

Enhanced Motion Correction in Magnetic Resonance Fingerprinting via a Deep Learning Approach

Mengze Gao
Department of Radiology
Stanford University
gaomz@stanford.edu

Xuetong Zhou
Department of Bioengineering
Stanford University
xuetongz@stanford.edu

Xiaozhi Cao *
Department of Radiology
Stanford University
xiaozhic@stanford.edu

Abstract

Magnetic resonance fingerprinting (MRF) is a novel quantitative imaging technique, while its relatively long scan time make its sensitive to motion artifacts. These artifacts can obscure critical anatomical details, potentially leading to misdiagnosis or necessitating repeated scans. The most conventional method to resolve motion artifacts in MRF involves estimating motion from navigator images and performing motion correction during reconstruction. However, acquiring relatively high-resolution navigator images impairs overall scan efficiency. Therefore, this study aims to improve the performance of motion estimation in MRF using lower-resolution, more scan time-efficient navigator images through a deep learning-based approach. A Convolutional Neural Network (CNN), inspired by ResNet, was designed and trained using an in vivo brain MRF dataset. Preliminary results indicate a better performance in motion estimation compared to traditional algorithms and a reduction of motion artifacts in brain MRF parameter maps. On top the network, data consistency loss and semi-supervised learning were explored to improve the performance and enhance the generability of the approach.

1. Introduction

Magnetic resonance fingerprinting (MRF) is a robust quantitative imaging technique [13]. An example of MRF sequence is presented in Figure 1, where high resolution spiral readout is repeated with varying scan parameters, such as flip angle (FA) and repetition time (TR), to achieve temporal incoherence so that tissues with different tissue parameters, such as T1, T2 and proton density (PD), can exhibit different signal evolution. This signal evolution is the so-called "fingerprint" of the tissue. By matching the acquired "fingerprint" of each image pixel to a pre-calculated

signal dictionary that contain the signal evolution of all possible combination of tissue parameters, we can then retrieve the information at each image pixel and finally generate a tissue parameter maps, including T1 and T2 maps, of the whole object. Due to its robustness and time efficiency compared with conventional quantitative imaging methods, MRF has huge potential in the diagnosis and studies of many neurological disease, such as Parkinson's disease (PD)[9], Alzheimer's disease (AD)[8], multiple sclerosis (MS)[14].

Motion is one of the major challenges in obtaining high quality tissue parameter maps. Due to relatively long scan time, patients' involuntary movement, breath, blood flow will all introduce errors into the acquired data, therefore compromising the accuracy of the resulting "fingerprints" and consequently the estimated parameter maps. Traditional methods address these challenges by acquiring a low-resolution water navigator image by the end of each acquisition group, as shown in Figure 1, and do motion correction for each group by image registration methods. However, the precision of the motion estimation is constrained by the resolution of these motion navigator images. Low-resolution navigator images can lead to inaccuracy of motion parameter estimation, thereby resulting in unresolved motion artifacts, such as blurring and distortion, in the final tissue parameter maps, and, more importantly, inaccurate parameter estimates. On the contrary, acquisition of a high-resolution motion navigator induces extra acquisition time, therefore diminishing the overall scan efficiency of MRF.

To balance the trade-off between acquisition time and precision of motion estimation in MRF, the purpose of this work is to enhance the accuracy of motion estimation derived from low-resolution but more scan time-efficient navigator images. A Convolutional Neural Network (CNN), inspired by ResNet, was designed and trained using an *in vivo* brain MRF dataset. The inputs of our model were low-resolution motion navigator images with and without simulated motion, and the outputs were estimated motion parameters, including translation and rotation along three

*Xiaozhi Cao is a postdoctoral scholar from Department of Radiology who is not enrolled in CS 231N.

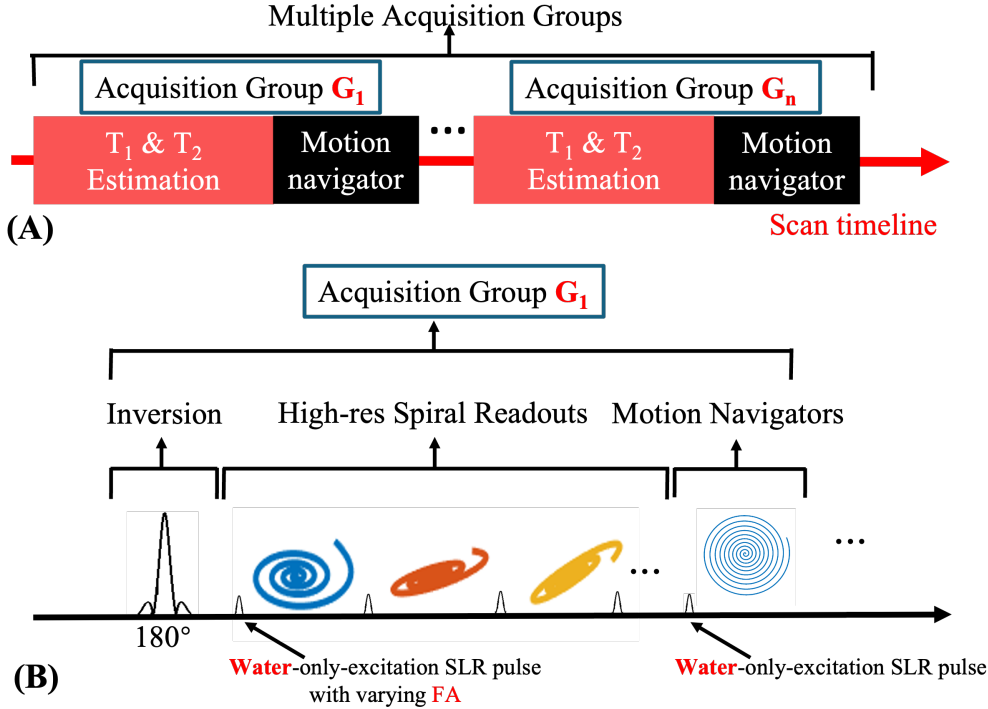


Figure 1. MRF Data Acquisition Pipeline. (A) Illustrates the entire MRF data acquisition process, while (B) zooms in on each acquisition group, displaying the specific acquisition patterns.

directions perspectively.

By obtaining more precise rigid motion parameters, we expect enhancing the overall quality of MRF tissue parameter maps in terms of high-resolution and artifact-free depiction, which can potentially enhance both the clinical feasibility and research significance of MRF. This work can further motivate for self-navigated MRF, wherein low-resolution navigator images could be extracted directly from the high-undersampled MRF data itself and be used to correct its corresponding acquisition group, therefore significantly improve the scan efficiency of MRF.

2. Related Work

2.1. Conventional Motion Correction in MRF

To our knowledge, the application of deep learning-based methods in MRF motion correction is very limited. Conventional MRF relies on specific sequence designs and sampling patterns to achieve motion insensitivity. Only in recent has the use of additional motion navigators to better address motion artifacts.[10]. The previous works have only been done on using image processing technique or optimization for motion estimation from the navigator images [6] [2].

2.2. Deep Learning-Based Motion Correction in MRI

Despite its limited usage in MRF, deep learning-based approaches are widely adopted for motion correction in Magnetic Resonance Imaging (MRI). Similar to MRF, MRI image quality is sensitive to motion during the scan, presenting challenges in clinical diagnosis.

Existing deep learning-based motion correction methods in MRI can be categorized into: 1) image-based approaches and 2) k-space (Fourier domain)-based approaches [20]. Image-based approaches use motion-corrupted images as input with networks trained to detect and reduce the motion artifact, thereby generating cleaner and sharper images. Multiple works have been done using the image-based approach, with different architectures such as fully convolutional network [19], U-Net [16] [12], and encoder-decoder structures [11]. These methods demonstrate promising reductions in motion artifacts. Advantages of image-based methods include flexibility in datasets and the simplicity of models, as the utilization of absolute value images. Training can be accomplished using most open-source MRI datasets or previously acquired images. However, image-based approaches relies on large training datasets, and the absence of a data consistency term raises concerns about network hallucination

Despite the image-based approaches, k-space-based

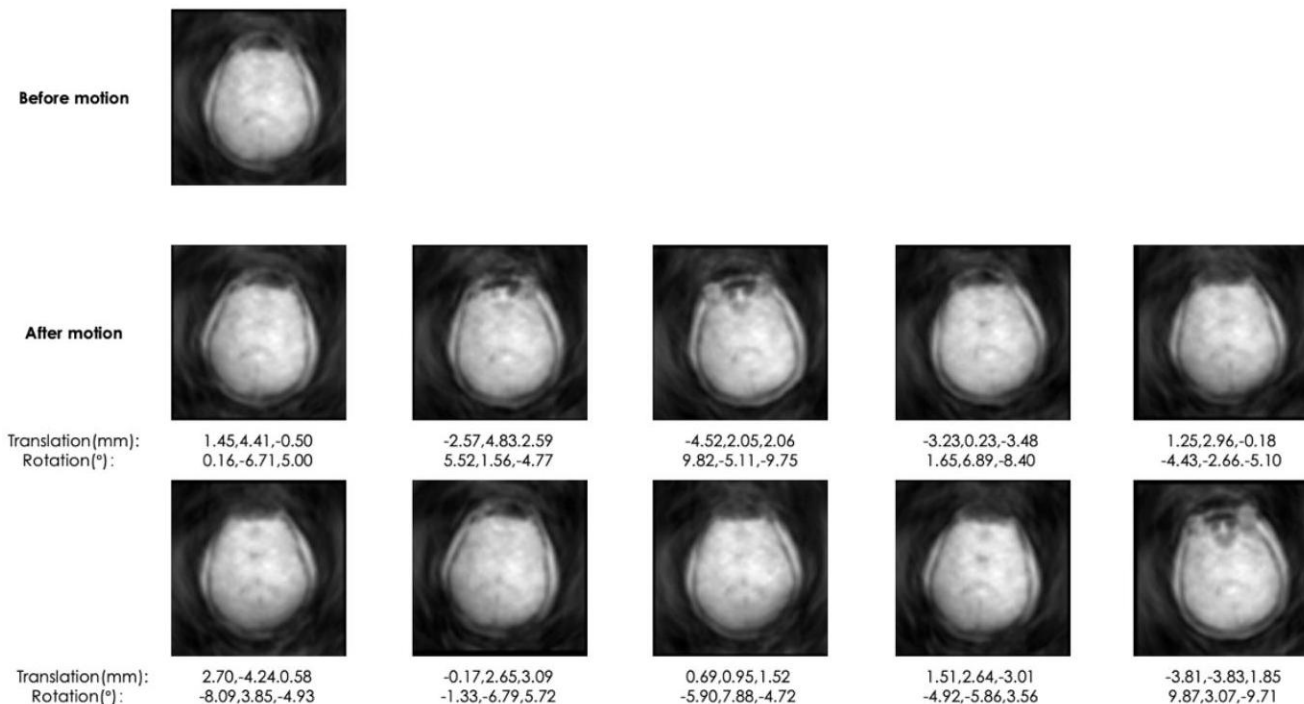


Figure 2. Simulated data by random translation and rotation to original image.

methods are now more widely used. MRI data is acquired in k-space, the Fourier domain of MR images, and properly integrating raw k-space data provides more information, enabling more robust performance. Due to the use of complex-valued k-space data, the output can include both motion-corrected images and rigid motion parameters. In k-space-based approaches, data consistency is often enforced within the models [18] [5] [3] by integrating the reconstruction forward model. Training can involve adding a data consistency loss term or iteratively performing motion correction and image reconstruction. Utilizing data consistency helps mitigate the hallucination effect, which is critical for clinical use.

3. Data

3.1. Data Acquisition and Preprocessing

In vivo brain MRF dataset was collected from 10 healthy volunteers on GE Signa Premier 3.0T system (GE Healthcare, Waukesha, WI) following the same acquisition scheme in Figure 1. The MRF spiral acquisition is 1-mm isotropic resolution with matrix size $256 \times 256 \times 256$. The motion navigator acquisition is 4-mm isotropic resolution with matrix size $64 \times 64 \times 64$. Each acquisition group contains 540 (500+40) MRF spiral readouts. 16 acquisition groups in total were collected in each volunteer scan, resulting in a total scan time of $(500 + 40) \times 16 \times \frac{12.5}{1000} = 108$ seconds.

Motion navigator images were reconstructed following the pipeline of density compensation, gridding to Cartesian coordinates, and 2D inverse Fourier transform. Then, each image was normalized by its maximum value independently.

3.2. Data Augmentation

According to our training task, 500 motion navigator image was randomly selected from the collected MRF dataset. To augment our database, we randomly generated 10 different motion states for each motion navigator image. Each motion state is represented by six motion parameters, translation and rotation along x , y , z respectively. The generated translations range between -5 to 5 mm, and rotations range from -10 to 10 degrees, reflecting the typical extent of motion experienced during real MRF scans. We then simulated navigator images with different motion states using k-space phase modulation and coordinate rotation.

Translation in image space is equivalent to adding phase in k-space, and rotation in image space is equivalent to rotate the k-space, according to basic Fourier Transform properties.

$$s(k_x, k_y, k_z) = s(k_x, k_y, k_z) \cdot e^{i2\pi(k_x \cdot \Delta x + k_y \cdot \Delta y + k_z \cdot \Delta z)}$$

where $s(k_x, k_y, k_z)$ indicates signal collected at specific k-space coordinates, Δx , Δy , Δz are the amounts of translation in the image domain along x , y and z .

$$M' = R_z(\gamma) \cdot R_y(\beta) \cdot R_x(\alpha) \cdot M$$

where M and M' are the k-space coordinates before and after rotation, α, β, γ are rotation angle along x, y and z axis and R is the rotation matrix. Examples of navigator images with simulated motions, with its ground truth image, are presented in Figure 2.

After the motion simulation, we had a total of 5000 datasets, each containing a ground truth image, image with simulated motion, and the corresponding motion parameters. We then randomly split these 5000 datasets into training, validation, and test sets in a ratio of 7:2:1. Mean value calculated on the training dataset was subtracted to all images.

4. Methods

4.1. Baseline Methods

Due to the differences between MRI images and MRF motion navigators in terms of acquisition schemes (e.g., low resolution, non-Cartesian sampling) and the subsequent interpretation of the image information, existing methods of resolving motion artifacts and estimating motion parameters on MRI are not good candidates for the baseline method. In this work, two open-source neuroimage processing toolbox, MCFLIRT [6] and AFNI[2], which are heavily used in research of brain MRF and functional MRI, were used as baseline methods. Both methods follow some optimization algorithms to search for the six rigid motion parameters that are described above. Motion correction were then applied to the navigator images using the estimated parameters.

4.2. Network Structure

Figure 3 shows the structure of the neural network. The input to the network consists of navigator images at different motion states, ground truth motion parameters, and their motion state 1 image, which is used as the reference of motion correction. The output of the convolutional neural network (CNN) is six motion estimation parameters: translation in the $x, y,$ and z directions, and rotation along the $x, y,$ and z axes, respectively.

We frame this as a regression problem. The network is composed of several combinations of convolutional layers, instance normalization, and ELU activation functions. Feature maps are reduced by a factor of two after each convolutional layer, followed by a ResNet block to ensure gradient flow throughout the network. The final output is produced after a convolutional layer with a kernel size of 1, an instance normalization layer, and a Tanh activation layer. Mean Squared Error (MSE) loss is employed, and Adam is used as the parameter optimizer.

Residual Learning Traditional deep networks often suffer from the vanishing gradient problem, where gradients become very small, hindering the ability of the network to

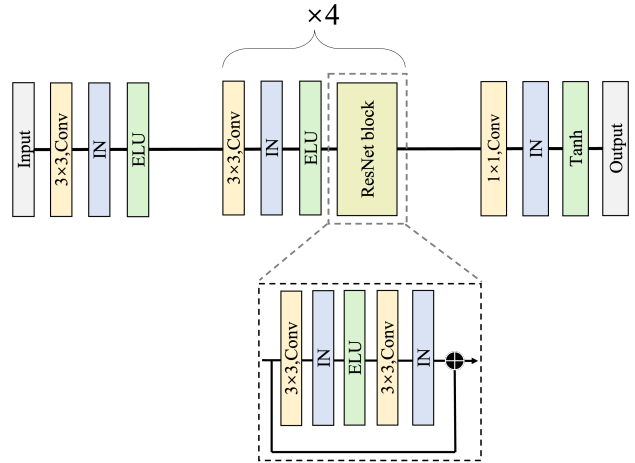


Figure 3. Structure of the Convolutional Neural Network.

learn. We here use ResNet [4] to mitigate this degradation problem by using shortcut connections that skip one or more layers. ResNet’s architecture, with its residual connections, enables the network to learn more complex and abstract features. This incremental learning approach leads to more effective feature extraction and representation.

Data Consistency As mentioned previously, enforcement on data consistency is often used in the application of MRI. The idea behind it is to minimize the difference between the results output by the model and the acquired k-space data. In general, the forward model of MRI reconstruction can be written as

$$y = Ex$$

where x is the reconstructed image and y is the data in k-space. The gridding operator E of the motion navigator with spiral trajectory is

$$E = W * NUFFT$$

W is the density compensation operator and $NUFFT$ represents non-uniform Fast Fourier transform using k-space coordinates.

As discussed above, there are many ways of achieve data consistent result in deep-learning based methods. For the consideration of training efficiency, we followed the idea in [18, 3]. No change was made during the training. At the test time, we used the output of the CNN and optimized the motion parameters by

$$\operatorname{argmin}_{\theta} \|y - EA(\theta)x\|_2^2$$

where $A(\theta)$ represent translation and rotation correction with given motion parameters θ .

The minimization is performed using a quasi-Newton search available with the built in *fminunc* function in

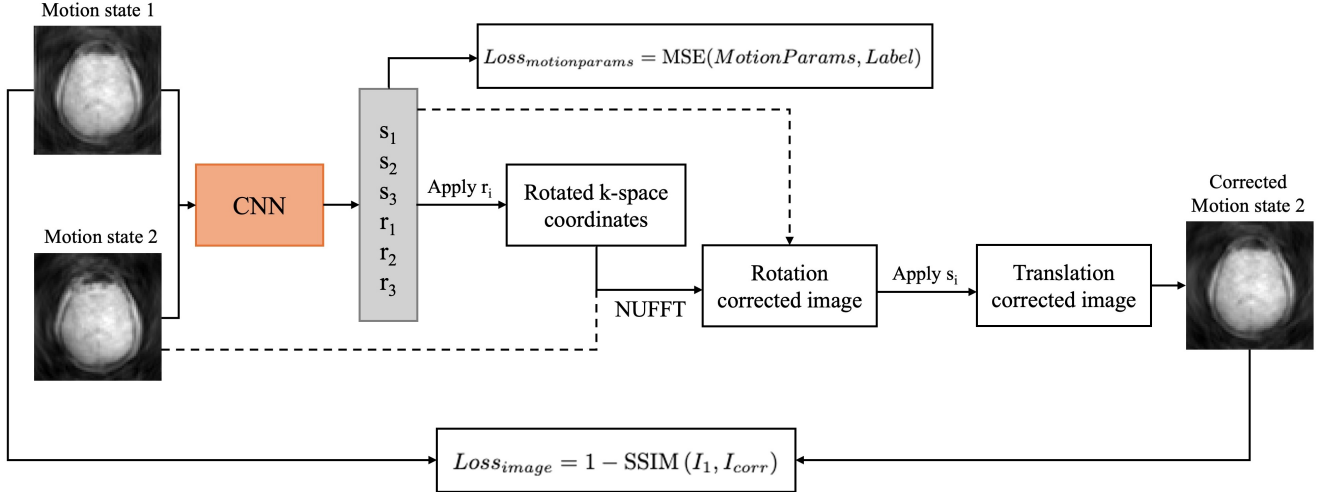


Figure 4. The network structure used for semi-supervised learning

MATLAB (Mathworks, Natick, MA). Motion parameters generated by CNN was used as initialization.

Semi-Supervised Learning In the medical imaging field, obtaining actual clinical data is often challenging. To maximize the use of limited datasets, semi-supervised learning (SSL) is increasingly utilized [15] in the tasks such as image segmentation [7] and reconstruction [21]. Here we also explored SSL in MRF motion estimation.

The network structure used in this study is shown in Figure 4. We employ the same CNN architecture as in the residual learning supervised network. Rotation parameters estimated by the network were applied to obtain a rotated k-space grid, followed by an NUFFT operation to correct the rotation for the motion state 2 image. Translation parameters were then applied to correct shifts, which should align perfectly with the motion state 1 image, despite some minor boundary artifacts.

The loss is expressed as follows,

$$Loss = 1 - SSIM(I_1, I_{corr}) + \lambda \cdot MSE(MotionParams, Label)$$

Mean square error (MSE) was used for the supervised training of motion parameters. For the unsupervised image loss, Structural Similarity Index (SSIM) was used instead of MSE. Unlike MSE, which merely measures pixel-wise differences, SSIM evaluates the structural similarity between images, this means that SSIM is more sensitive to changes in textures and edges, making it better suited for maintaining visual quality under such conditions. These factors contribute to higher perceptual quality and more effective image comparison. The second loss term, which compares network estimates and labels, is balanced by tuning $\lambda = 100$, yielding improved results.

To mimic the scenario where limited number of dataset was given, we trained our model on 500 datasets, which is

only one-tenth of the training data size used in supervised residual learning cases.

5. Experiments and Results

5.1. Hyperparameter Tuning

To enhance the robustness of our model and reduce overfitting, Weights and Biases (wandb) [1] was applied to tune the parameters such as batch size, learning rate, weight decay, and the number of channels in each convolutional layer. We conducted 100 sweeps, randomly sampling different hyperparameter pairs within reasonable ranges. For each sweep, we trained the model for 100 epochs. As shown in Figure 5, batch size and learning rate were the two most important hyperparameters in our case. The number of channels in the first few convolutional layers had a slightly greater influence on model performance compared to the last few layers. After 100 sweeps, we selected the best hyperparameter pair, named "wovensweep-5," based on the validation loss, and used for further training.

5.2. Training Results

We trained the model selected after hyperparameter tuning for 2000 epochs. As shown in Figure 6, the validation loss trend aligns well with the training loss, and the test loss is 0.003, which falls within the range of the training and validation losses. This indicates good model performance and minimal overfitting.

To compare our method with conventional motion estimation approaches, we calculated the mean and standard deviation of the absolute error on the test data, as shown in Table 1. Our deep learning approach outperformed MCFLIRT and AFNI in every motion parameter estimation. Our network demonstrates a strong capability to automati-

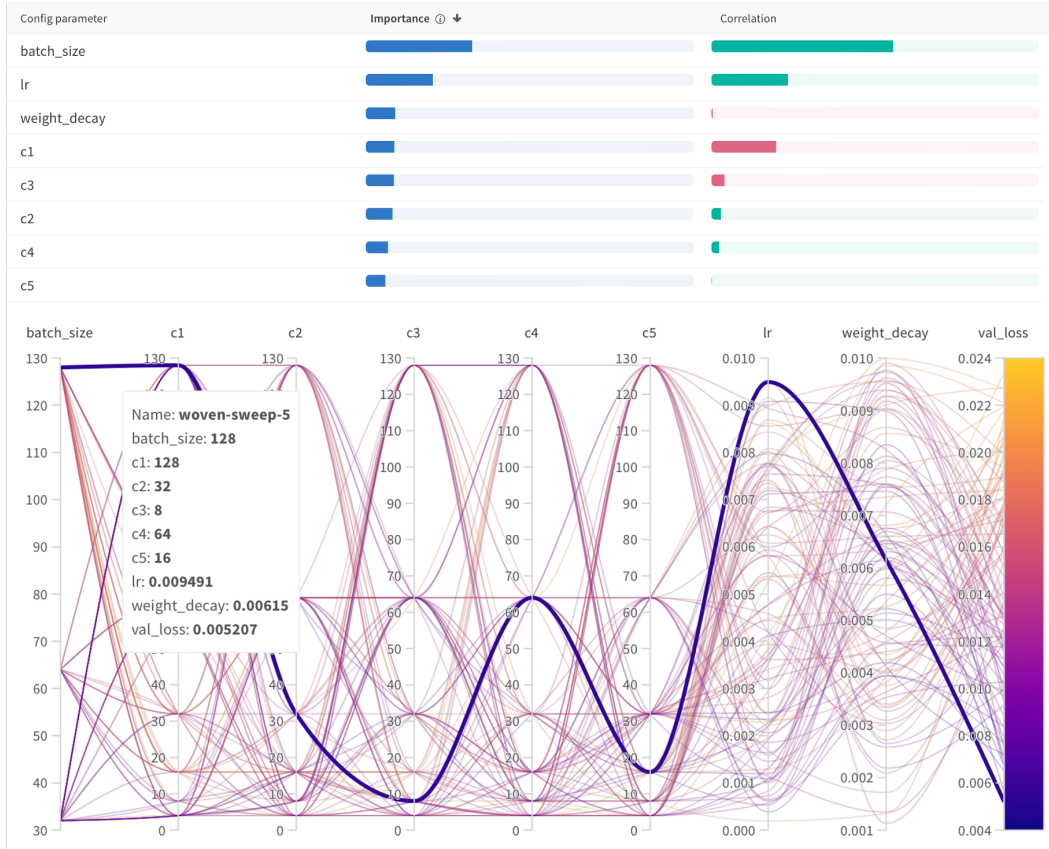


Figure 5. Hyperparameter Tuning of the propose CNN. The upper image displays the importance and correlation of different hyperparameters in minimizing validation loss, while the lower image shows the combination of different hyperparameters and the resulting validation loss.

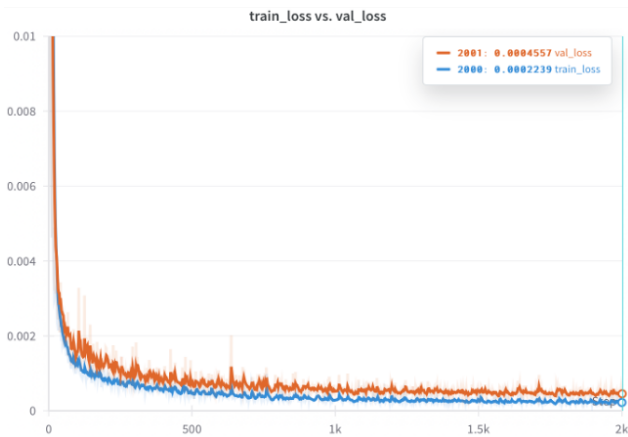


Figure 6. Training Procedure.

cally learn and extract hierarchical features from the data, capturing both local and global patterns. This allows it to model complex, non-linear relationships effectively, making it highly proficient in understanding and predicting motion in various contexts. Adding the optimization of motion parameters with data consistency loss can slightly improve

the performance of the motion estimation on both mean and standard deviation. The SSL approach gave higher errors on the estimated motion parameters, compared to the residual learning CNN. However, it still outperformed the two conventional methods. Considering the very limited data used, the result is still promising and indicating the feasibility of SSL in this task.

To further visualize the motion estimation results, we compared the performance of MCFLIRT, AFNI, and our proposed method in estimating motion between two given motion states, as shown in Figure 7. We used the estimated motion parameters to transform the motion state 2 images back to motion state 1. To prevent from too many duplicated images, only the results of the residual learning CNN are shown here. The results demonstrate that, using the motion parameters estimated by our method, the motion state 2 images were accurately transformed back to motion state 1, showing the good performance of our approach.

5.3. Motion Correction in MRF Parameter Maps

To further evaluate the robustness of our method, we applied our model and the baseline methods to the postpro-

	Absolute Translation Error (mm)			Absolute Rotation Error ($^{\circ}$)		
MCFLIRT	0.64 ± 0.44	0.47 ± 0.31	0.46 ± 0.33	0.32 ± 0.23	0.41 ± 0.30	0.32 ± 0.21
AFNI	0.23 ± 0.18	0.23 ± 0.16	0.21 ± 0.16	0.24 ± 0.21	0.43 ± 0.31	0.46 ± 0.41
RL	0.15 ± 0.12	0.19 ± 0.15	0.15 ± 0.13	0.16 ± 0.11	0.22 ± 0.17	0.21 ± 0.18
RL + DC	0.11 ± 0.08	0.14 ± 0.12	0.15 ± 0.10	0.14 ± 0.12	0.22 ± 0.14	0.15 ± 0.11
SSL	0.24 ± 0.16	0.25 ± 0.16	0.27 ± 0.15	0.22 ± 0.21	0.31 ± 0.23	0.27 ± 0.19

Table 1. Absolute Mean and Standard Deviation of the Error in x, y, z Translation and Rotation Estimation.

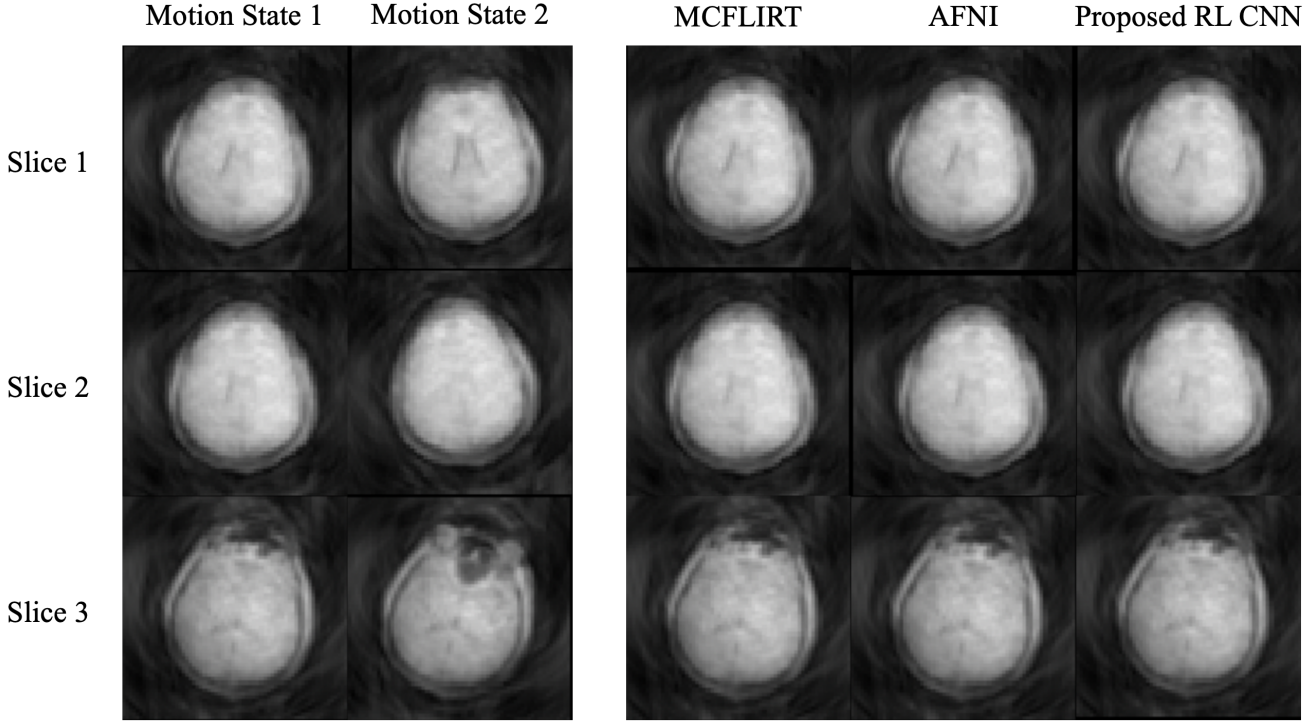


Figure 7. Estimation of Motion Parameters Using Different Approaches. Motion parameters are estimated from state 1 to state 2 and used to convert state 2 images back to state 1.

cessing of a brain MRF data. Given the motion navigator images from each acquisition group, we estimated the motion parameters relative to the first acquisition group using our network. These estimated motion parameters were then used to correct the acquired MRF data for each group, followed by SENSE image reconstruction [17] and MRF dictionary matching [13].

Figure 8 shows the quantitative parameter maps (T1, T2, and PD) generated without motion correction, with conventional motion estimation, and with our proposed methods. Conventional motion correction methods select the best correction results using traditional motion estimation approaches. The parameter maps corrected using our proposed method show significant reduction of motion artifacts and clear delineation of tissue structures, particularly near the cerebrospinal fluid and gray matter areas indicated by red arrows.

6. Discussion and Conclusion

In this work, we explored the potential of using deep learning-based methods to estimate and correct motion in MRF. A CNN architecture was proposed and trained, which demonstrated superior performance compared to two conventional methods. The accuracy of the estimated motion can be further improved by enforcing data consistency at test time, which provides a slight enhancement in motion estimation.

Given the limited size of the simulated training data and the absence of labeled navigator images in real-world scenarios, we also investigated the feasibility of semi-supervised learning approaches to enhance the performance and generality of the model. With a very small training dataset, our SSL shows promising results.

One of the limitations of this method is its reliance on labeled data for guidance during training. An interesting angle for future research involves using more advanced net-

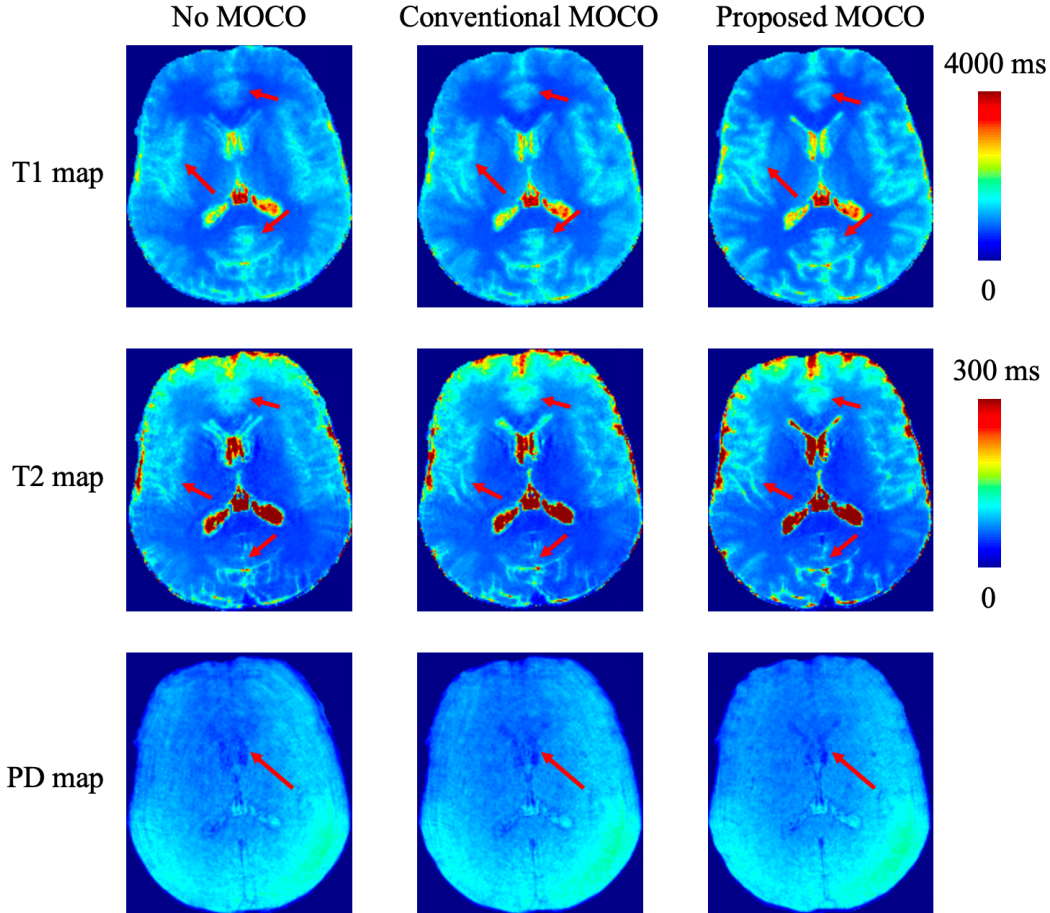


Figure 8. Quantitative Parameter Maps Before and After Motion Correction (MOCO).

work architectures, such as transformers or pretrained models, to eliminate the need for supervised loss and transition to a self-supervised learning paradigm. This shift would be particularly beneficial in scenarios where abundant unlabeled data is available, further enhancing the applicability and effectiveness of the model.

For the purpose of this project, all the motion-corrupted images were simulated in order to acquire the ground truth motion parameters. However, in realistic cases, patient motion often includes a mix of rigid and non-rigid movements, and these are further complicated by other hardware imperfections and field inhomogeneity during scanning. All these factors might bring significant challenges to the robustness and applicability of the proposed method. Therefore, further work can be done by evaluating the motion correction performance on a large amount of *in vivo* case and improve the generality of the methods.

7. Contributions and Acknowledgements

Mengze Gao developed and trained the CNN and semi-supervised learning model, and wrote the paper. Xuotong Zhou developed the data consistency loss and wrote the paper.

We thank Dr. Xiaozhi Cao for his assistance in MRI data collection and for his contributions to our discussions. His support and expertise were crucial to the successful completion of this project.

References

- [1] Weights & biases. <https://docs.wandb.ai/company/academics#cite-weights-and-biases>.
- [2] R. W. Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.
- [3] M. W. Haskell, S. F. Cauley, B. Bilgic, J. Hossbach, D. N. Splitthoff, J. Pfeuffer, K. Setsompop, and L. L. Wald. Network accelerated motion estimation and reduction (namer): convolutional neural network guided retrospective motion

- correction using a separable motion model. *Magnetic resonance in medicine*, 82(4):1452–1461, 2019.
- [4] 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*, pages 770–778, 2016.
- [5] J. Hossbach, D. N. Splitthoff, S. Cauley, B. Clifford, D. Polak, W.-C. Lo, H. Meyer, and A. Maier. Deep learning-based motion quantification from k-space for fast model-based magnetic resonance imaging motion correction. *Medical physics*, 50(4):2148–2161, 2023.
- [6] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.
- [7] R. Jiao, Y. Zhang, L. Ding, B. Xue, J. Zhang, R. Cai, and C. Jin. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Computers in Biology and Medicine*, page 107840, 2023.
- [8] V. C. Keil, S. P. Bakoeva, A. Jurcoane, M. Doneva, T. Amthor, P. Koken, B. Mädler, W. Block, R. Fimmers, K. Fliessbach, et al. Mr fingerprinting as a diagnostic tool in patients with frontotemporal lobe degeneration: a pilot study. *NMR in Biomedicine*, 32(11):e4157, 2019.
- [9] V. C. Keil, S. P. Bakoeva, A. Jurcoane, M. Doneva, T. Amthor, P. Koken, B. Mädler, G. Lüchters, W. Block, U. Wüllner, et al. A pilot study of magnetic resonance fingerprinting in parkinson’s disease. *NMR in Biomedicine*, 33(11):e4389, 2020.
- [10] J. W. Kurzwaski, M. Cencini, L. Peretti, P. A. Gómez, R. F. Schulte, G. Donatelli, M. Cosottini, P. Cecchi, M. Costagli, A. Retico, et al. Retrospective rigid motion correction of three-dimensional magnetic resonance fingerprinting of the human brain. *Magnetic Resonance in Medicine*, 84(5):2606–2615, 2020.
- [11] J. Lee, B. Kim, and H. Park. Mc2-net: motion correction network for multi-contrast brain mri. *Magnetic Resonance in Medicine*, 86(2):1077–1092, 2021.
- [12] J. Liu, M. Kocak, M. Supanich, and J. Deng. Motion artifacts reduction in brain mri by means of a deep residual network with densely connected multi-resolution blocks (drn-dcmb). *Magnetic resonance imaging*, 71:69–79, 2020.
- [13] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold. Magnetic resonance fingerprinting. *Nature*, 495(7440):187–192, 2013.
- [14] T. R. Mostardeiro, A. Panda, N. G. Campeau, R. J. Witte, N. B. Larson, Y. Sui, A. Lu, and K. P. McGee. Whole brain 3d mr fingerprinting in multiple sclerosis: a pilot study. *BMC medical imaging*, 21(1):88, 2021.
- [15] Y. Ouali, C. Hudelot, and M. Tami. An overview of deep semi-supervised learning. *arXiv preprint arXiv:2006.05278*, 2020.
- [16] K. Pawar, Z. Chen, J. Seah, M. Law, T. Close, and G. Egan. Clinical utility of deep learning motion correction for t1 weighted mprage mr images. *European Journal of Radiology*, 133:109384, 2020.
- [17] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.
- [18] N. M. Singh, N. Dey, M. Hoffmann, B. Fischl, E. Adalsteinsson, R. Frost, A. V. Dalca, and P. Golland. Data consistent deep rigid mri motion correction. In *Medical Imaging with Deep Learning*, pages 368–381. PMLR, 2024.
- [19] K. Sommer, A. Saalbach, T. Brosch, C. Hall, N. Cross, and J. Andre. Correction of motion artifacts using a multiscale fully convolutional neural network. *American Journal of Neuroradiology*, 41(3):416–423, 2020.
- [20] V. Spieker, H. Eichhorn, K. Hammernik, D. Rueckert, C. Preibisch, D. C. Karampinos, and J. A. Schnabel. Deep learning for retrospective motion correction in mri: a comprehensive review. *IEEE Transactions on Medical Imaging*, 2023.
- [21] M. Yurt, O. Dalmaz, S. Dar, M. Ozbey, B. Tınaz, K. Oguz, and T. Çukur. Semi-supervised learning of mri synthesis without fully-sampled ground truths. *IEEE Transactions on Medical Imaging*, 41(12):3895–3906, 2022.