EasyChair Preprint

№ 1227

Multi-scale Network with the deeper and wider residual block for MRI motion artifact correction

Wei-Liang Zhang, Qin-Yan Zhang, Ji-Jiang Yang and Qing Wang

June 24, 2019

Multi-scale Network with the deeper and wider residual block for MRI motion artifact correction

Weiliang Zhang[1,3], Qinyan Zhang[1,3], Jijiang Yang[2], and Qing Wang[2]

[1]Automation School.Beijing University of Posts and Telecommunications,Beijing,China
[2]Research Institute of Information Technology.Tsinghua University,Beijing,China
[3]Beijing Engineering Research Center of Post Intelligent Equipment,
(Postal Scientific Research and Planning Academy,Beijing 100096,China)
[2]Corresponding author: yangjijiang@tsinghua.edu.cn

***Abstract*—Magnetic resonance imaging (MRI) motion artifact is common in clinic which affects the doctor to accurately locate the lesion and diagnose the condition. MRI motion artifact is caused by the physiological movements of the patient while scanning the organ. Most of the current methods do artifact suppression and image restoration on the inverse Fourier transform level. They are neither effective nor efficient and can not be utilized in clinic. In this paper, the method that transfers deep learning into this domain with adopting a novel approach in Multi-scale mechanism for MRI motion artifact correction was proposed. What' more, a newer residual block with the deeper and wider architecture was proposed. With the deeper and wider residual block, the correction effect is greatly improved. The Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) were adopted as the evaluation metrics. In short, our model is trainable in an end-to-end network, can be tested in real-time and achieves the state-of-the-art results for MRI motion artifact correction.**

***Index Terms*—MRI motion artifact correction, deep learning, multi-scale, residual block**

I. Introduction

Magnetic resonance imaging (MRI) technology is widely used in many medical institutions. MRI technology can clearly show the internal structure of the human body such as organs, bones and blood. At the same time, MRI has no radiation and X-ray damage. Therefore, MRI not only can help doctors effectively identify tumors and necrotic lesions but also be not harmful to human's body. The motion artifacts will be produced in the process of generating magnetic resonance images which will affect the quality of the images seriously, especially for the edge of the object in the images. The details of the images will be lost, then it will be difficult for doctors to locate the key parts of the lesion which will have an extremely adverse impact on the diagnosis. Motion artifact in MRI is mainly caused by physiological movements of the patient during the process of generating images, such as atrial fibrillation, muscle spasm, and so on. These motions cause phase shifts during imaging. In the process of the inverse Fourier transform, these shifts will cause image signals of the same period to be combined at different locations. See Fig. 1 for the principle.

The correction of MRI motion artifact has received many attentions in recent years, and many different algorithms have been developed by the researchers. However, most of these methods do artifact suppression and image restoration on the inverse Fourier transform level. These methods can only do motion artifact correction for the images with the single blur kernel and it's time-cost and involves heuristic parameter-tuning and expensive computation leading to difficulties in implementation. Actual MRI with motion artifact is complex with many blur kernels that makes these traditional methods inapplicable.

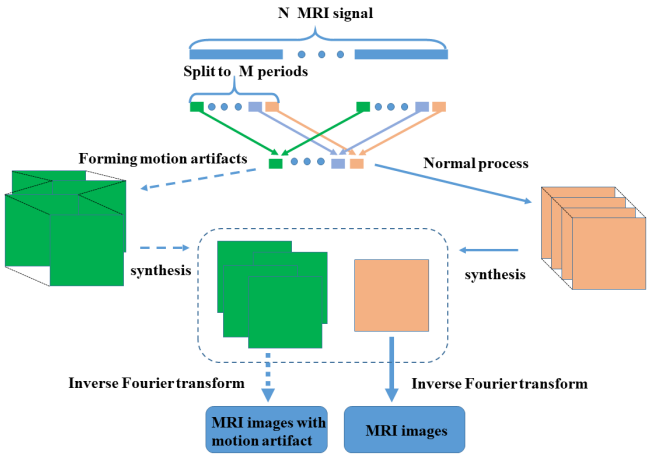

Fig. 1. The synthesis process of MRI images. The normal process is on the right that the signal from the same periods will be filled in the same places. However, if there is irregular atrial fibrillation during scanning, the signals will be stacked incorrectly in different places that the process was shown on the left.

Deep learning has achieved great development in recent years. Deep learning techniques are choosen to solve the problem of MRI motion artifact in our paper. In this paper, an end-to-end convolutional neural network structure was developed which takes blurred images as input to train the mapping between the blurred images and the sharp images through reducing the loss function, then utilized the trained model to restore images with motion artifacts.

Utilizing CNN for image deblurring has been widely studied and applied. Our method which utilizes CNN for MRI motion artifact correction is motivated by utilizing CNN for image

deblurring. The following formula (1) defines motion blur:

$$L_B = K * L_S + N, \tag{1}$$

,where L_B is a blurred image, K is a blur kernel, L_S is a sharp image, $*$ denotes the convolution operation and N is an external noise.

Through learning the formation principles of MRI motion artifact and the definition of motion blur, the facts that they have essential similarity in principle was found. Organ motion represents the blur kernel of the motion blur, the frequency domain of the inverse Fourier transform can be regarded as the image latent space. These provide theoretical support for the application of deep learning for MRI motion artifact correction.

This paper first introduced the background in I, then the related works about the paper are introduced in II, algorithm implementation and details are introduced in III. What's more, experimental results and analysis are written in IV. At last, some summary of the paper was made and some future work plans were done in V.

Our paper has the following contributions:

First, the connection between the MRI motion artifact correction and the image motion deblur in theory was found in our research. What's more, CNN has been utilized for MRI motion artifact correction successfully and got state-of-the-art results.

Second, an end-to-end convolutional neural network architecture for MRI motion artifact correction which takes the pair of blur image and sharp image as input to train a mapping between the blur images and the sharp images was developed. This network was named as Remove MRI Motion Artifact Network (RMMAN). This network uses the multi-scale information of the images to fully extract the feature information of the images of different scales. After many training iterations, good results could be achieved.

Third, ResNet is the basic network of our architecture, and ResBlock is widely used in various computer vision tasks. The residual network has the ability to make the network deeper so as to be good for parameter learning, so it was selected in our method. A deeper and wider ResBlock architecture was designed which greatly enhances the ability of the network to extract features, and it was called the deeper and wider ResBlock (DW-ResBlock) in our paper. Through utilizing this architecture, the same results while reducing the depth of the network could be achieved. After applying this architecture, our restored images not only get better results than the original network on the evaluation metric but also have a more actual edge and detail information, which makes the restored images are closer to the ground-truth images visually.

II. Related Works

A. Image deblurring

As mentioned in I, the methods of the image deblurring could be divided into two types: blind deblurring (unknowing blur kernel) and non-blind deblurring (knowing blur kernel).

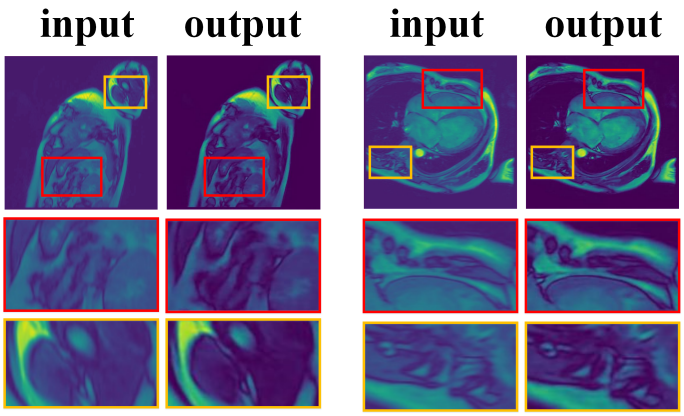

Fig. 2. Our results. The above two sets of images are MRI images of the patient's heart, but they are shot at different angles, so they don't look very similar. The column on the left is the input images, the column on the right is the output images. The first line is the MRI images, and the next two lines are the enlarged images.

Most of the early algorithms which rely on the classical Lucy-Richardson algorithm [1] were used to process the motion blur for images with clear blur kernel K. These algorithms are effective for the images with a single and clear blur kernel, while most of the image blur kernels in daily life are not the single kernels, on the contrary they are complex and diverse.

CNN has been widely used in the study of image deblurring and has achieved good results. Li Xu et al. [2] presented a generalized and mathematically sounded L0 sparse expression, they also presented a new effective method to deal with motion deblurring. Jian Sun et al. [3] used convolutional neural networks to predict the probability distribution of motion blur and used the Markov random field model to infer non-uniform motion blur fields and got good results. Orest Kupyn et al. [4] presented an end-to-end learning approach for motion deblurring, which is based on conditional GAN and content loss. This method adopts GAN idea which takes the blur images as input and generates sharp images, achieving the state-of-the-art results by a visual appearance at that time. Sainandan Ramakrishnan et al. [5] also adopt GAN as their basic framework. Furthermore, they added sparse connections and dense connections in their networks. This structure effectively reduces the computation time and gets good results.

B. MRI motion artifact correction

MRI motion artifact is very common in clinic. It is usually caused by the movements of the patient, such as atrial fibrillation and muscle spasm, which will affect the doctor's judgment of the patient's condition, so, it is very necessary to remove motion artifact. YH Tseng et al. [6] presented a new post-processing algorithm to deal with more general motion artifact. This algorithm corrects blurry images by constantly iterating through the knowledge of the image. Van, Anh T et al. proposed a new k-space and image space combination (KICT) method [7] to eliminate motion artifact and avoid incomplete correction of phase error. Huang Min et al. [8]

improved the MRI motion artifact correction method based on minimum entropy constraint, which improved the correction effect. Wu Chunli et al. [9] presented an image correction algorithm which combined Fourier projection algorithm and genetic algorithm for handling MRI motion artifact correction. This method has higher image clarity and faster imaging speed at that time. So far as today, the effects of motion artifact correction are extremely tiny.

C. multi-scale architecture

The idea of Multi-scale architecture is widely utilized in various tasks of deep learning, which could make the network has a more powerful ability to extract features. Seungjun Nah et al. [10] presented multiscale loss function that mimicked conventional coarse-to-fine approaches. Furthermore, they proposed a new large-scale dataset that providesd pairs of realistic blurry image and the corresponding ground truth sharp image that were obtained by a high-speed camera. With the proposed model and dataset, they got achieved exciting results not only qualitatively, but also quantitatively. Juncheng Li et al. [11] proposed a novel multi-scale residual network (MSRN) to fully exploit the image features. The architecture was well designed which using different convolution kernels and superimposing gradients of different scales. Qifeng Chen et al. [12] adopted a cascaded refinement network to generate realistic streetscape maps. This architecture could be applied to high-resolution images with strong adaptability. Multi-scale architecture could improve the ability to extract features without increasing network parameters, so our network adopted this architecture.

III. Proposed Methods

In this section, the proposed method was introduced in a comprehensive and detailed manner. The new residual block which was named DW-ResBlock in our paper was illustrated firstly. The DW-ResBlock was compared to the previous residual block and the benefits of such an architecture were introduced in principle. After that, the overall architecture was illustrated. It takes a sequence of different scale blurry images as input that were resized from the original image and restores a sequence of different scale sharp images. The sharp one at the original size as the final output is adopted. What's more, the loss functions were demonstrated and finally, the evaluation metric was introduced.

A. ResBlock and DW-ResBlock

The residual network has epoch-making significance in the field of deep learning, which addresses the problem that CNN is difficult to train an effective model. As we all know, the deeper network has a powerful ability to learning. However, with the network getting deeper and deeper, the problem of vanishing gradient and exploding gradient will occur when the network is training. The residual network has just solved this issue, so it is widely utilized in various tasks of deep learning, such as object detection, object segmentation. Recently, residual networks [13]–[15] exhibit excellent performance in computer vision problems of super-resolution. Ledig et al. [15] successfully applied the residual network to the super-resolution problem with developing SRResNet.

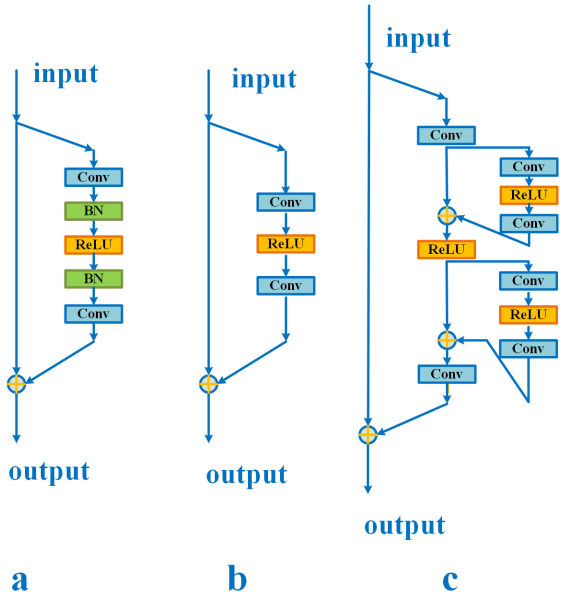

Fig. 3. Comparison of residual blocks in SRResNet (a), EDSR (b), and our proposed (c). Each convolution layer's kernel size is 5 and the activation function is ReLU. "+" represents the splicing operation.

In Fig. 3, the different residual network architectures which contains SRResNet [15] ,EDSR [16] and our proposed networks were illustrated. Compared to SRResNet, EDSR removes the batch normalization layers that make the network simple. In SRResNet, authors thought that since batch normalization layers normalize the features, they got rid of range flexibility from networks by normalizing the features, it was better to remove them. With the simple architecture, GPU memory usage is also sufficiently reduced since the batch normalization layers consume the same amount of memory as the preceding convolutional layers.

Removing the batch normalization layers was proved to be effective in the task of MRI motion artifact correction, so based on it, a deeper and wider residual network was proposed which was called DW-ResBlock in this paper. A Convolution-ReLU-Convolution architecture was added between the original convolution layer and ReLU activation function. A powerful ability to extract image features could be gotten with using the deeper and wider residual network. With utilizing DW-ResBlock, a bigger receptive field could also be gotten by utilizing the same number of ResBlock. Then the deep information of the MRI motion artifact images can be fully extracted. The better results could be achieved not only in evaluation metric but also in vision.

B. Overall architecture

As mentioned in I, a multi-scale network architecture was adopted in the coarse-to-fine strategy. In Fig. 4, each scale can be seen as an image deblurring subtask which takes blur

images as input and generates corresponding sharp images as output. The formula (2) can represent this subtask.

$$I_i = Net(B_i, I_{i-1}; Net_{par}) \tag{2}$$

Where i is the scale index, and i = 1 is the first scale with the smallest image size. After testing, the three scales achieved good results in balancing network parameters and evaluation metric. B_i and I_i are the blur images and restored images at the i-th scale, respectively. Net_{par} represents the training parameters of this subnetwork.

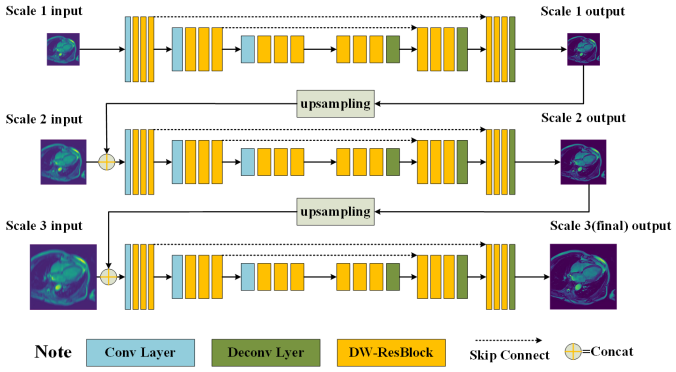

Fig. 4. Overall architecture of our model. Our architecture has three scales, in which the input of scale 1 is the initial input and the output of scale 3 is the final output. Each scale is a subnetwork which has three downsampling structures and three upsampling structures. The downsampling structure contains one 5 $\times$ 5 convolution layer and three DW-ResBlocks. The upsampling structure contains three DW-ResBlocks and one 5 $\times$ 5 convolution layer. The image size (S) is resized by the scale ratio(r) that S_i = S_{i-1} * r. + represents the splicing operation. And lines with arrows represent skip connections

Each subnetwork is a symmetric architecture with the same number of downsampling structures and upsampling structures. Downsampling structure contains one convolution layer and n DW-ResBlocks. In this paper, n is 3 is adopted because n = 1...5 have been tried and the network with n = 3 get the best result. Similarly, the upsampling structure contains m DW-ResBlocks and one deconvolution layer, m = 3 is chosen in our paper after testing. The dotted line with an arrow represents the skip connection which connecting the feature maps with the same dimension. The output image of the previous scale was upsampled and spliced with the blurred image of the next scale to input the next scale subnetwork. For example, the ***scale 1 output*** was upsampled and the upsample one and the ***scale 2 input*** were spliced. The spliced result will be taken as the input of the 2-nd scale. The output of the 3-rd scale is the final output of the whole network architecture. What's more, the convolution kernel size will also affect the final result because the bigger kernel size can get a bigger receptive field which will have a better ability to extract feature. However, our methods can't simply increase the size of the convolution kernel because if the kernel size is too large, the parameters of the network will become too much, so much so that the memory usage of GPU is unaffordable or there will be the problem of overfitting. Similarly, if the kernel size is too small, the features could not extract very well which will lead to the network's ability to express not very good then this will not get good results too. After doing the experiment, 5 was chosen as the size of the convolution kernel.

C. Loss function

The loss function is used to measure the learning ability of the training model. For the task of image deblurring, **L1 loss** and **L2 loss** are two classical choices. **L2 loss** has a stronger convergence ability which makes the model difficult to be overfitting. So **L2 loss** was adopted for each scale, the specific definition is as follows:

$$\mathcal{L} = \sum_{i=1}^{n} \frac{1}{N_i} \left|\left| I^i - I_g^i \right|\right|^2 \tag{3}$$

Where I^i represents the output of the i-th scale and I_g^i is the ground truth images with the corresponding size. N_i is the batch size of the i-th scale, and in this paper, all scale's batch size is the same. What's more, the **L1 loss** and **adversarial loss** [17] have also been tried, but **L2 loss** has the best performance, so **L2 loss** is adopted.

D. evaluation metrics

Our goal is to correct MRI motion artifact. With the restored images and ground truth images, some evaluation metrics are needed to use to determine the similarity between them. Structural Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR) are classical evaluation metrics which are effective for the task of evaluating the quality of the model. SSIM was first introduced by Zhou Wang et al. [18] who come from Laboratory for Image and Video Engineering. SSIM is a measure of the similarity between two images. Python's library could be used to calculate SSIM. The peak signal-to-noise ratio (PSNR) is an objective measure of image distortion or noise level. The greater the PSNR and SSIM between the two images, the more similar the two images are. In addition to the above two classic evaluation indicators, a new evaluation metric which is called Inception Score(IS) [19] has appeared to evaluate the clarity and diversity of generated images recently. In order to calculate the IS value of the image, the Google's inception v3 model [20] is adopted. The following formula (4) defines Inception Score.

$$IS(G) = exp(\underset{x \sim p_g}{E} D_{KL}(p(y|x)||p(y))) \tag{4}$$

Where $\underset{x \sim p_g}{E}$ represents the distribution of images generated from the generator. y y is the vector obtained after the images are input to inception v3. What's more, $p(y|x)$ is probability of y under the condition of x, $p(y)$ is the probability of y and $D_{KL}(p(y|x)||p(y))$ is the KL divergence between $p(y|x)$ and $p(y)$. In general, the larger the value of the inception score, the better the quality of the generated model.

IV. Experiments

In this section, the details and process of our experiments were introduced, containing the dataset used for our model and the settings of the model parameter. The results from both qualitative and quantitative aspects were given. What's more, not only the comparison of the original MRI motion artifact images and the restored images were given but also the evaluation metric containing SSIM and PSNR to evaluate the structural similarity between motion artifact input and restored output were given.

Because the MRI image is different from the image from daily life and most of the current methods to address the MRI motion artifact are not open-source, so lots of comparisons with other methods could not be done. So a comparison with the DeblurGAN, RMMAN (our network) with Fig. 3-b and RMMAN (our network) with Fig. 3-c in SSIM and PSNR were given. Because the purpose of MRI motion artifact correction is to make it easier for the doctor to diagnose the condition. Highlighting the edge of the images helps to diagnose the condition after we consulted doctors. So our network is better to have the ability to highlight the edge information of the images. Fortunately, the restored images have a good edge detection effect as shown in Fig. 5.

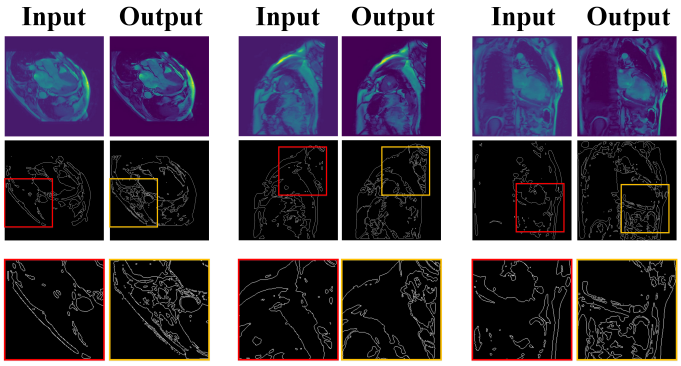

Fig. 5. The above pictures are collected from the heart of the patient, but they are shot at different angles. The column on the left is the input images, the column on the right is the output images. The first line is the MRI images, the second line is the corresponding edge detection results, and the third line is the enlarged partial edge detection results. As we can see, our network makes the output images have better edge detection results and more details.

A. Training Data

Our data for training are MRI images which are scanned at the different angles of the patients' heart. The cause of the MRI motion artifact is arrhythmia mostly. Fig. 6 shows the data in detail. There are thousands of images without motion artifact which are used for training and 1825 images with real motion artifact which are used for testing.

B. Image Preprocessing and Motion artifact generation

The collected MRI images are DICOM format which can't be taken as the input directly. So the images with DICOM format are transferred to the three-channel color images. What's more, some transformations which contain random rotation, translate and zoom were also adapted to make the dataset rich for training a comprehensive model. In general, the size of our final images is 512×512.

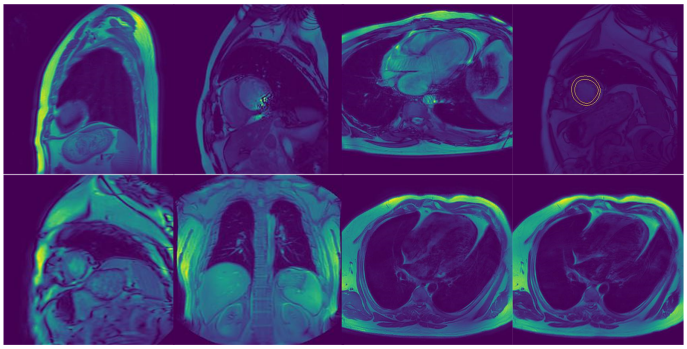

Fig. 6. Dataset:MRI images scanned from different parts of the heart

In the task of image super-resolution, researchers have the public dataset with image pairs for training and testing, such as DIV2K dataset, GoPro dataset, Set5, Set14 and so on. However, our research can't get image pairs which contain MRI motion artifact images and corresponding images without motion artifact. Other researchers who have no public dataset in their domain was also confronted with this problem. Some researchers even use Generative Adversarial Networks (GAN) to train a model for generating the approximate real image pairs. The idea describes by [4] was adopt in our methods. MRI motion artifact from sharp MRI was simulated via average random several frames moving along the motion orientation of diaphragm. The trajectory is generated by the Markov process. For example, the next point of the trajectory is generated randomly based on the current point, but the moving orientation should be consistent with atrial fibrillation. This method helps us to create a realistic image with MRI motion artifact. Then, the generated MRI motion artifact images and original sharp images were combined into image pairs for training. As for the testing phase, the real MRI motion artifact images were utilized for testing our model.

C. Evaluation on different methods

In this section, a comparison of three methods which contained DeblurGAN, RMMAN(our network) with Fig. 3-b and RMMAN (our network) with Fig. 3-c were given. Starting from the DeblurGAN, various settings were changed gradually to perform ablation tests. DeblurGAN is a network based on Generative Adversarial Networks. DeblurGAN has a generator which is used to generate the corresponding sharp images and a discriminator which is used to judge the quality of the generated images. Some adjustments to DeblurGAN were made to fit our dataset and task. RMMAN (our network) with Fig. 3-b was also adopted as our comparison.

The Tab.I shows the results of three methods in Evaluation. Our proposed method (RMMAN with Fig. 3-c) achieved the highest evaluation scores than the other two methods.

The repair capabilities of the networks with different convolution kernel size were also compared to prove our remarks

TABLE I
MEAN PEAK SIGNAL-TO-NOISE RATIO, STRUCTURAL SIMILARITY MEASURE

Evaluation Metric	Models		
	DeblurGAN	***RMMAN with Fig. 3-b***	***proposed***
PSNR	31.22	34.78	**34.97**
MSSIM	0.9324	0.9269	**0.9369**
Inception Score	1.2682	1.2643	**1.2770**

in section III-B. The Tab.II shows the results of three kernel size in Evaluation.

TABLE II
THE RESULTS OF DIFFERENT KERNEL SIZE

Evaluation Metric	Models		
	kernel size = 3	***kernel size = 7***	***kernel size = 5***
PSNR	32.69	32.54	**34.97**
MSSIM	0.9197	0.9232	**0.9369**
Inception Score	1.2585	1.2643	**1.2770**

D. Runtime

All of our models were implemented with using Tensorflow [21] deep learning framework and perform the training on Texla K80 GPU. For model training, Adam solver [22] with β_1 = 0.9, β_2 = 0.999 and $\epsilon = 10^{-8}$ are used. The learning rate is exponentially decayed from initial value of 0.0001 to $1e^{-6}$at 2,000 epochs using power 0.3. All the models are trained with batch size = 10. The code released by [23] was utilized as our framework.

2,000 epochs are enough for convergence, which takes about 192 hours for training. In addition to comparably robust and visual reality correction results, our proposed method is the first real-time, which will take 1.71s to generate a sharp image in the size of 512×512 on the GPU. A transfer from JPEG to DICOM post-process would take up to an additional 160.8 ms on the CPU. Comparing to existing complex and time-cost correction methods that need several iterations to get results, our method has large advantages.

V. CONCLUSION

In this paper, a deeper and wider residual block (DW-ResBlock) was developed and why it is effective in the task of MRI motion artifact correction was explained. What's more, an end-to-end correction system which based on the idea of multi-scale was proposed and the DW-ResBlock was adopted in the system. The method in deep learning domain was successfully transferred to the task of MRI motion artifact correction and it got state-of-the-art results both qualitatively and quantitatively. However, whether the results have been up to the clinical standard? We need to consult doctors for further professional evaluation in medicine. As mentioned in section IV-D, the time to generate a sharp image is about 1.71s. We think this is still not fast enough, so we will do more research to make the algorithm faster. We believe the DW-ResBlock can be applied to other image processing tasks and we will do research in this field in the future.

REFERENCES

[1] M. K. Singh, U. S. Tiwary, and Y. H. Kim, "An adaptively accelerated lucy-richardson method for image deblurring," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 1–10, 2008.

[2] X. Li, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2013.

[3] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," no. CVPR, pp. 769–777, 2015.

[4] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," 2017.

[5] S. Ramakrishnan, S. Pachori, A. Gangopadhyay, and S. Raman, "Deep generative filter for motion deblurring," 2017.

[6] Y. H. Tseng, J. N. Hwang, and C. Yuan, "Motion artifact correction of mri via iterative inverse problem solving," in *Image Processing, Icip-94, IEEE International Conference*, 1994.

[7] A. T. Van, D. C. Karampinos, J. G. Georgiadis, and B. P. Sutton, "K-space and image space combination for motion artifact correction in multicoil multishot diffusion weighted imaging," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2008, pp. 1675–1678, 2008.

[8] M. Huang, X. J. Qin, and L. I. Qing-Yuan, "Research and realization of correction method of mri motion artifact," *Chinese Journal of Magnetic Resonance Imaging*, 2013.

[9] C. Wu, Z. Du, Y. Kun, and W. Yun, "An improved algorithm of translational motion artifact correction for mri," in *Control & Decision Conference*, 2015.

[10] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," pp. 257–265, 2016.

[11] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *IEEE Conference on European Conference on Computer Vision*, 2018.

[12] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," 2017.

[13] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2016.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2016.

[16] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Computer Vision & Pattern Recognition Workshops*, 2017.

[17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv preprint arXiv:1703.10593*, 2017.

[18] W. Zhou, B. Alan Conrad, S. Hamid Rahim, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans Image Process*, vol. 13, no. 4, pp. 600–612, 2004.

[19] S. Barratt and R. Sharma, "A note on the inception score," 2018.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Computer Vision & Pattern Recognition*, 2016.

[21] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, and X. Zheng, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] T. Xin, H. Gao, W. Yi, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," 2018.