MULTI-SCALE ADAPTIVE TRANSFORMER FOR IMAGE DEBLURRING IN MAGNETIC PARTICLE IMAGING

Fan Yang[1,2,3†], *Liwen Zhang*[1,2†], *Jiaxin Zhang*[1,2,3], *Zechen Wei*[1,2,3], *Xin Yang*[1,2], *Jie Tian*[4,5⋆], *Hui Hui*[1,2,3⋆]

[1]CAS Key Laboratory of Molecular Imaging, Institute of Automation, Beijing, China
[2]Beijing Key Laboratory of Molecular Imaging, Beijing, China
[3]University of Chinese Academy of Sciences, Beijing, China
[4]Key Laboratory of Big Data-Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology, Beijing, China
[5]School of Engineering Medicine
School of Biological Science and Medical Engineering, Beihang University, Beijing, China

ABSTRACT

Magnetic particle imaging (MPI) is a novel medical imaging technology, which uses the nonlinear magnetization response of super-paramagnetic iron oxide (SPIOs) to describe the distribution of SPIOs in the target region. X-space method is a widely used image reconstruction method of MPI, which directly reconstruct in spatial domain. However, due to its inherent limitation of disregarding the physical properties of real systems, native X-space reconstructed MPI images are usually very blurry which limits the subsequent clinical use. In this study, we propose a novel multi-scale adaptive transformer (MSA-transformer) for image deblurring and recover the detail information. We perform experiments on both simulated datasets and real data to evaluate of our method. The experimental result shows, our method outperforms existing methods including methods used in MPI and some prevalent restoration methods used in computer vision.

Index Terms— Magnetic particle imaging, X-space, deep learning, transformer, image deblurring

1. INTRODUCTION

Magnetic particle imaging (MPI), a emerging medical imaging technique, can sensitively detect the magnetization response of magnetic nanoparticles in vivo [1]. MPI has many advantages, including high sensitivity and contrast, no depth limitation and no ionizing radiation [2, 3]. With the continuous development of MPI theory and instruments, MPI has shown promising in the application of vascular imaging [4], tumor detection [5], cell tracking [6].

Currently, system matrix (SM) reconstruction [7] and X-space [8] reconstruction are two main approaches for MPI image reconstruction. The SM method offers high imaging accuracy, however, measuring the complete SM is often time-consuming. In contrast, X-space method directly reconstruct in spatial domain achieving fast imaging. Nevertheless, due to influence of point-spread function and the limitation of assumptions in X-space theory [8, 9], the reconstruction images of X-space method are usually blurry which causes negative impact on the subsequent use of MPI images. Image deblurring is a widely studied problem in the field of image processing and low-level computer vision. The purpose of image deblurring is to recover a sharp image from the blurry one [10]. Traditional image deblurring utilize deconvolution algorithms such as Lucy-Richardson method or Wiener deconvolution to restore high quality images. However, these methods are often not effective enough in complex real-world scenarios with noise. With the rise of deep learning, many researchers use deep learning to solve the image deblurring problem [10]. transformer, a network architecture achieves great success in natural language processing, is increasingly used in image processing tasks for its powerful modeling capabilities. Self-attention is the core component of transformer architecture that is effective in modelling global-range correlation, but may neglect the local details such as the edge information which is quite important in X-space reconstruction. Convolution operation shifts the kernel across the image detect the repetitive patterns which could help to restore local information and provide a complementary effect to self-attention.

In this paper, we propose a multi-scale adaptive transformer (MSA-transformer) to explicitly extract local and global information and adaptively fuse them together to effectively reduce blur in native MPI images. We use different types of self-attention to extract global information and proposed a local channel attention module to extract local information. After that, we combine the information of different scales with channel-wise adaptive mix-up. We verify the performance and generalization of our model on differ-

† Equal contribution; ⋆ Corresponding author

ent simulated MPI datasets, our method outperform all other methods which is prevalent in image restoration.

2. METHOD

2.1. Network structure

Our proposed MSA-transformer is shown in Fig.1, which is an encoder-decoder architecture with residual connection. We first use combination of convolution layer and residual convolution blocks to extract shallow features and downsamples the image resolution. Next, we operate our multi-scale adaptive transformer (MSAT) blocks on the smallest and second-smallest scales. MSAT consists of multi-scale attention (MSA) module and multi-scale feed-forward network (MSFN) module. For MSA module, we use two types of self-attention to extract global information. The first one is intra-strip attention from Stripformer. The strip attention could capture blur patterns and decompose them into different orientations which is benefitial in solving the anisotropic problem of MPI images. The second one is multi-Dconv head transposed attention (MDTA) block from Restormer, which applies self-attention across feature dimension and computes cross-covariance across feature channels. In addition to using self-attention to extract global-range information, we apply a local channel attention (LCA) block to focus on local information which is composed of convolutional layers and channel-attention block playing a complementary role to self-attention. we combine the extracted global and local feature with channel-wise adaptive mix-up which is a channel-wise learnable parameter. Besides, We also introduce a multi-scale feed-forward network (MSFN) consisting of depth-wise convolutions and normal convolutions to replace the original feed-forward network in transformer. After using MSAT blocks to extract and combine local and global information in high level, we adopt transpose convolution for up-sampling. Its output features are combined with the feature maps from encoder on the same scale using channel-wise adaptive mix-up as well. Finally, our network ends with residual blocks and a convolution layer with a residual connection to the input image.

2.2. Loss function

We employ a combination of three loss functions to train our network: Charbonnier loss, Edge loss and Contrastive loss [11]. The combined use of the three loss functions can make the model pay attention to different aspects and improve the recovery effect of the model on MPI images.

Charbonnier loss:

$$L_{char} = \sqrt{\|Y - \hat{Y}\|^2 + \varepsilon^2}, \tag{1}$$

where Y is the real high quality MPI image and $\hat{Y}$ is the predicted high quality MPI image, ε is set to 10^3. Charbonnier loss is a commonly used loss function in image restoration tasks.

Edge loss:

$$L_{edge} = \sqrt{\|\Delta Y - \Delta \hat{Y}\|^2 + \varepsilon^2}, \tag{2}$$

where Δ is the Laplacian opeator. Edge loss encourages the model focus on the edge information and recovery more accurate edges.

Contrastive loss:

$$L_{con} = \frac{L1(\psi(Y) - \psi(\hat{Y}))}{L1(\psi(X) - \psi(\hat{Y}))}, \tag{3}$$

where ψ is the extracted hidden features from fixed pre-trained VGG-19. L1 is the L1 norm. Contrastive loss pulled the deblurred MPI image $\hat{Y}$ close to the clear ground truth while pushing $\hat{Y}$ away from the blurred input X in latent feature space. The overall loss function of LGA-transformer is

$$L = L_{char} + \lambda_{edge} L_{edge} + \lambda_{con} L_{con}, \tag{4}$$

where we set $\lambda_{edge} = 0.05$ and $\lambda_{con} = 0.0005$.

2.3. Implementation Details

We use images reconstructed in low gradient field (3T/m) as input and deconvoluted images in high gradient (6T/m) as label. The number of stacks of Strip-SA-LCA blocks and MDTA-LCA blocks of our model is set to be 6 and 3 in the smallest and second-smallest scales. We initialize the model weights using Xavier initialization. For the training process, the network operates on 120×120 input images with a batch size of 16. We utilize the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The network is trained for 200 epochs and the initial learning rate is set to 1e-4 and cosine decay to 1e-5. To ensure fairness, our method and all comparison methods are used the same training settings. The models is trained using the PyTorch framework.

2.4. Metrics

To evaluate the performance of our LGA-transformer, we utilize the Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to provide quantitative measurements of the quality of the restored MPI images.

3. EXPERIMENTS

3.1. Dataset

In this study, two simulated dataset are used to train and evaluate the performance and generalization of model. The first one is Mixed National Institute of Standards and Technology database (MNIST) dataset as simulated phantoms including 10000 images. The original images were resiezed

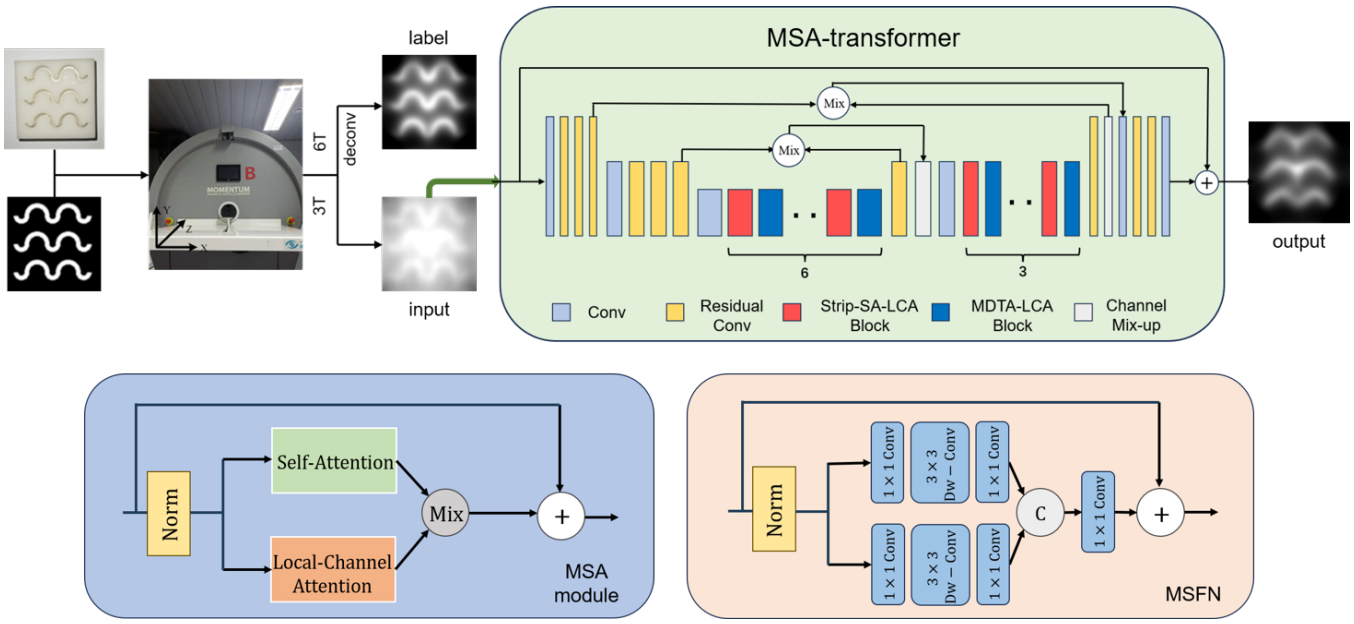

Fig. 1. Architecture of our proposed MSA-transformrer network and details of the two modules of MSAT block. Strip-SA-LCA-Block and MDTA-LCA are different types of MSAT blocks with different types of self-attention. Mix represent channel-wise adaptive mix-up which is a channel-wise learnable parameter.

to 120×120 in order to agree with the simulation parameters. In MNIST dataset, 8000 are used for training, 1000 for validation and 1000 for testing. The second one is Phantoms dataset, we construct different types of phantoms including: resolutions, vessels, vessel stenosis tumors and shape combinations to verify the generalization performance of our model. There are a total of 1000 phantoms images and we only use them for testing. We utilized the Langevin equation to simulate the magnetization response of MNPs in MPI. The Langevin function can be written as follow:

$$M(r,t) = M_0(r) \times [coth(\frac{m|B(r,t)|}{k_B T}) - \frac{k_B T}{m|B(r,t)|}], \quad (5)$$

where $B(r,t)$ is the total magnetic field at position r and time t, m is the magnetic moment and $M_0(r) = \frac{N(r)m}{\Delta V}$ is the saturation magnetization of the particles contained in the volume ΔV, $N(r)$ is the number of particles each contributing a magnetic moment $m = \frac{1}{6}\pi D^3 M_{sat}$. To reconstruct the MPI image, we applied the X-space method and generated the simulated MPI image using MATLAB. To make our simulations more realistic and to make our model more robust to real noise, we used a commercial MPI equipment (MOMENTUM CT, Magnetic Insight Inc., Alameda, CA, USA). to collect real noise images. We randomly crop the acquired noise images and add them to the native image with different proportion σ. The MPI simulation parameters are lists in Table 1. We also uses real phantom data to verify the effectiveness of our method.

Table 1. MPI simulation parameters

Parameters	Value	Unit
Permeability of vacuum	$4\pi \times 10^{-7}$	NA^{-2}
Nanoparticle diameter	20	nm
Saturation magnetization	4.77×10^{-5}	Am^{-1}
Nanoparticle diameter	6.75×10^{-18}	Am^2
Kelvin temperature	293.15	K
Boltzmann constant	1.38×10^{-23}	JK^{-1}
Gradient	3,6	Tm^{-1}
Field of view	30×30	mm^2
Noise SNR	$20 - 30$	dB

3.2. Experiment result

To demonstrate the effectiveness of our proposed method in removing the blurred artifacts and improving spatial resolution of native MPI images. We conducted experiments on different datasets and compared our approach with previous method used in MPI such as FDS-MPI [12] and some powerful transformer-based image restoration model like SwinIR [13], Stripformer [11] and Restormer [14]. We show examples of deblurred MPI images in Fig.2. The first row is a digit phantom result from MNIST test set. We can see that except FDS-MPI, other methods are better to restore the shape of handwritten digits. However, the images recovered by Restormer and Stripformer have visually broken parts that are inconsistent with the high-quality images. Although the SwinIR restores a more complete shape, there are some irreg-

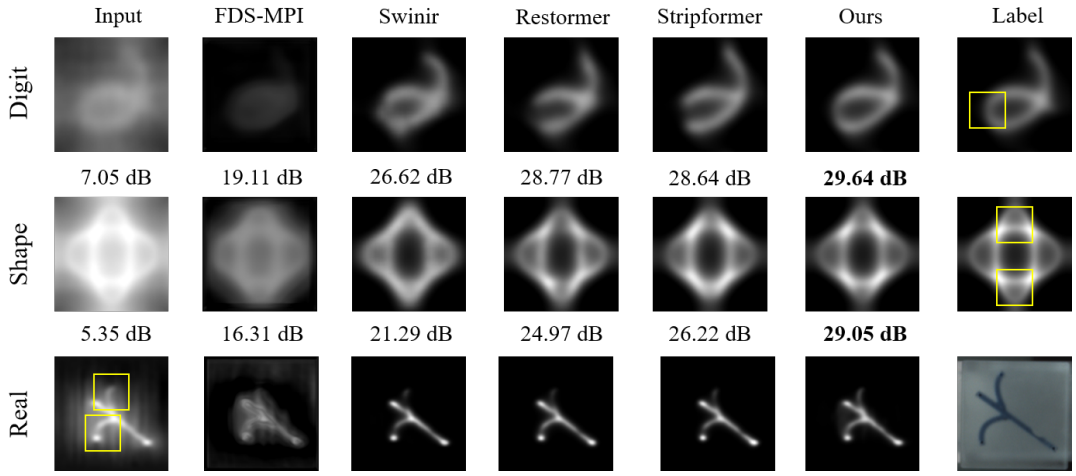

Fig. 2. Examples of deblurring effect comparison of different methods on different types of data.

ular blurring artifacts. The second row is an shape phantom example from Phantoms dataset. As can be seen from the results, all methods have recovered the shape of the phantom. Although FDS-MPI recovered the shape, it fails to deblur the image. Other methods achieves better deblurring effect, but there are still differences in details. Focusing on the boundary of the two holes at the top and the bottom of the star shape(yellow box), compared to other methods, our method restores the boundary more clearly which makes the two holes appear more distinct. From the quantitative indicators, we can also see that our method outperform all other methods. The last row is an example in real situation. Visually, our method produces much clearer branches than other methods which shows the effectiveness of our method in real scenarios. We calculated the average quantization results on the MNIST test set and the Phantoms test set, which are shown in Table 2. Our method achieves the highest PSNR and SSIM on both dataset, which indicates that our model has superior performance and generalization ability to accurately recover images form blurred and noisy data.

Table 2. Performance evaluation for different methods

Method	MNIST		Phantoms	
	PSNR	SSIM	PSNR	SSIM
FDS-MPI [12]	20.16	0.5082	18.41	0.4137
SwinIR [13]	31.17	0.9534	27.99	0.9152
Stripformer [11]	33.21	0.9711	30.02	0.9367
Restormer [14]	33.68	0.9737	30.76	0.946
LGA-transformer (ours)	**34.87**	**0.979**	**31.03**	**0.9479**

4. CONCLUSION

In this paper, we proposed MSA-transformer for effectively deblurring and improving the spatial resolution of MPI images. Through the combination of self-attention and local attention adaptive, our network can effectively extract multi-scale information, model global information, and focus on local details to achieve more effective MPI image deblurring. In the simulation, we use quantitative indicators to demonstrate the superiority of our approach. In the real situation, our network achieves better result visually which shows the effectiveness of our method for deblurring MPI images in real scenario.

5. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant: 62027901, 81827808, 81930053,81227901, 82202269. Beijing Natural Science Foundation: JQ22023; CAS Youth Innovation Promotion Association under Grant Y2022055. The authors would like to acknowledge the instrumental and technical support of Multi-modal Biomedical Imaging Experimental Platform, Institute of Automation, Chinese Academy of Sciences.

6. COMPLIANCE WITH ETHICAL STANDARDS

No ethical approval was required for this study.

7. REFERENCES

[1] Bernhard Gleich et. al., "Tomographic imaging using the nonlinear response of magnetic particles," *Nature*, vol. 435, no. 7046, pp. 1214–1217, 2005.

[2] Tobias Knopp and Thorsten M Buzug, *Magnetic particle imaging: an introduction to imaging principles and scanner instrumentation*, Springer Science & Business Media, 2012.

[3] Tobias Knopp, Nadine Gdaniec, and Martin Möddel, "Magnetic particle imaging: from proof of principle to preclinical applications," *Physics in Medicine & Biology*, vol. 62, no. 14, pp. R124, 2017.

[4] AP Khandhar, P Keselman, SJ Kemp, RM Ferguson, PW Goodwill, SM Conolly, and KM Krishnan, "Evaluation of peg-coated iron oxide nanoparticles as blood pool tracers for preclinical magnetic particle imaging," *Nanoscale*, vol. 9, no. 3, pp. 1299–1306, 2017.

[5] Liron L Israel, Anna Galstyan, Eggehard Holler, and Julia Y Ljubimova, "Magnetic iron oxide nanoparticles for imaging, targeting and treatment of primary and metastatic tumors of the brain," *Journal of Controlled Release*, vol. 320, pp. 45–62, 2020.

[6] Kyung Oh Jung, Hunho Jo, Jung Ho Yu, Sanjiv Sam Gambhir, and Guillem Pratx, "Development and mpi tracking of novel hypoxia-targeted theranostic exosomes," *Biomaterials*, vol. 177, pp. 139–148, 2018.

[7] Alper Güngör, Baris Askin, Damla Alptekin Soydan, Emine Ulku Saritas, Can Barış Top, and Tolga Çukur, "Transms: Transformers for super-resolution calibration in magnetic particle imaging," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3562–3574, 2022.

[8] Patrick W Goodwill and Steven M Conolly, "Multidimensional x-space magnetic particle imaging," *IEEE transactions on medical imaging*, vol. 30, no. 9, pp. 1581–1590, 2011.

[9] Zechen Wei, Yanjun Liu, Tao Zhu, Xin Yang, Jie Tian, and Hui Hui, "Bss-tfnet: Attention-enhanced background signal suppression network for time-frequency spectrum in magnetic particle imaging," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023.

[10] Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li, "Deep image deblurring: A survey," *International Journal of Computer Vision*, vol. 130, no. 9, pp. 2103–2130, 2022.

[11] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin, "Stripformer: Strip transformer for fast image deblurring," in *European Conference on Computer Vision*. Springer, 2022, pp. 146–162.

[12] Yaxin Shang, Jie Liu, Liwen Zhang, Xiangjun Wu, Peng Zhang, Lin Yin, Hui Hui, and Jie Tian, "Deep learning for improving the spatial resolution of magnetic particle imaging," *Physics in Medicine & Biology*, vol. 67, no. 12, pp. 125012, 2022.

[13] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1833–1844.

[14] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.