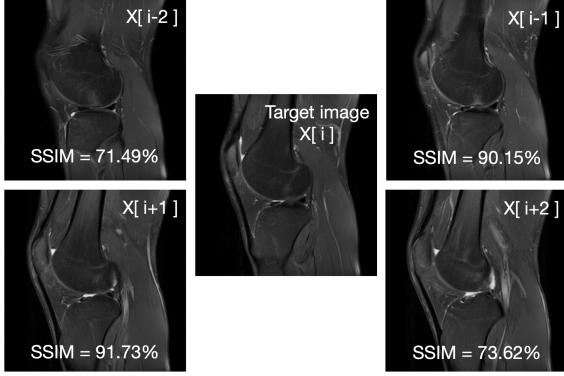


Figure 4. Examples of how adjacent images have shared structure information with target image.

els. In the complex convolutional layer, for a input matrix $X = X_{real} + iX_{imag}$ and a complex convolutional filter $\delta = \delta_{real} + i\delta_{imag}$, where X_{real} , X_{imag} , δ_{real} and δ_{imag} are real-valued matrices, the convolution is performed as:

$$\begin{aligned}\delta * X &= (\delta_{real} + i\delta_{imag}) * X_{real} + iX_{imag} \\ &= (\delta_{real} * X_{real} - \delta_{imag} * X_{imag}) \\ &\quad + i(\delta_{real} * X_{imag} + \delta_{imag} * X_{real})\end{aligned}\quad (2)$$

Therefore, as shown in Figure 2, k-space generator has two different convolutional filters in each layer while there is only real convolutional layers in real-space generator. As shown in Figure 3, traditional skip-connections (copy and crop, feature concatenation) in U-Net are also maintained to fix the information loss in the up-convolution phase. Despite the different calculation of convolutions, the dimension (number of channels) change in both generators is the same except for the bottleneck block.

The data consistency between the two spaces is preserved with a pair of loss functions that push data in both spaces to approximate the artifact-clean ground-truth. Details on the loss functions will be discussed in Section 5.

3.2. Contextual Fusion

With a continuous scan in one run, MRI data is produced as ordered series. Spatially, every image represents a transverse section at a certain height with a small gap between each other. In terms of Structural Similarity (SSIM), as shown in Figure 4, the closest adjacent images (one slice prior or after the target image) can have a relatively high SSIM score of over 90% with the target image, while farther images (two slices prior or after) would have a relatively medium SSIM score of about 70%. Therefore, closely adjacent MRI images have shared auxiliary information of the same organ and are capable of fixing the detail loss in the target image with possibly undistorted information.

Adapted from the work of Al-masni *et al.* [1], we use stacked U-Nets to fuse the contextual information from ad-

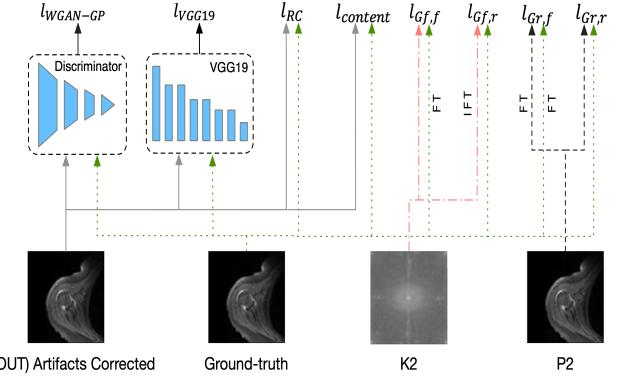


Figure 5. The rough illustration of how all loss functions are calculated. K2 and P2 are the output of k-space and real-space generators.

jacent slices of the same subject. As shown in Figure 2, the output of real-space generator and the adjacent images are processed with a feature extractor. For better fusion performance, there are two U-Nets for the prep stage and the final stage of contextual fusion. Before the prep U-Net, the extracted features are concatenated as a whole input. The output of prep U-Net will then be concatenated with the extracted features earlier for the final U-Net.

Except for the first convolutional layer that deals with different input dimension, both U-Nets have the same structure and layer configurations as the real-space generator. The feature extractor is a small network with one convolutional layer (kernel size: 3×3 , feature map: $1 \rightarrow 64$) followed by Layer Norm and ReLU. It achieves satisfying results without adding up calculation complexity.

3.3. Loss Functions

Multiple loss functions are adapted for optimization goals of each component within FGAN. A rough illustration of all loss functions is shown in Figure 5. The total loss of FGAN can be summarized as:

$$\begin{aligned}Loss &= l_{GAN} + l_{Gf} + l_{Gr} \\ &= (l_{RC} + \lambda_1 l_{WGAN-GP} + \lambda_2 l_{VGG19} + \lambda_3 l_{Content}) \\ &\quad + (\lambda_4 l_{Gf,f} + \lambda_5 l_{Gf,r}) + (\lambda_6 l_{Gr,f} + \lambda_7 l_{Gr,r})\end{aligned}\quad (3)$$

where λ_1 to λ_7 represent the balance coefficients of different losses. l_{GAN} stands for the losses that use the final output of the network as input, while l_{Gf} and l_{Gr} mean the losses that are devoted to the optimization of k-space and real-space generators.

l_{GAN} is composed of 4 loss functions:

1) l_{RC} is the regional consistency loss that reflects the regional similarity between two images. During the processing of artifacts correction, the texture and physiological details may face a certain damage. Here we use a simple expression of regional similarity as an additional loss function

to reduce the regional information loss. It can be formulated as:

$$l_{RC} = \sum_{i=1}^N \|X_i * \Omega_{3,3} - Y_i * \Omega_{3,3}\|_F^2 \quad (4)$$

where N represents the batch size and $\Omega_{3,3}$ stands for a 3×3 matrix with all elements set as 1. X_i and Y_i represent the i -th sample of the damaged MRI images and its corresponding clean ground-truth.

2) $l_{\text{WGAN-GP}}$ is a better replacement for traditional adversarial loss in GAN-based networks. It is adapted from WGAN-GP [11] and can be formulated as:

$$\begin{aligned} l_{\text{WGAN-GP}} &= \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [D(\hat{x})] - \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] \\ &\quad + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \end{aligned} \quad (5)$$

where $x \sim \mathbb{P}_r$ means the sample from the ground-truth data distribution, \mathbb{P}_g is the model output distribution and $\mathbb{P}_{\hat{x}}$ is the distribution sampling uniformly along straight lines between pairs of points sampled from \mathbb{P}_r and \mathbb{P}_g . This loss function is proved to boost the model stability in training.

3) In l_{VGG19} , we replace the original feature extractor of Perceptual Loss [16] with VGG19 [21].

4) l_{Content} is the MSE loss for the final output and ground-truth. We find that the combination of l_{VGG19} (feature-level) and l_{Content} (pixel-level) leads to better detail recovery.

l_{Gf} and l_{Gr} have a shared pair of loss functions $l_{G,f}$ and $l_{G,r}$, which preserves the consistency between two spaces. They are respectively the k-space and the real-space consistency loss for the output of generators in both spaces compared to the ground-truth.

To balance the order of magnitudes between $l_{G,f}$ and $l_{G,r}$, we use L1 Loss as $l_{G,f}$ and MSE Loss as $l_{G,r}$. The ground-truth image and the output of real-space generator will be pre-processed with Fourier Transform for $l_{G,f}$, while the output of k-space generator will be pre-processed with Inverse Fourier Transform for $l_{G,r}$. Both $l_{G,f}$ and $l_{G,r}$ can be formulated as:

$$l_{Gf,f} = \sum_{i=1}^N \left\| \underbrace{G_f(\mathcal{F}(X_i)) - \mathcal{F}(Y_i)}_{\text{output of k-space generator}} \right\|_1 \quad (6)$$

$$l_{Gf,r} = \sum_{i=1}^N \left\| \mathcal{F}'(G_f(\mathcal{F}(X_i))) - Y_i \right\|_F^2 \quad (7)$$

$$l_{Gr,r} = \sum_{i=1}^N \left\| \underbrace{G_r(\mathcal{F}'(G_f(\mathcal{F}(X_i)))) - Y_i}_{\text{output of real-space generator}} \right\|_F^2 \quad (8)$$

$$l_{Gr,f} = \sum_{i=1}^N \left\| \mathcal{F}(G_r(\mathcal{F}'(G_f(\mathcal{F}(X_i))))) - \mathcal{F}(Y_i) \right\|_1 \quad (9)$$

where N stands for the batch size. G_f and G_r are respectively the generators in k-space and real-space. \mathcal{F} and \mathcal{F}'

represent Fourier Transform and Inverse Fourier Transform. X_i and Y_i stand for the i -th sample of the damaged MRI images and its corresponding clean ground-truth.

4. Experiment

Since it is difficult to acquire paired data clinically, simulation of MRI motion artifacts is necessary to fulfill the training demand of all methods. The details of artifact simulation will be discussed in Section 4.1. Besides the paired data obtained from publicly available datasets, we also introduce a real motion-corrupted MRI dataset provided by Sichuan Provincial People's Hospital. A more comprehensive introduction of the datasets will be presented in Section 4.2. Peak Signal-to-Noise Ratio (PSNR) and Structured Similarity (SSIM) are used for quantitative evaluations. In all experiments, hyper-parameters are set as: $\lambda_1 = 0.01$, $\lambda_2 = 1 \times 10^{-6}$, $\lambda_3 = \lambda_5 = \lambda_7 = 10$, $\lambda_4 = \lambda_6 = 0.1$.

4.1. Motion Artifact Simulation

To generate highly realistic imitation of motion artifacts, we adapted the method proposed by AI-masni *et al.* [1]. Rigid motions can be regarded as a combination of rotational motion and transverse motion. This method applies rotational motion on 3 axis (x , y , z) in the range of $[-7^\circ, +7^\circ]$ and transverse motion on 3 planes (x^2 , y^2 , z^2) in the range of $[-7\text{mm}, +7\text{mm}]$. The degree of rotational motion and transverse motion is controlled by (t_x, t_y, t_z) , (ρ_x, ρ_y, ρ_z) . In order to generate motion artifacts with different intensity, we select the coefficients randomly from a range of number and amplitude of motions. Details are shown in Table 1.

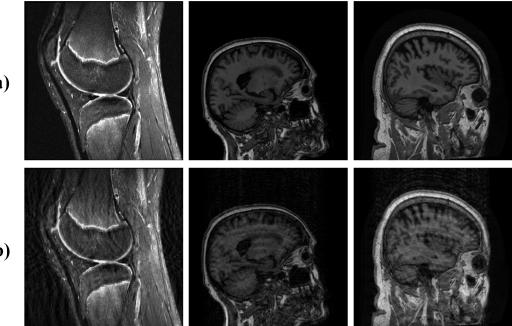


Figure 6. The example of simulated motion-corrupted images. (a) The original artifact-free images; (b) The corresponding simulated damaged images.

To make the simulated artifacts more realistic, we perform motion simulation in both k-space and real-space. The rotational motion is performed in real-space where artifact-free images are rotated according to the random coefficients. Then, we sample k-space lines that are transformed from the rotated images through Fourier Transform. In the meantime, the transverse motion is simulated by applying

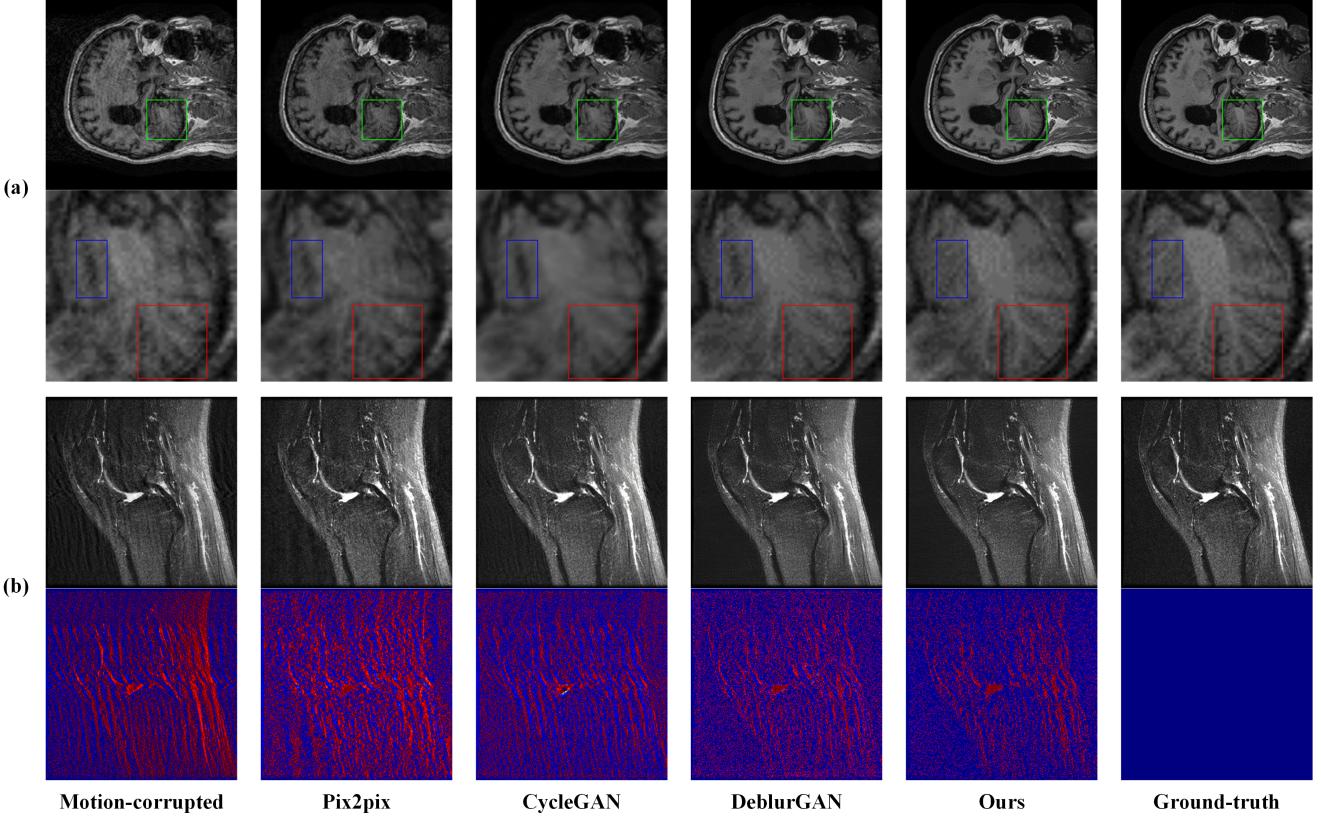


Figure 7. The artifact-corrected results from different models on different datasets. (a) Results on ADNI. The second line is an enlarged version of the content in the green box from the first line. The blue and red boxes indicate the regions that demonstrate the difference between each result; (b) Results on MRNet. The fourth line is the error images between corresponding results and the ground-truth image in the third line. It is colored for a better comparison.

Artifact Intensity	Rotational Motion	Transverse Motion	Repetition (N)
mild	$[-3^\circ, +3^\circ]$	$[-3mm, +3mm]$	$2 \sim 4$
moderate	$[-5^\circ, +5^\circ]$	$[-5mm, +5mm]$	$4 \sim 6$
severe	$[-7^\circ, +7^\circ]$	$[-7mm, +7mm]$	$5 \sim 7$

Table 1. The range of parameters for different artifact intensity.

a linear phase shift directly to the k-space of artifact-free image. This procedure is repeated for N times and N represents the number of twisted lines in k-space. Since motion simulation is performed at three different levels of motion intensity, N varies according to the level of motions. Each level corresponds to a specific range of N : mild, moderate and severe correspond to $10 \sim 40$, $40 \sim 60$ and $50 \sim 90$ lines, respectively. The deformed k-space lines caused by the rotational and transverse motions are accumulated to form the final K-space of motion-corrupted images. The example of simulated motion-corrupted image is shown in Figure 6.

4.2. Dataset

Two public datasets and one clinical dataset have been adopted in our experiments. MRNet is a MRI image dataset on knees, which is introduced in the competition [5] held by Stanford University. It has been widely used in research on MRI image processing [4, 23, 7]. This dataset contains complete series of MRI scan from 1370 patients, among which 18111 images are selected for training, 1672 for validation and 141 for testing. After manual screening, some slices with poor imaging quality are removed. We choose the images of 'MP-RAGE' protocol selected from the research-only data release of Alzheimer's Disease Neuroimaging Initiative(ADNI) as the second dataset. The training, validation and test set contain 42596, 5790 and 1920 images, respectively. We use the motion artifact simulation method described in section 4.1 to generate paired motion-corrupted and clean MRI images based on the two aforementioned datasets and construct the simulation datasets. All images in the simulation datasets are resized into 256×256 and normalized. Moreover, the ability and feasibility of the proposed model have been verified

by evaluating the performance on correcting real motion artifacts of clinical datasets from Sichuan Provincial People’s Hospital, which contains T1 and PD images of knee and ankle joints of 47 patients, with a total of 628 images.

4.3. Comparison with State-of-the-art Methods

To provide a comprehensive view, we compare FGAN with state-of-the-art image restoration methods: Pix2pix [14], CycleGAN [25] and DeblurGAN [17]. Each model is trained and tested with the same experiment setting.

In order to provide quantitative comparisons among different models, we perform two full-reference quality assessments, Peak-Signal-Noise-Rate (PSNR) and structural similarity (SSIM). On publicly available datasets, 8 experiments have been conducted. The comparison results on the test sets of both datasets have been summarized in Table 2. The visualization of results is shown in Figure 7.

According to both PSNR and SSIM, the performance rank on MRNet is: our model, CycleGAN, DeblurGAN and finally Pix2pix. Our method outperforms the other models by 10.68% in terms of PSNR, averagely. The quantitative performance rank is also clearly proved by the error images presented in the (b) part of Figure 7. Among all the artifact-corrected results on MRNet, although certain number of artifacts are corrected, Pix2pix introduce additional noise into the output. CycleGAN removed more scattered degrading noise than DeblurGAN, but it still fails to compete with ours in the correction of longitudinal artifacts.

Model	MRNet		ADNI	
	PSNR	SSIM	PSNR	SSIM
Pix2pix	27.6541	0.9899	33.2498	0.9407
CycleGAN	29.3880	0.9980	32.2777	0.8935
DeblurGAN	28.2961	0.9904	34.3147	0.9580
Ours	31.4654	0.9949	36.4845	0.9706

Table 2. Quantitative comparisons between different models tested on MRNet and ADNI.

Images from ADNI are reconstructed from head scan and therefore have much larger information density. PSNR and SSIM are still in consistence considering the performance rank. While our model still outperforms the other models by an average of 9.69% in terms of PSNR, Pix2pix beats CycleGAN in both PSNR and SSIM. As shown in the (a) part of Figure 7, CycleGAN does remove more artifacts than Pix2pix, however, it also smooths the image and damages the structural and textural information. Comparing to DeblurGAN, our method restores more details, especially these important regional information like gaps and boundaries between tissues.

Furthermore, in order to verify the feasibility of these model under real conditions, we use the models trained on

MRNet to test the motion-corrupted MRI images of knees provided by Sichuan Provincial People’s Hospital. Due to the complex mixture of different types of artifacts, noise and other degrading damage in real MRI images, SSIM and PSNR fluctuate dramatically among the output images. Therefore, we represent the visualizations in Figure 8 for auxiliary assessment. It can be observed that images obtained by Pix2pix and CycleGAN are more blurry than others and DeblurGAN introduces evident noise into the left region of the image, especially around the gap. Although our model has demonstrates its outstanding ability in artifact removal, it is clear that the image obtained by our method still contains a few artifacts.

4.4. Ablation Study

In this section, we perform module ablation study to validate the effectiveness of each proposed component. According to the structure illustrated in Figure 2, there are three components in FGAN: the k-space generator, the real-space generator and the stacked U-Nets. We remove three components respectively and set the corresponding loss functions as zero to perform each ablation experiment on MRNet. Apart from the incomplete network, all the experiment settings remain the same as in Section 4.3. The quantitative results of ablation study are shown in Table 3 and the visualization is illustrated in Figure 9, where No-K and No-R denote the k-space and real-space generators are removed respectively, No-Stack represents the absence of stacked U-Nets, and Origin is the proposed FGAN.

The effect of k-space generator In the ablation study on k-space generator, we set $l_{Gf,f}$ and $l_{Gf,r}$ as zero and the input motion-corrupted image is directly processed by the real-space generator. It is clear to see that the value of SSIM is not affected, while there is a great drop in terms of PSNR. This means that the structural information is still well-preserved, but the overall noise is much more than the result of complete FGAN. This is in consistence with the visualization shown in Figure 8. It is evident to see that there exists much noise around the central area. This is mainly because that k-space generator makes use of the frequency information and perform artifact correction in k-space. Hence, due to the N-to-N mapping relationship between k-space and real-space (the image), the overall noise intensity is well-contained.

The effect of real-space generator In the ablation study on real-space generator, we set $l_{Gr,r}$ and $l_{Gr,f}$ as zero and the output of k-space generator is processed with Inverse Fourier Transform to serve as the input of stacked U-Nets. In Table 3, we find that both PSNR and SSIM undergo a mild drop. It indicates that the general performance in artifact correction is restrained due to the absence of real-space generator. As shown in Figure 8, the image of No-R contains slightly more noise and less details than those

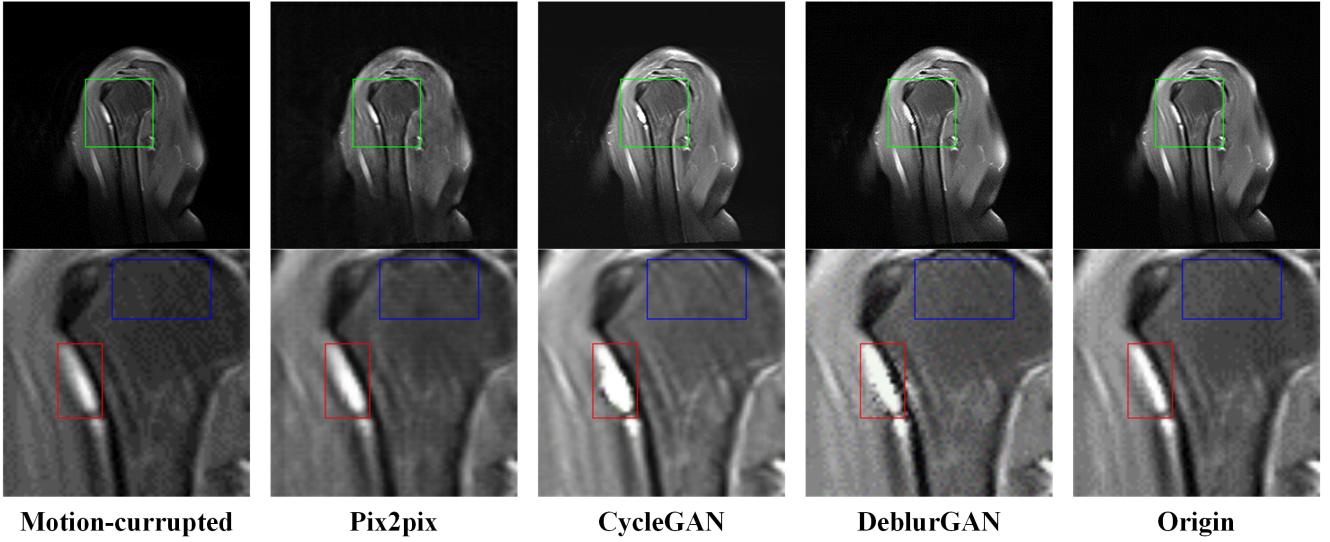


Figure 8. The comparison among real data and results from different models. We add the colored boxes for the same functions as in Figure 7.

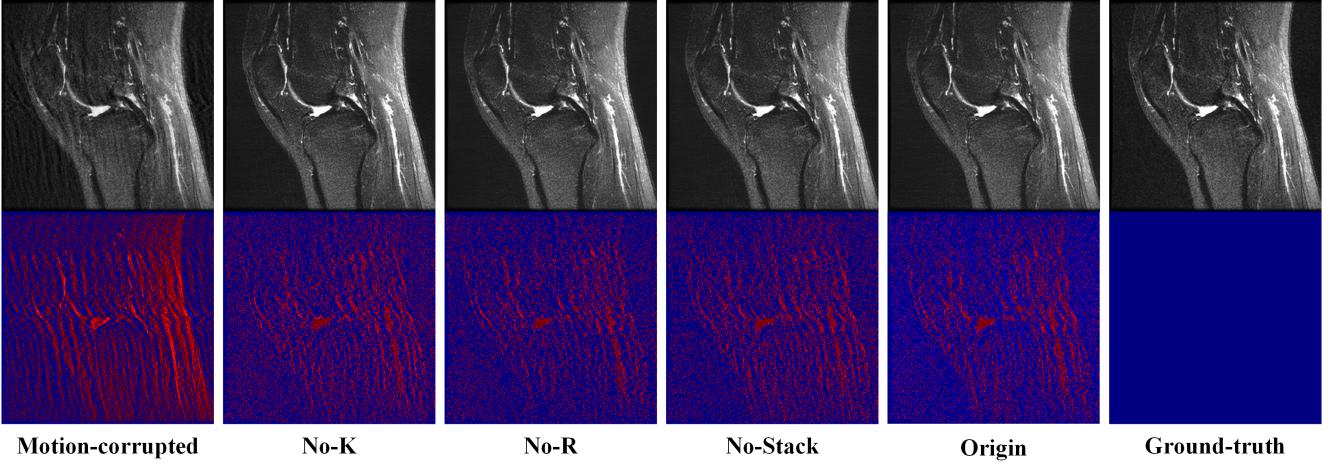


Figure 9. The visual analysis of ablation study.

from the image of Origin. The real-generator works as a traditional end-to-end model for image restoration that only utilizes the real-space data. Without this, the damaged image is then generally less denoised and deblurred and leads to lower PSNR and SSIM.

The effect of stacked U-Nets No loss is set as zero in the ablation study on stacked U-Nets, but the final output used in l_{GAN} is replaced with the output of real-space generator. Similar to the ablation study on real-space generator, values of both PSNR and SSIM decline from those of the complete FGAN, except the drop is much bigger than No-R. In Figure 8, No-Stack also has more noise and less details than Origin, but the specific degrading areas are different with No-R. The sharper decline proves that the contextual fusion implemented through stacked U-Nets greatly

remedies the detail loss and exploits the possibly undamaged structural information from the adjacent images.

Model	PSNR	SSIM
No-K	30.7120	0.9949
No-R	31.0800	0.9933
No-Stack	30.6085	0.9921
Origin	31.4654	0.9949

Table 3. The quantitative analysis of ablation study.

5. Conclusion

In this paper, we develop a GAN-based network named FGAN for the correction of motion artifact in MRI images. It combines fusions of cross-space and contextual

information through cascaded generators and stacked U-Nets. We compare FGAN with state-of-the-art methods on both simulated motion-corrupted images obtained from publicly available MRI datasets and real damaged MRI images provided by the hospital. The results demonstrate that FGAN outperforms state-of-the-art methods considering SSIM, PSNR and visual analysis. Moreover, the effectiveness of each proposed component in FGAN is validated through ablation study.

References

- [1] Mohammed A. Al-masni, Seul Lee, Jaeuk Yi, Sewook Kim, Sung-Min Gho, Young Hun Choi, and Dong-Hyun Kim. Stacked u-nets with self-assisted priors towards robust correction of rigid motion artifact in brain mri. *NeuroImage*, 259:119411, 2022.
- [2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
- [3] Karim Armanious, Sergios Gatidis, Konstantin Nikolaou, Bin Yang, and Thomas Kustner. Retrospective correction of rigid and non-rigid mr motion artifacts using gans. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1550–1554. IEEE, 2019.
- [4] Niamh Belton, Ivan Welaratne, Adil Dahlan, Ronan T. Hearne, Misgina Tsighe Hagos, Aonghus Lawlor, and Kathleen M. Curran. Optimising knee injury detection with spatial attention and validating localisation ability. In Bartłomiej W. Papież, Mohammad Yaqub, Jianbo Jiao, Ana I. L. Namburete, and J. Alison Noble, editors, *Medical Image Understanding and Analysis*, pages 71–86, Cham, 2021. Springer International Publishing.
- [5] Nicholas Bien, Pranav Rajpurkar, Robyn L. Ball, Jeremy Irvin, Allison Park, Erik Jones, Michael Bereket, Bhavik N. Patel, Kristen W. Yeom, Katie Shpanskaya, Safwan Halabi, Evan Zucker, Gary Fanton, Derek F. Amanatullah, Christopher F. Beaulieu, Geoffrey M. Riley, Russell J. Stewart, Francis G. Blankenberg, David B. Larson, Ricky H. Jones, Curtis P. Langlotz, Andrew Y. Ng, and Matthew P. Lungren. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of mrnet. *PLOS Medicine*, 15(11):1–19, 11 2018.
- [6] Kecheng Chen, Jiaxin Huang, Jiayu Sun, Yazhou Ren, and Xiaorong Pu. Task-driven deep learning for ldct image denoising. In *The Fourth International Symposium on Image Computing and Digital Medicine, ISICDM 2020*, page 35–39, New York, NY, USA, 2021. Association for Computing Machinery.
- [7] Yi-Ting Chou, Ching-Ting Lin, Ting-An Chang, Ya-Lun Wu, Cheng-En Yu, Tsung-Yu Ho, Hui-Yi Chen, Kai-Cheng Hsu, and Oscar Kuang-Sheng Lee. Development of artificial intelligence-based clinical decision support system for diagnosis of meniscal injury using magnetic resonance images. *Biomedical Signal Processing and Control*, 82:104523, 2023.
- [8] Ben A Duffy, Wenlu Zhang, Haoteng Tang, Lu Zhao, Meng Law, Arthur W Toga, and Hosung Kim. Retrospective correction of motion artifact affected structural mri images using deep learning of simulated motion. In *Medical Imaging with Deep Learning*, 2018.
- [9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [10] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: A survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2022.
- [11] Ishaaq Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [12] Daniel Christopher Hoinkiss, Peter Erhard, Nora-Josefin Breutigam, Federico von Samson-Himmelstjerna, Matthias Günther, and David Andrew Porter. Prospective motion correction in functional mri using simultaneous multislice imaging and multislice-to-volume image registration. *NeuroImage*, 200:159–173, 2019.
- [13] Jiaxin Huang, Kecheng Chen, Jiayu Sun, Xiaorong Pu, and Yazhou Ren. Cross domain low-dose ct image denoising with semantic information alignment. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4228–4232, 2022.
- [14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [15] Conghui Ji. *Removal of Medical Motion Artifact Based on Generative Adversarial*. PhD thesis, in Chinese, Chongqing University of Posts and Telecommunications.
- [16] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 694–711, Cham, 2016. Springer International Publishing.
- [17] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [18] Ilkay Oksuz, James Clough, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Rene Botnar, Daniel Rueckert, Julia A Schnabel, and Andrew P King. Cardiac mr motion artefact correction from k-space using deep learning-based reconstruction. In *Machine Learning for Medical Image Reconstruction: First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1*, pages 21–29. Springer, 2018.
- [19] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional networks. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

- tional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
 - [21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
 - [22] Nick Todd, Oliver Josephs, Martina F. Callaghan, Antoine Lutti, and Niklaus Weiskopf. Prospective motion correction of 3d echo-planar imaging data for functional mri using optical tracking. *NeuroImage*, 113:1–12, 2015.
 - [23] Balagopal Unnikrishnan, Cuong Nguyen, Shafa Balaram, Chao Li, Chuan Sheng Foo, and Pavitra Krishnaswamy. Semi-supervised classification of radiology images with noteacher: A teacher that is not mean. *Medical Image Analysis*, 73:102148, 2021.
 - [24] Shanshan Wang, Huitao Cheng, Leslie Ying, Taohui Xiao, Ziwen Ke, Hairong Zheng, and Dong Liang. Deepcomplexmri: Exploiting deep residual network for fast parallel mr imaging with complex convolution. *Magnetic Resonance Imaging*, 68:136–147, 2020.
 - [25] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.