International Journal on Magnetic Particle Imaging
Vol 11, No 1, Suppl 1, Article ID 2503012, 3 Pages

Proceedings Article

Background signal suppression using a transformer-based masked autoencoder for magnetic particle imaging

Zechen Wei[a] · Xin Yang[a] · Jie Tian[b] · Hui Hui[a,*]

[a]CAS Key Laboratory of Molecular Imaging, Institute of Automation, Beijing, China
[b]School of Engineering Medicine & School of Biological Science and Medical Engineering, Beihang University, Beijing, China
[*]Corresponding author, email: hui.hui@ia.ac.cn

Abstract

Magnetic Particle Imaging (MPI) is an emerging imaging technique that utilizes the nonlinear response of superparamagnetic iron oxide nanoparticles to generate an image of their spatial distribution. To achieve high-quality MPI images, it is crucial to suppress background noise. In this work, we propose a transformer-based masked autoencoder for learning the relationships between harmonic components to improve noise suppression. Experimental results demonstrate that the proposed method effectively reduces background noise across varying levels.

I. Introduction

The application of magnetic particle imaging (MPI) is heavily reliant on the image quality, which in turn depends on both the signal quality of the hardware system and the reconstruction algorithm employed [1]. However, environmental factors and system-induced noise can degrade the acquired signals, leading to a reduction in image quality [2]. Noise manifests in various patterns, and effectively suppressing different types of noise simultaneously remains a challenge due to their inherent randomness and diversity. In previous work, we proposed a self-attention-based method [3], which dealt with the two-dimensional (2D) time-frequency spectrum [4] obtained from short-time Fourier transform (STFT) of the temporal signal to suppress varying levels of background noise. In the current work, we introduce a transformer-based masked autoencoder [5]. We pre-train the encoder to learn the relationships between different harmonic components using a set of simulated system matrices (SMs), and then transfer this knowledge to the background noise suppression task. Our experimental results show that the fine-tuned model performs more effectively in suppressing background noise.

II. Method

II.I. Datasets

We generated a set of SMs corresponding to MPI systems with varying gradients and scan trajectories to serve as our pre-training dataset. The field of view is fixed at 24×24, with 168 frequency components extracted, resulting in an SM size of 168×576. Simultaneously, we created 0-1 binary matrices of equal size to serve as mask matrices. Additionally, phantom images containing numbers and letters were used to represent the particle concentration distribution. Field-free point scanning signals were generated, followed by window function framing and STFT to obtain time-frequency spectrums, replicat-

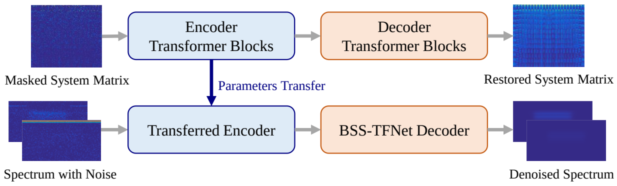

Figure 1: Illustration of the transformer-based masked autoencoder and the transferring process to background signal suppression task.

ing the process described in [3, 4]. Noise was added to the temporal signals to generate noisy spectrums.

II.II. Network Structure

We pre-trained the model to learn harmonic relationships by restoring randomly masked SMs. Based on the Transformer architecture [5], we randomly set values in the SM to zero and divided the SM into equal-sized, non-overlapping patches. For both the encoder and decoder, we employed a series of standard Transformer blocks. Mean squared error was used as the loss function during pre-training. After pre-training, the model was fine-tuned for the spectrum denoising task, leveraging the learned harmonic knowledge to process the harmonic data of the time-frequency spectrum. For the decoder, we used BSS-TFNet [3], an end-to-end spectrum enhancement network for background noise suppression. The pre-trained encoder replaced the feature extraction module in BSS-TFNet. The hyperparameters of the network were adjusted according to the size of the spectrum, with Mean Absolute Error (MAE) as the loss function during the fine-tuning phase.

III. Results

We validated our method on datasets with varying levels of background noise and compared it with BSS-TFNet [3]. Both methods were trained using datasets with a signal-to-interference ratio (SIR) of 5 dB and a signal-to-noise ratio (SNR) of 10 dB, and were tested on higher noise levels. Experimental results, shown in Table 1 as mean values ± standard deviations, reveal that the network with pre-trained encoders consistently achieved the best performance across all datasets. In contrast, models without pre-trained encoders (direct) performed slightly worse, indicating that pre-training with harmonic knowledge enhances noise suppression. Additionally, the performance degradation of BSS-TFNet can be attributed to the limited training dataset, which hindered effective harmonic feature extraction.

Table 1: Quantitative results of different networks.

Method	Spectrum Amplitude	Reconstructed Image
	MAE ($\times 10^{-3}$)	Peak Signal-to-Noise Ratio ↑
SIR=0 dB, SNR=10 dB		
Input	35.70±11.72	11.09±1.54
BSS-TFNet	3.47±1.20	28.83±4.38
Direct	1.57±1.04	36.07±4.87
Pre-trained	1.12±0.66	38.88±4.32
SIR=5 dB, SNR=10 dB		
Input	27.84±6.99	13.02±1.69
BSS-TFNet	1.49±0.46	35.95±3.62
Direct	0.88±0.30	40.53±3.89
Pre-trained	0 .67±0.24	43.08±4.49
SIR=5 dB, SNR=0 dB		
Input	85.89±23.00	10.30±0.83
BSS-TFNet	3.62±1.43	30.72±3.80
Direct	2.31±0.94	31.90±4.05
Pre-trained	1.40±0.44	37.26±4.47

IV. Conclusions

We have introduced a deep learning-based approach for enhancing MPI signals, where the encoder is pre-trained on masked SMs to learn the relationships between different harmonics. The pre-trained encoder is then transferred to the background noise suppression task. Experimental results demonstrate that the pre-trained model improves noise suppression performance. Therefore, our method can assist in obtaining high-quality MPI images, when combined with X-space reconstruction techniques.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grants: 62027901, 92359304, U23A6011; Beijing Natural Science Foundation under Grants: JQ22023, L242064; CAS Youth Innovation Promotion Association under Grant Y2022055.

Author's statement

Authors state no conflict of interest.

References

[1] Z. Wei, et al., *First non-human primate-sized magnetic particle imaging system based on digital-scanned focus field*, IEEE Trans. Instrum. Meas., vol. 73, pp. 5023611, 2024.

[2] T. Knopp, et al., *Dictionary-based background signal estimation for magnetic particle imaging*, IEEE ISBI, pp. 1540-1543, 2021.

[3] Z. Wei, et al., *BSS-TFNet: Attention-enhanced background signal suppression network for time-frequency spectrum in magnetic particle imaging*, IEEE Trans. Emerging Top. Comput. Intell., vol. 8, no. 2, pp. 1322-1336, 2024.

[4] Z. Wei, et al., *Recovery of fundamental frequency component in magnetic particle imaging using an attention-based neural network*, Int. J. Magn. Part. Imaging, vol. 9, no. 1, pp. 2303033, 2023.

[5] K. M. He, et al., *Masked autoencoders are scalable vision learners*, IEEE CVPR, pp. 15979-15988, 2022.

10.18416/ijmpi.2025.2503012