DeepRetroMoCo

Mobarak-Abadi, Mahdi; Mahmoudi-Aznave, Ahmad; Dehghani, Hamed; Zarei, Mojtaba; Vahdat, Shahabeddin; Doyon, Julien ; Khatibi, Ali

DOI:
10.1101/2022.09.06.506787

DeepRetroMoCo: Deep neural network-based Retrospective Motion Correction Algorithm for Spinal Cord functional MRI

Mahdi Mobarak-Abadi [a], Ahmad Mahmoudi-Aznave [a], Hamed Dehghani [b], Mojtaba Zarei [a], Shahabeddin Vahdat [c], Julien Doyon [d], Ali Khatibi [e,f*]

a Institute of Medical Science and Technology Shahid Beheshti University, Velenjak, Tehran, Iran

b Neuro Imaging and Analysis Group (NIAG), Research Center for Molecular and Cellular Imaging (RCMCI), Tehran University of Medical Sciences, Tehran, Iran

c Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, USA

d McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

e Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), University of Birmingham, UK

f Centre for Human Brain Health, University of Birmingham, UK

* Corresponding Author

Ali Khatibi, PhD

Centre of precision Rehabilitation for Spinal Pain

University of Birmingham, Birmingham, UK

m.khatibitabatabaei@bham.ac.uk

Abstract

There are unique challenges in the preprocessing of spinal cord fMRI data, particularly voluntary or involuntary movement artifacts during image acquisition. Despite advances in data processing techniques for movement detection and correction, there are challenges in extrapolating motion correction algorithm developments in the brain cortex to the brainstem and spinal cord. We trained a Deep Learning-based convolutional neural network (CNN) via an unsupervised learning algorithm, called DeepRetroMoCo, to detect and correct motions in axial T2*-weighted spinal cord data. Spinal cord fMRI data from 27 participants were used for training of the network (135 runs for training and 81 runs for testing). We used average temporal signal-to-noise-ratio (tSNR) and Delta Variation Signal (DVARS) of raw and motion-corrected images to compare the outcome of DeepRetroMoco with sct_fmri_moco implemented in the spinal cord toolbox. The average tSNR in the cervical cord was significantly higher when DeepRetroMoco was used for motion correction compared to sct_fmri_moco method. Average DVARS was lower in images corrected by DeepRetroMoco than those corrected by sct_fmri_moco. The average processing time for DeepRetroMoco was also significantly shorter than sct_fmri_moco. Our results suggest that DeepRetroMoCo improves motion correction procedures in fMRI data acquired from the cervical spinal cord.

Keywords: fMRI; Spinal Cord; Motion Correction; Deep Learning, Unsupervised

1. Introduction

Spinal cord functional magnetic resonance imaging (fMRI) has become increasingly popular for exploring intrinsic neural networks and their role in pain modulation, motor learning and sexual arousal (Alexander et al., 2016; Kinany et al., 2019). There are unique challenges in data acquisition and preprocessing, such as relatively small cross-sectional dimension, the variable articulated structure of the spine between individuals, low signal intensity in standard gradient-echo echo-planar T2*-weighted fMRI and voluntary (bulk motion) or involuntary (fluctuation of cerebrospinal fluid due to respiration and heartbeat) movements during image acquisition (Dehghani, Oghabian, Batouli, Arab Kheradmand, & Khatibi, 2020; Kinany et al., 2022; Powers, Ioachim, & Stroman, 2018). Spinal cord motions can cause signal alterations across volumes, which decrease the temporal stability of the signal and ultimately increase false positive and negative discovery rates (Cohen-Adad, Piche, Rainville, Benali, & Rossignol, 2007; Dehghani, Weber, Batouli, Oghabian, & Khatibi, 2020; Stroman, Figley, & Cahill, 2008).

Despite advances in fMRI motion correction, there are problems in extrapolating the motion correction algorithm developments in the brain to the brainstem and spinal cord. In brain fMRI, we generally utilize six degrees of freedom rigid-body registration of a single volume to a reference, which can be a preselected volume or an average volume (Maknojia, Churchill, Schweizer, & Graham, 2019; Oakes et al., 2005). This method is non-robust and insufficient for spinal cord fMRI preprocessing due to the non-rigid motion of the spinal column and physiological motion from swallowing and the respiratory cycle (Dehghani, Oghabian, et al., 2020; Fratini, Moraschi, Maraviglia, & Giove, 2014). Along with the release of the Spinal Cord Toolbox (SCT), sct_fmri_moco was introduced for motion correction in the spinal cord (De Leener et al., 2017). The basis of sct_fmri_moco is slice-by-slice regularized registration for spinal cord algorithm (SliceReg) that estimates slice-by-slice translations of axial slices while ensuring regularization constraints along the z-axis (Paquin et al., 2018).

In the past few years, we have seen an interest in the application of artificial intelligence in medical image processing (Anaya-Isaza, Mera-Jiménez, & Zequera-Diaz, 2021; Varoquaux & Cheplygina, 2022; Wen et al., 2018). In spinal cord imaging, deep learning has been used for the segmentation of the spinal cord and CSF in structural T1 and T2 weighted images. DeepSeg as a fully automated framework based on convolutional neural networks (CNNs), is proposed to apply spinal cord morphometry for segmenting the spinal cord, as part of SCT (Gros et al., 2019; Perone, Calabrese, & Cohen-Adad, 2018; Prados et al., 2017). More recently, the K-means clustering algorithm has been used for the segmental spinal cord in the thoracolumbar region (Sabaghian, Dehghani, Batouli, Khatibi, & Oghabian, 2020). A robust and automated CNN model with two temporal convolutional layers is introduced for motion correction in brain fMRI, and the following regression employs derived motion regressors. (Yang, Zhuang, Sreenivasan, Mishra, & Cordes, 2019)

Studies in the field of registration are generally divided into two categories: learning-based and non-learning based. In the non-learning category, extensive work has been done in the field of 3D medical image registration (Ashburner, 2007; Avants, Epstein, Grossman, & Gee, 2008; Bajcsy & Kovačič, 1989; Dalca, Bobu, Rost, & Golland, 2016; Sokooti, Saygili, Glocker, Lelieveldt, & Staring, 2016; Thirion, 1995). Some models are based on optimizing the field space of displacement vectors, which include elastic models (Bajcsy & Kovačič, 1989; Kybic & Unser, 2003), statistical parametric mapping (Ashburner, Andersson, & Friston, 2000) , free-form deformations with b-spline (Ashburner et al., 2000), and demons(Thirion, 1995). Common formulations include Large Diffeomorphic Distance Metric Mapping (LDDMM) (Ceritoglu et al., 2010; Risser et al., 2011), DARTEL (Ashburner, 2007) and standard symmetric normalization (SyN) (Avants et al., 2008). There are several recent articles in learning-based studies that have suggested neural networks for registering medical images, and most of them require ground truth data or any additional information such as segmentation results (Krebs et al., 2017;

Rohé, Datar, Heimann, Sermesant, & Pennec, 2017; Sokooti et al., 2017; Yang, Bian, Yu, Ni, & Heng, 2017)

To the best of our knowledge, no prior study utilized AI for motion correction in the spinal cord fMRI. This study aimed to train a deep learning-based CNN via unsupervised learning to detect and correct motions in axial T2*-weighted spinal cord data. We hypothesize that our method can improve the outcome of motion correction and reduces the preprocessing time as compared to the existing methods.

2. Material and Methods

2.1 Methods

2.1.1. Unsupervised deep learning network architecture

Convolutional Neural Network (CNN) architecture:

Assume *M* and *F* are two images of the same slice defined in the *N-dimensional* spatial domain $\Omega \subset R^N$. We are focusing on *N = 2* because the type of data we are using is "functional," containing single-channel grayscale data. Additionally, our network focuses on the Axial view. The fixed image *F* is the reference volume, so it can be the first, middle, average, or any of the volumes, and *M* is the rest of the time-series images. Before training the network, we align *F* and *M* using our fixing Centerline method, which we describe in the following section, so that the only misalignment between the volumes is nonlinear. We then use a convolutional neural network structure similar to UNet (Isola, Zhu, Zhou, & Efros, 2017; Ronneberger, Fischer, & Brox, 2015) to model a $N_\theta(F, M) = \emptyset$ function, which includes an encoder and decoder with skip connection (Figure 1): where $\emptyset$ is the register map between the two input images and the θ learned parameter of the network. In this map, for each voxel $p \in \Omega$, there is a position where *F(p)* and the warped image $M(\emptyset(p))$ have the same anatomical position. Therefore, our network takes the concatenated images *F* and *M* as input and calculates the registration flow field based on

θ. In the next step, it uses the spatial transformation operator to warp the moving image based on the flow field and evaluates the similarity between *M* and *F* and θ update. Figure 3 shows our introduced architecture and an integrated input by concatenating *F* and *M* in two channels of the 2D image.

In both the encoder and decoder stages, we use two-dimensional convolution with a *3x3* kernel size and leaky Relu activation. The hierarchical properties of the concatenated image pair are captured by the convolution layer, which is required to estimate $\emptyset$. We also use stride convolution to decrease the spatial dimensions and get to the smallest layer. During the encoding steps, features are extracted by down sampling, and during the decoding and up sampling steps, the network propagates the trained features from the previous step directly to the layer that generates the registry by using a skip connection. A decoder's output size $(\emptyset)$ is equal to the input image *M*.

We used two architectures to examine a trade-off between speed and accuracy. These two structures are DRM_1 and DRM_2, which differ in size at the end of the decoder. DRM_1 uses more layers at the end of the decoder, and more channels are used throughout the model (Figure 1).

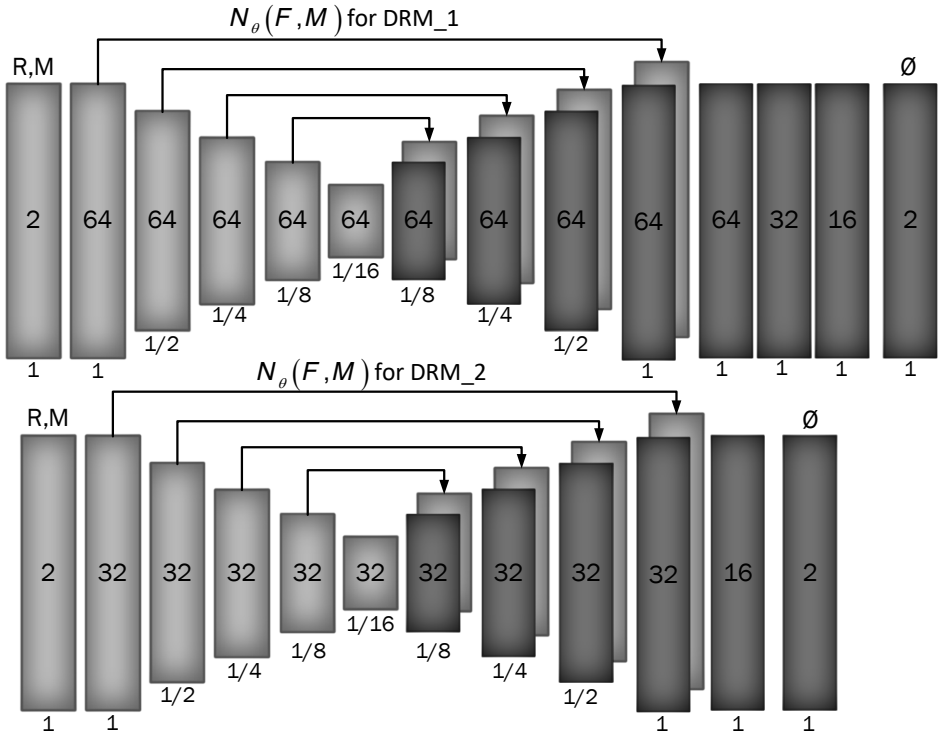

Figure 1- Proposed convolutional architectures implementing $g_\theta(F, M)$. Each rectangle shows a 2D volume in which two fixed and moving images are connected. The number of channels inside each rectangle is shown and the spatial resolution is printed below it according to the input volume. The first model has a larger architecture and more channels than the second model.

To find the optimal theta parameter, we used the stochastic gradient descent method to minimize the loss function $\mathcal{L}$:

$$\hat{\theta} = argmin_\theta \left[E_{(F,M)\sim D} \left[\mathcal{L}\left(F, M, g_\theta(F,M)\right) \right] \right] \quad (1)$$

Where *D* is the data scatter. It should be noted that we do not need supervisor information such as Atlas or T1 images.

The $\mathcal{L}_{Unsupervised}$ consists of two parts: $\mathcal{L}_{sim}$, which measures the similarity between *F* and $M(\emptyset)$, and $\mathcal{L}_{reg}$, which measures the smoothness of the registration field. So, our total loss function is as follows:

$$\mathcal{L}(F, M, \emptyset) = \mathcal{L}_{sim}\left(F, M(\emptyset)\right) + \lambda \mathcal{L}_{reg}\left(\emptyset\right) \quad (2)$$

And λ is the regulation parameter.

We used two different cost functions for $\mathcal{L}_{sim}$: mean square error and normalize cross-correlation, which is a common metric due to robust intensity variations. The first cost function, the mean square error is as follows:

$$MSE = \frac{1}{(Image\ sigma)^2} \times \frac{1}{N} \times \sum_{p_i \in \Omega} \left(F(p_i) - M\left(\emptyset(p_i)\right)\right)^2$$

Here p_i is the position of the pixels and Image sigma is equal to one in this work. Also, the fact that MSE is close to zero indicates better alignment. The second cost function, normalizing cross-correlation, is as follows:

$$CC = \sum_{p \in \Omega} \frac{\left(\sum_{p_i}\left(F(p_i) - \hat{F}(p)\right)\left(M(\emptyset(p_i)) - \hat{M}(\emptyset(p))\right)\right)^2}{\left(\sum_{p_i}\left(F(p_i) - \hat{F}(p)\right)\right)\left(\sum_{p_i}\left(M(\emptyset(p_i)) - \hat{M}(\emptyset(p))\right)\right)} \quad (3)$$

Let $F(p_i)$ and $M\left(\emptyset(p_i)\right)$ be the image intensities of fixed and moving images respectively, and $\hat{F}(p)$, $\hat{M}\left(\emptyset(p)\right)$ be the local mean at position p, respectively. The local mean is computed over a local n^2 window centered at each position p with *n=9* in this work. Therefore, p_i represent the position within n^3 local windows centered at *p*.

By minimizing $\mathcal{L}_{sim}$, we seek to approximate $M\left(\emptyset(p)\right)$ from $F(p)$, but it may cause a discontinuity in $\emptyset$, so we used Spatial gradients to regulate the deformation field between the voxel's neighborhood, as follows:

$$\mathcal{L}_{regularization} = \sum_{p \in \Omega} \|\nabla \emptyset(p)\|^2 = \sum \left\|\frac{\partial \emptyset}{\partial x}\right\|^2 + \left\|\frac{\partial \emptyset}{\partial y}\right\|^2 + \left\|\frac{\partial \emptyset}{\partial z}\right\|^2 \quad (4)$$

This cost function is applied to the network's output vectors and controls the size of the vectors by deriving the vectors in each direction.

Spatial Transformation Function:

The Spatial Transformation Function method learns the optimal parameter by minimizing the difference between the warped image and the reference image. In this way it learns

the transformation from the training network and creates a sample grid by the predicted transformation parameters θ, which is a set of points that samples the input image to give the output image (Jaderberg, Simonyan, Zisserman, & Kavukcuoglu, 2015). Therefore, to perform a spatial transformation on the input image, the Spatial Transformation Function sampler must take a number of sample points of moving image *M* to produce the transformed image $M(\emptyset)$. Also, because the $M(\emptyset)$ image is only defined in integer locations, we perform bilinear interpolation in 8 neighboring voxels for the sampler.

2.1.2. Fixing centerline as preprocessing

Before predicting with the network, we align the data in each slice over time using a centerline in the spinal cord. We used the spinal cord toolbox to extract the centerlines, but we had to use interpolation to adjust the points because some of them were outside the anticipated range or were missing. We organized the centerline coordinates by utilizing third-degree b-spline interpolation after identifying the outliers using the interquartile range approach to discover the lower and upper boundaries of the centerline coordinates.

In this step, data is only corrected in 2D, *x* and *y*, and we can choose one or both dimensions (*x*, *y*, both) for alignment in this approach (

Figure 2). we chose to only correct along the y-direction due to two reasons: 1) we observed that most of the movements in the Axial view of spinal cord data are in the y-direction, also consistent with more distortions along the phase encoding direction. 2) using numerical analysis of the coordinates of the center lines in the x and y directions based on the specific reference, we reached the average variance of 0.52 for x and 1.1 for y, so the scattering of the middle coordinates for the x-direction is about twice less than the y-direction.

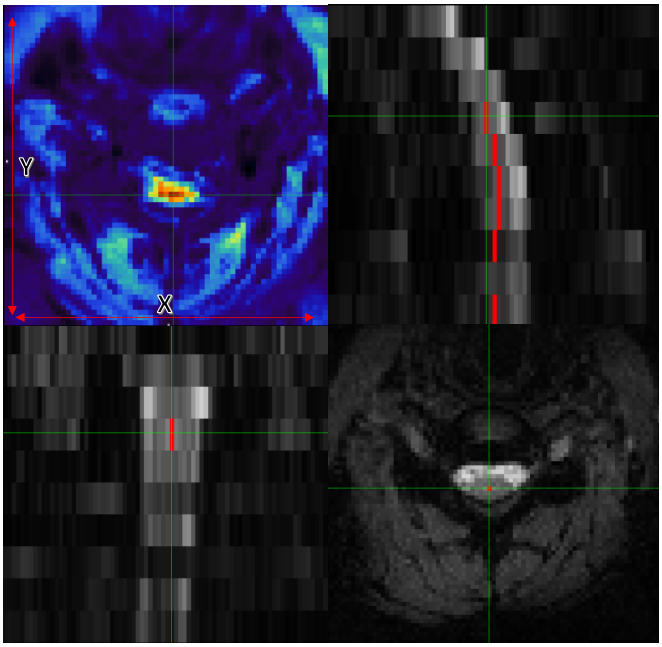

Figure 2- Axial (bottom right), Coronal (bottom left), and Sagittal (top right) views of data with the centerline. The image on the top left also shows tSNR with the x and y guidelines.

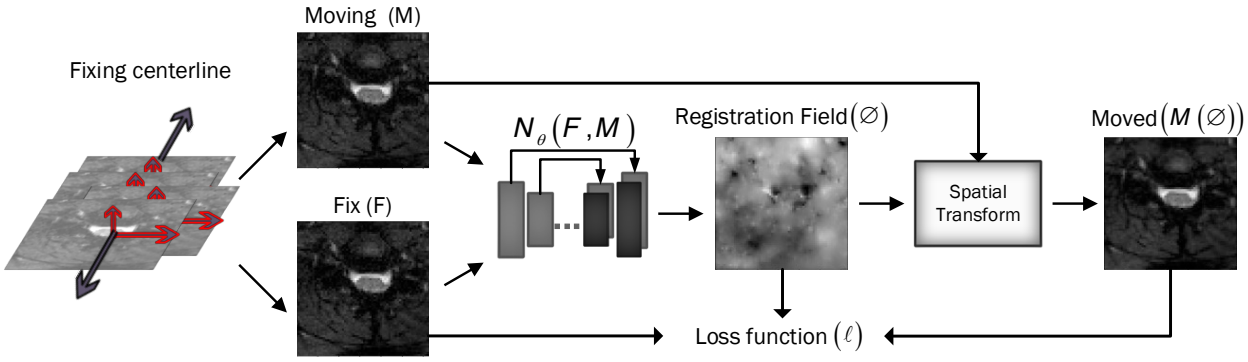

Figure 3- Overview of DeepRetroMoco. As a preprocess, we align the data in two dimensions based on the centerline, and then we register the moving image (*M*) to the fixed image (*F*) by learning function parameters (N_θ). During training, an ST was used to warp the moving image with the registration field, and in this operation, the loss function compares $\mathbf{M}(\emptyset)$ and *F* using the smoothness of $\emptyset$.

2.2. Experiments

2.2.1. Dataset

The data used for this experiment includes 30 subjects with T2*-weighted MRI scans acquired from 3T TIM Trio Siemens scanner (Siemens Healthcare, Erlangen, Germany) equipped with a 32-channel head coil and a 4-channel neck coil was used for the imaging to investigate the functional activity in the brain and the spinal cord (Khatibi et al., 2022). All subjects were scanned twice. Five runs were collected in the first session and three runs in the second session. Sessions were acquired one week apart. This resulted in 240 runs. We only used the data from the neck coil and cervical spinal cord in this study.

The dataset included 8–10 slices that covered the cervical spinal cord from C3 to T1 spinal segmental levels and were orientated parallel to the spinal cord at the C6 level. The FoV of the slices was $132 \times 132\ mm^2$, with voxel sizes of $1.2 \times 1.2 \times 5\ mm^3$ and a 4 mm gap between them. The flip angle was 90 °, and the bandwidth per pixel was 1263 Hz, resulting in an echo spacing of 0.90 ms. 7/8 partial Fourier and parallel imaging (R = 2, 48 reference lines) was utilized again, resulting in a 43.3 ms echo train length and a 33 ms echo time. Finally, the TR for all slices was 3140 ms, with the exception of three subjects, who had TRs of 3050 ms or 3200 ms (depending on each participant's coverage within the field of view). We eliminated the lost data due to low quality and differences in data points when compared to other data, and selected data from 27 subjects and 216 runs: 135 runs for training and a total of 81 runs for testing datasets. For training and validation, the training data was divided by 70 to 30%. The data from the validation part was utilized to select and assess the proposed models.

2.2.2. Evaluation

Since there is no gold standard for direct evaluation of functional registration or motion correction performance, we used two functional measures to check the signal strength of

each subject or to examine signal variations in the group of volumes after predicting them by the network.

1-temporal signal to noise ratio (tSNR):

Temporal SNR (tSNR) is used to quantify the stability of the BOLD signal time series and is calculated by dividing the mean signal by the standard deviation of the signal over time.

$$tSNR = \frac{\bar{S}}{\sigma_{t\_noise}} \quad (5)$$

where $\bar{S}$ is the mean signal over time and σ is the standard deviation across time. A better motion correction algorithm will result in greater tSNR values by reducing signal variations in the BOLD time-series due to motion.

2-DVARS:

DVARS (D, temporal derivative of time courses, VARS, variance over voxels) shows the signal rate changes in each fMRI data frame. In an ideal data series, its value depends on the temporal standard deviation and temporal autocorrelation of the data (Nichols, 2017) and calculates the changes in the values of each voxel at each time point compared to its previous time point (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012). DVARS was calculated in the whole image to find a metric that demonstrated the standard deviation of temporal difference images in the 4D raw data (Power et al., 2014). DVARS was calculated using the following equation:

$$DVARS(\Delta I)_i = \sqrt{\left(\Delta I_i(x)\right)^2} = \sqrt{\left(I_i(x) - I_{i-1}(x)\right)^2} \quad (6)$$

In this equation, $\Delta \mathrm{I}_i(x)$ is used as local image intensity on the frame. DVARS could result in more accurate modeling of the temporal correlation and standardization because it is

obtained by the most short-scale changes (Nichols, 2017). The best value for this parameter is zero, and the closer it is to zero, the better the result.

We extracted the tSNR and DVARS parameters of output results by using the SCT toolbox and the FSL toolbox (Woolrich et al., 2009). For more accurate analysis of the tSNR parameter, we manually segmented the data into two parts, spinal cord and CSF, using the FSLeyes toolbox. Analyses compared the outcome of SCT and our method (DeepRetroMoco).

2.2.3. Statistical Analysis

All statistical analyses were carried out using IBM SPSS Statistics (V. 25 IBM Corp., Armonk, NY, USA) with $\alpha < 0.05$ as the statistical significance threshold. The Kolmogorov-Smirnov test was used to determine the normality of the parameters. For statistically significant results, the mean of normal data for each method was processed using one-way ANOVA with repeated measures in within-subjects comparison, followed by a multiple comparison post-hoc test with Bonferroni correction.

2.2.4. Implementation

In our experiment, we trained our deep learning network with and without using Fixing Centerline as preprocessing for the network. The number of our training epochs is 200 with 150 iterations for each epoch. We used Keras with TensorFlow backend (Abadi et al., 2016) on NVIDIA GEFORCE RTX 1080 to train our network, and it took an average of 23 hours to run the network. We also used Google Colab to review models and learn various parameters, which we worked on much faster.

The optimization parameter we used was Adam, with a learning rate of $1e^{-4}$ (Kingma & Ba, 2015). We trained our two simple (DRM_2) and complex (DRM_1) models designed by two different cost functions (NCC, MSE) and different lambda up to convergence time.

In our study, we designed a data generator to deliver fMRI data to the network. The data generator randomly selects the subject and slice and then selects a pair (fixed and moving) from the corresponding volume to the size of the data batch.

In comparison between the models, we chose the model that has better results in terms of our desired metric (tSNR) on the validation data. Then, we select one of the cost functions. Our code and model parameters are available online at https://github.com/mahdimplus/DeepRetroMoco

Table 1- Average tSNR for two types of our model, DRM-1 and DRM-2. Standard deviations are in parentheses. The averages are computed over all validation data. And in both models, regardless of the type of cost function, the first model is selected. (df=39)

Model	*Loss type*	*Mean. tSNR (std)*	*F-value*	*P-value*
DRM_1	MSE	6.18(0.6)	12.408	<0.001
DRM_2		5.96(0.7)		
DRM_1	NCC	10.13(1)	2.632	<0.05
DRM_2		10.07(1.3)		

3. Result

3.1 Model selection

Table 1 displays the average of our method's tSNR values in the validation data utilizing two distinct cost functions. The first model, DRM 1, outperforms DRM 2 in both Losses MSE and NCC by a slight margin. Furthermore, when the validation data of two cost functions in the first model are examined, NCC with an average of 10.13 ± 1, has better outcomes for the motion correction target based on the tSNR and statistical analysis, $t(39) = 2.63$, $p<0.05$.

3.2 Test statistic report

A repeating one-way measurement ANOVA was used to compare the influence of motion correction techniques on test data in sct_fmri_moco (De Leener et al., 2017) and DeepRetroMoco, a deep neural network-based motion correction tool.

Table 2- Summary of tSNR as an Image quality parameter between different motion correction methods (df=4)

tSNR		*Mean (std)*	*F value*	*p-value*
Spinal cord	Main Image	7.1043(2.41)	1004.249	<0.001
	sct_fmri_moco	12.90(2.44)		
	DeepRetroMoCo	16.072(3.09)		
CSF	Main Image	4.0387(1.17)	938.842	<0.001
	sct_fmri_moco	7.1469(1.31)		
	DeepRetroMoCo	10.3156(2.25)		

Table 3- Summary of DVARS as an Image quality parameter between different motion correction methods (df=4)

DVARS	*Mean (std)*	*F value*	*p-value*
Main Image	.0343(.009)	176.446	<0.001
sct_fmri_moco	.0316(.009)		
DeepRetroMoCo	.0182(.006)		

In a statistical comparison of tSNR parameters in the spinal cord, this parameter increased significantly from 7.104 ± 2.41 to 16.072 ± 3.09 arbitrary units (AU) (Table 3). Mauchly's Test of Sphericity revealed that the assumption of sphericity had been violated, $\chi^2(9) = 2.324$, $p < .313$, and thus a Greenhouse-Geisser correction was used. The motion correction algorithm had a significant effect on the tSNR parameter in the spinal cord, $F(2, 160) = 862.572$, $p < .0001$. Post hoc multiple comparisons using the Bonferroni correction revealed that the DeepRetroMoCo had a significantly higher mean tSNR in the spinal cord than the other motion correction method and raw data ($p<.0001$). Figure 4 depicts the significant difference between the groups using a violin plot.

The tSNR in CSF increased significantly from 4.038 ± 1.17 to 10.315 ± 2.25 arbitrary units (AU) (Table 2 and Table 3). Mauchly's Test of Sphericity revealed that the sphericity assumption had been violated, $\chi^2(9) = 27.772$, $p < .0001$, and thus a Greenhouse-Geisser correction was applied. The motion correction algorithm had a significant effect on the tSNR parameter in CSF $F(2, 160) = 949.72$, $p < .0001$. Post hoc multiple comparisons using the Bonferroni correction revealed that the DeepRetroMoCo's mean tSNR in CSF was significantly higher than the other motion correction method and raw data ($p<.0001$). Figure 4 depicts the significant difference between the groups using a violin plot.

DVARS decreased statistically significantly from 0.034 ± 0.009 to 0.182± 0.006 arbitrary units (AU) (Table 4). Mauchly's Test of Sphericity revealed that the sphericity assumption had been violated, $\chi^2(9) = 64.966$, $p < .0001$, and thus a Greenhouse-Geisser correction was applied. The motion correction algorithm had a significant effect on the DVARS parameter, $F(2, 160) = 309.349$, $p < .0001$. Post hoc multiple comparisons using the Bonferroni correction revealed that the DeepRetroMoCo had significantly lower DVARS than the other motion correction methods and raw data ($p<.0001$). Figure 4 depicts the significant difference between the groups using a violin plot.

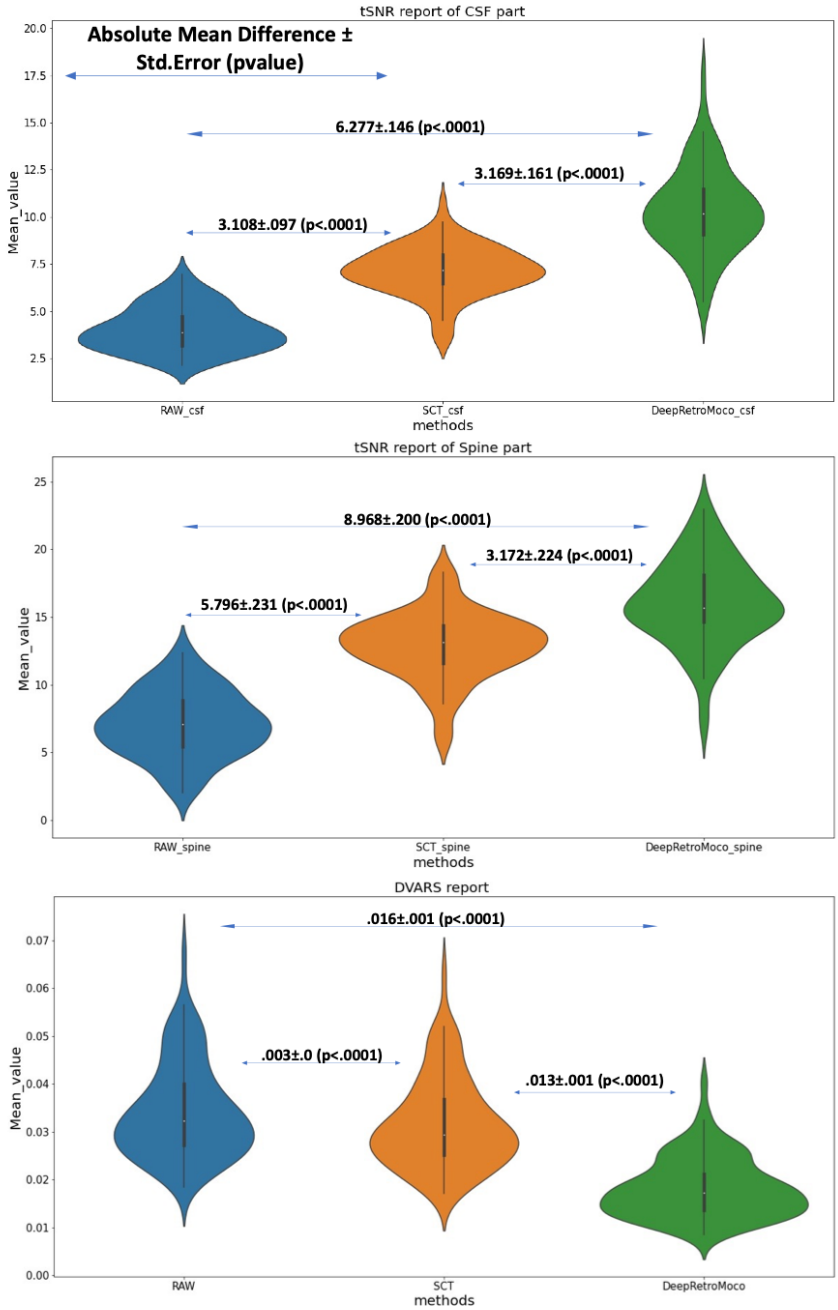

Figure 4- This figure depicts the mean and standard deviation of the SNR on the Spine and CSF sections (two top figures) that were manually segmented, as well as DVARS (bottom figure) with three types of results. RAW data that hasn't been corrected, SCT results, and DeepRetroMoco results are the three groups. The absolute mean difference + standard error (p value) between groups is also reported.

*. The mean difference is significant at the .05 level.

3.3 Compare with other methods

Further, we used the FSL gold standard method and MC_FLIRT for the estimation of movements in three groups of our data. Group one contains the RAW data which is not corrected for motion. Group two includes the result of the SCT toolbox, sct_fmri_moco and the third group contains the DeepRetroMoCo results. The reference volume was set to the first volume and the DOF was 6.

We used the MSE parameter to analyze the results, setting the reference line for actual movement to zero and the predicted movement to the numbers reported by FSL. As reported in Table 5, the raw data had the most movement in all directions, followed by the SCT and DeepRetroMoCo results, which had the most movement respectively.

Table 4- Mean square error of 3 groups of our data in 6 directions such as Translation in X, Y, Z and Rotation in X, Y, and Z directions.

MSE	*Translation*			*Rotation*		
Dir	X	Y	Z	RX	RY	RZ
DeepRertoMoco	1.41e-07	2.06e-07	5.43e-07	0.0010	0.0001	0.0023
SCT	5.54e-06	3.91e-07	1.47e-06	0.0012	0.0103	0.0054
RAW data	5.61e-06	1.15e-05	1.55e-06	0.1435	0.1073	0.0017

3.4 Processing speed

The implementation and calculation are carried out in a workstation with Intel® Core (TM) i7-4720HQ CPU at 2.60Hz and 16.0GB memory. No explicit parallelization was implemented in the Python script. The computation time of the motion correction procedure in sct_fmri_moco and DeepRetroMoco changes with the number of volumes of fMRI raw data. Average computation times (± Stdv) were 222.54±63.64 seconds and

131.91±35.94 seconds for sct_fmri_moco and DeepRetroMoco respectively and demonstrates a significant reduction of ~40.72% in computation time. This operation for SCT contains the slice-by-slice registration plus regularization across the Z, and for DeepRetroMoCo contains fixing centerline plus registration via a network.

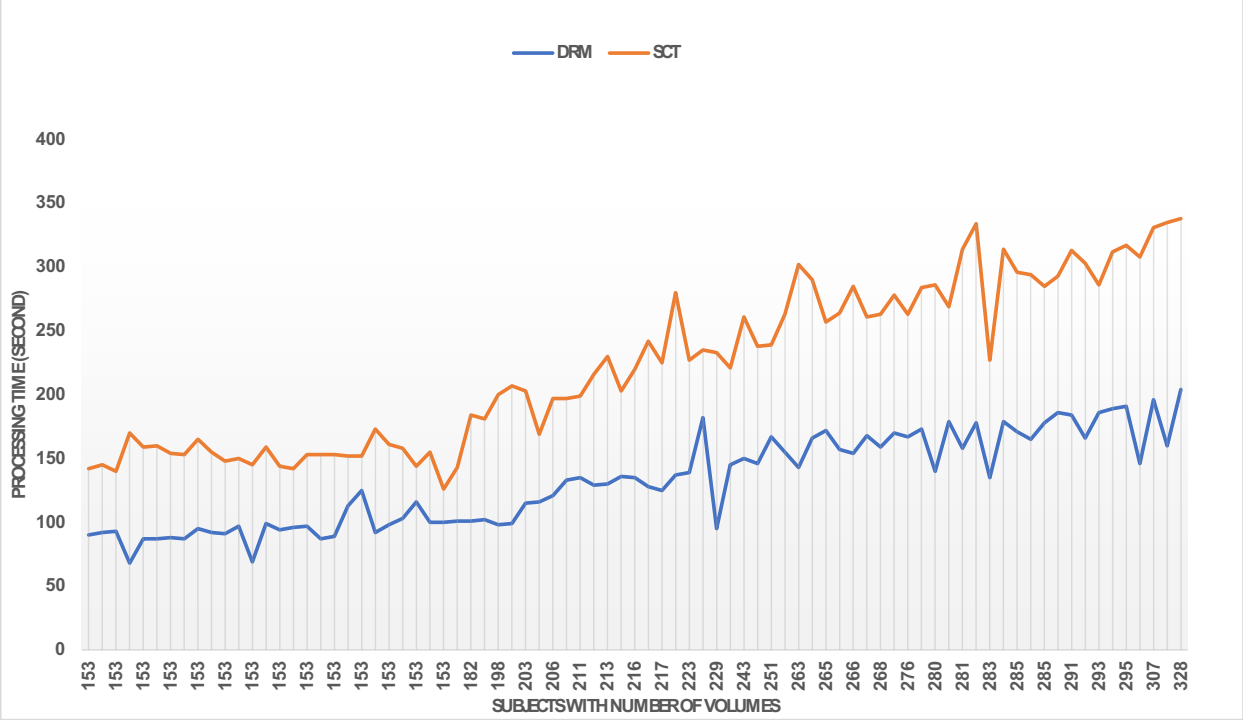

Figure 5- Comparing the speed of the two methods sct_fmri_moco (SCT) and DeepRetroMoco (DRM). Processing time is measured in seconds to correct the motion on all volumes.

3.5 Regularization Analysis

With different lambda parameters, we examined the mean tSNR for the test data. With the NCC cost function, the optimal tSNR for model 1 occurred when lambda was 0.01. In this section, the mean tSNR is applied to the entire spinal cord; lambda=0 indicates no regularization. As shown, the results deteriorate dramatically as the regularization term is increased. As a result, lambda's actions do not help to improve performance and may have a negative impact on the results for the NCC cost function and the first model, which is more complex.

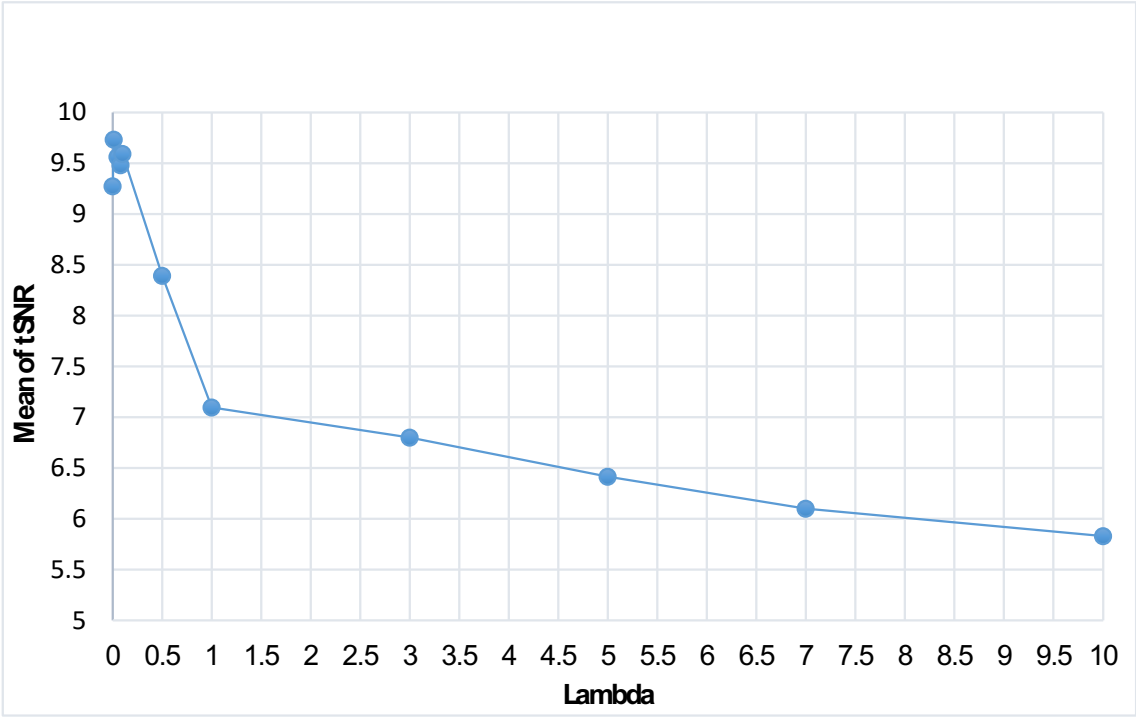

Figure 6- Effect of different λ modes for DRM_1 based on tSNR. Lambda 0.01 has a maximum tSNR and shows the best results.

3.6 Correlation Coefficient Analysis

We further calculated the Pearson correlation coefficient (CC) between the corrected and reference volumes to examine the similarity of the two images in a linear fashion. For

both the raw data and the motion-corrected data, we calculated the average CC per subject. Since linear relationships between variables are preserved by a linear transformation, the correlation between x and y would remain unchanged following a linear translation (Carroll et al., 1985). We found a large average CC value in our algorithm (0.90±0.02) compared to the raw data (0.70±0.17), which shows that our trained network (DeepRetroMoco) uses a significant proportion of linear transformations due to our rigorous regularization.

4. Discussion

Since the spinal column's voluntary and non-voluntary movements lead to non-optimal shimming, the effects of motion artefacts cannot be fully eliminated even after perfect conventional retrospective motion correction of successive functional volumes in image space (Eippert, Kong, Jenkinson, Tracey, & Brooks, 2017). If spinal column movements are small, motion correction is a useful step to improve the data quality for subsequent statistical data analysis. Our findings demonstrate that deep learning-based motion correction, DeepRetroMoco improves the quality of spinal cord fMRI data acquired in the axial field of view that has an effect on the pre-processing step. These improvements are at least in part due to improved tSNR and DVARS parameters compared to conventional algorithms introduced in the SCT data processing toolbox.

Instead, here we aimed to use a deep learning-based method potential to decrease preprocessing step for spinal cord fMRI data that strongly affected by motion. We found significant differences in the time of processing to implement DeepRetroMoco compared to the sct_fmri_moco algorithm.

As previously mentioned, the majority of leaning-based methodologies require additional data or ground truth. We don't need this information, which is another clear distinction between our approach and earlier research. The previous two works (Li & Fan, 2017; Vos,

Berendsen, Viergever, Staring, & Išgum, 2017) reported unsupervised methods that are close to ours. Both use the CNN neural network with spatial transformation function (Jaderberg et al., 2015), which warps images on top of each other and has significant problems: they only operate on a limited subset of volumes and only support small transformations. In addition a recent study (Balakrishnan, Zhao, Sabuncu, Dalca, & Guttag, 2018) and our network improved the problems mentioned and helped to solve them by designing a satisfactory model in the spinal cord data. Other methods (Vos et al., 2017) use regularization that is determined only by interpolation methods.

The DeepRetroMoco replaces a costly optimization problem for each image pair, with a function optimization that is collected over a data set during a training step. This notion could be replaced with previous motion correction algorithms, especially on spinal cord data which traditionally relies on complex, non-learning-based optimization algorithms for each input. Although implementing this network requires a one-time network training on a single NVIDIA TITAN X GPU with training data, it takes less than a second to register a pair of images. Due to the growing need for medical images for further investigation in less time, our solution, which is a learning-based method, is preferable to non-learning-based methods.

5. Limitation and Future Works

The acquisition of spinal cord fMRI data is made in two ways: GRE-EPI acquisition sequence in axial and FSE or SE-HASTE acquisition sequence in sagittal field of view. The field of view and data-set orientation were axial in this study, and all motion correction methods and preprocessing steps were performed specifically on axially oriented data in the cervical spine; however, some studies performed spinal cord fMRI acquisition in the sagittal orientation. Respiration and heartbeat cause changes in CSF flow around the spinal cord, which results in magnetic susceptibility during the GRE-EPI image acquisition and preprocessing steps.

Furthermore, we had access to two variables during this method: the centerline reference and the fixed image reference. It was set to the first volume in our network. We discovered that the proper selection of these two parameters could have a significant impact on the final results. Because our network is flexible enough to accept any reference, including first, mean, middle, and any other desired volume, we propose that the best reference for each data be selected by designing the appropriate method for future work.

6.Conclusion

Owing to the bulk and physiological motion corrupted spinal cord fMRI data, the statistical significance of the activation maps decreases, and the likelihood of false activations increases. As a result, a motion correction algorithm is required for acceptable single and group fMRI data analysis. In this study, we proposed DeepRetroMoco, an unsupervised learning-based approach based on advanced CNN models, that requires no supervised information such as ground truth registration fields or anatomical landmarks. Additionally, when compared to conventional methods, the use of DeepRetroMoco motion correction in spinal cord fMRI appears to be impressive in terms of increasing tSNR, reducing false positives, and increasing sensitivity, especially in cases of significant spinal cord motion. Furthermore, the statistical evaluation of DVARS as an fMRI quality measure, as well as the time of implementation on a cervical spinal cord fMRI dataset, demonstrated the superiority of the proposed framework in our experimental study and also is an easy-to-integrate tool for more accurate and faster motion correction for denoising in spinal cord fMRI applications.

Reference

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Alexander, M. S., Kozyrev, N., Bosma, R. L., Figley, C. R., Richards, J. S., & Stroman, P. W. (2016). fMRI Localization of Spinal Cord Processing Underlying Female Sexual Arousal. *J Sex Marital Ther, 42*(1), 36-47. doi:10.1080/0092623x.2015.1010674

Anaya-Isaza, A., Mera-Jiménez, L., & Zequera-Diaz, M. (2021). An overview of deep learning in medical imaging. *Informatics in Medicine Unlocked, 26*, 100723. doi:https://doi.org/10.1016/j.imu.2021.100723

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*(1), 95-113. doi:10.1016/j.neuroimage.2007.07.007

Ashburner, J., Andersson, J. L., & Friston, K. J. J. H. b. m. (2000). Image registration using a symmetric prior—in three dimensions. *9*(4), 212-225.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal, 12*(1), 26-41. doi:10.1016/j.media.2007.06.004

Bajcsy, R., & Kovačič, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing, 46*(1), 1-21. doi:https://doi.org/10.1016/S0734-189X(89)80014-3

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Dalca, A. V., & Guttag, J. (2018, 18-23 June 2018). *An Unsupervised Learning Model for Deformable Medical Image Registration.* Paper presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Carroll R.J., Lombard, F. (1985) A Note on N Estimators for the Binomial Distribution, *Journal of the American Statistical Association*, 80:390, 423-426, DOI: 10.1080/01621459.1985.10478134

Ceritoglu, C., Wang, L., Selemon, L. D., Csernansky, J. G., Miller, M. I., & Ratnanather, J. T. (2010). Large Deformation Diffeomorphic Metric Mapping Registration of Reconstructed 3D Histological Section Images and in vivo MR Images. *Front Hum Neurosci, 4*, 43. doi:10.3389/fnhum.2010.00043

Cohen-Adad, J., Piche, M., Rainville, P., Benali, H., & Rossignol, S. (2007, 22-26 Aug. 2007). *Impact of realignment on spinal functional MRI time series.* Paper presented at the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Dalca, A. V., Bobu, A., Rost, N. S., & Golland, P. (2016). Patch-Based Discrete Registration of Clinical Brain Images. *Patch Based Tech Med Imaging (2016), 9993*, 60-67. doi:10.1007/978-3-319-47118-1_8

De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Louis Collins, D., . . . Cohen-Adad, J. (2017). SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *NeuroImage, 145*, 24-43. doi:https://doi.org/10.1016/j.neuroimage.2016.10.009

De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Louis Collins, D., . . . Cohen-Adad, J. (2017). SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage, 145*(Pt A), 24-43. doi:10.1016/j.neuroimage.2016.10.009

Dehghani, H., Oghabian, M. A., Batouli, S. A. H., Arab Kheradmand, J., & Khatibi, A. (2020). Effect of Physiological Noise on Thoracolumbar Spinal Cord Functional Magnetic Resonance Imaging in 3T Magnetic Field. *BCN, 11*(6), 737-752. doi:10.32598/bcn.11.6.1395.1

Dehghani, H., Weber, K. A., Batouli, S. A. H., Oghabian, M. A., & Khatibi, A. (2020). Evaluation and Optimization of Motion Correction in Spinal Cord fMRI Preprocessing. *bioRxiv*, 2020.2005.2020.103986. doi:10.1101/2020.05.20.103986

Eippert, F., Kong, Y., Jenkinson, M., Tracey, I., & Brooks, J. C. W. (2017). Denoising spinal cord fMRI data: Approaches to acquisition and analysis. *NeuroImage, 154*, 255-266. doi:10.1016/j.neuroimage.2016.09.065

Fratini, M., Moraschi, M., Maraviglia, B., & Giove, F. (2014). On the impact of physiological noise in spinal cord functional MRI. *J Magn Reson Imaging, 40*(4), 770-777. doi:10.1002/jmri.24467

Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., . . . Cohen-Adad, J. (2019). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with

convolutional neural networks. *NeuroImage, 184*, 901-915. doi:10.1016/j.neuroimage.2018.09.081

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967-5976.

Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. (2015). *Spatial transformer networks*. Paper presented at the Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, Montreal, Canada.

Khatibi, A., Vahdat, S., Lungu, O., Finsterbusch, J., Büchel, C., Cohen-Adad, J., . . . Doyon, J. (2022). Brain-spinal cord interaction in long-term motor sequence learning in human: An fMRI study. *Neuroimage, 253*, 119111. doi:https://doi.org/10.1016/j.neuroimage.2022.119111

Kinany, N., Pirondini, E., Martuzzi, R., Mattera, L., Micera, S., & Van de Ville, D. (2019). Functional imaging of rostrocaudal spinal activity during upper limb motor tasks. *Neuroimage, 200*, 590-600. doi:10.1016/j.neuroimage.2019.05.036

Kinany, N., Pirondini, E., Mattera, L., Martuzzi, R., Micera, S., & Van De Ville, D. (2022). Towards reliable spinal cord fMRI: Assessment of common imaging protocols. *NeuroImage, 250*, 118964. doi:https://doi.org/10.1016/j.neuroimage.2022.118964

Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *CoRR, abs/1412.6980*.

Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F. C., Miao, S., . . . Kamen, A. (2017). *Robust non-rigid registration through agent-based action learning.* Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.

Kybic, J., & Unser, M. (2003). Fast parametric elastic image registration. *IEEE Transactions on Image Processing, 12*(11), 1427-1442. doi:10.1109/TIP.2003.813139

Li, H., & Fan, Y. J. a. p. a. (2017). Non-rigid image registration using fully convolutional networks with deep self-supervision.

Maknojia, S., Churchill, N. W., Schweizer, T. A., & Graham, S. J. (2019). Resting State fMRI: Going Through the Motions. *Front Neurosci, 13*, 825. doi:10.3389/fnins.2019.00825

Nichols, T. E. (2017). Notes on Creating a Standardized Version of DVARS. *arXiv: Applications*.

Oakes, T. R., Johnstone, T., Ores Walsh, K. S., Greischar, L. L., Alexander, A. L., Fox, A. S., & Davidson, R. J. (2005). Comparison of fMRI motion correction software tools. *NeuroImage, 28*(3), 529-543. doi:10.1016/j.neuroimage.2005.05.058

Paquin, M. Ê., El Mendili, M. M., Gros, C., Dupont, S. M., Cohen-Adad, J., & Pradat, P. F. (2018). Spinal Cord Gray Matter Atrophy in Amyotrophic Lateral Sclerosis. *AJNR. American journal of neuroradiology, 39*(1), 184-192. doi:10.3174/ajnr.A5427

Perone, C. S., Calabrese, E., & Cohen-Adad, J. (2018). Spinal cord gray matter segmentation using deep dilated convolutions. *Scientific Reports, 8*(1), 5966. doi:10.1038/s41598-018-24304-3

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage, 59*(3), 2142-2154. doi:10.1016/j.neuroimage.2011.10.018

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage, 84*, 320-341. doi:10.1016/j.neuroimage.2013.08.048

Powers, J. M., Ioachim, G., & Stroman, P. W. (2018). Ten Key Insights into the Use of Spinal Cord fMRI. *Brain sciences, 8*(9), 173. doi:10.3390/brainsci8090173

Prados, F., Ashburner, J., Blaiotta, C., Brosch, T., Carballido-Gamio, J., Cardoso, M. J., . . . Cohen-Adad, J. (2017). Spinal cord grey matter segmentation challenge. *NeuroImage, 152*, 312-329. doi:https://doi.org/10.1016/j.neuroimage.2017.03.010

Risser, L., Vialard, F., Wolz, R., Murgasova, M., Holm, D. D., & Rueckert, D. (2011). Simultaneous Multi-scale Registration Using Large Deformation Diffeomorphic Metric Mapping. *IEEE Transactions on Medical Imaging, 30*(10), 1746-1759. doi:10.1109/TMI.2011.2146787

Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., & Pennec, X. (2017). *SVF-Net: learning deformable image registration using shape matching.* Paper presented at the International conference on medical image computing and computer-assisted intervention.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation.* Paper presented at the MICCAI.

Sabaghian, S., Dehghani, H., Batouli, S. A. H., Khatibi, A., & Oghabian, M. A. (2020). Fully automatic 3D segmentation of the thoracolumbar spinal cord and the vertebral canal from T2-weighted MRI using K-means clustering algorithm. *Spinal Cord, 58*(7), 811-820. doi:10.1038/s41393-020-0429-3

Sokooti, H., Saygili, G., Glocker, B., Lelieveldt, B. P. F., & Staring, M. (2016, 2016//). *Accuracy Estimation for Medical Image Registration Using Regression Forests.* Paper presented at the Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, Cham.

Sokooti, H., Vos, B. d., Berendsen, F., Lelieveldt, B. P., Išgum, I., & Staring, M. (2017). *Nonrigid image registration using multi-scale 3D convolutional neural networks.* Paper presented at the International conference on medical image computing and computer-assisted intervention.

Stroman, P. W., Figley, C. R., & Cahill, C. M. (2008). Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem. *Magn Reson Imaging, 26*(6), 809-814. doi:10.1016/j.mri.2008.01.038

Thirion, J.-P. (1995). *Fast Non-Rigid Matching of 3D Medical Images*.

Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: methodological failures and recommendations for the future. *npj Digital Medicine, 5*(1), 48. doi:10.1038/s41746-022-00592-y

Vos, B. D. d., Berendsen, F. F., Viergever, M. A., Staring, M., & Išgum, I. (2017). End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 204-212): Springer.

Wen, D., Wei, Z., Zhou, Y., Li, G., Zhang, X., & Han, W. (2018). Deep Learning Methods to Process fMRI Data and Their Application in the Diagnosis of Cognitive Impairment: A Brief Overview and Our Opinion. *Frontiers in Neuroinformatics, 12*. doi:10.3389/fninf.2018.00023

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., . . . Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage, 45*(1 Suppl), S173-186. doi:10.1016/j.neuroimage.2008.10.055

Yang, X., Bian, C., Yu, L., Ni, D., & Heng, P.-A. (2017). *Class-balanced deep neural network for automatic ventricular structure segmentation.* Paper presented at the International workshop on statistical atlases and computational models of the heart.

Yang, Z., Zhuang, X., Sreenivasan, K., Mishra, V., & Cordes, D. (2019). Robust Motion Regression of Resting-State Data Using a Convolutional Neural Network Model. *Front Neurosci, 13*, 169. doi:10.3389/fnins.2019.00169