

# Stacked U-Nets with self-assisted priors towards robust correction of rigid motion artifact in brain MRI



Mohammed A. Al-masni<sup>a</sup>, Seul Lee<sup>a</sup>, Jaeuk Yi<sup>a</sup>, Sewook Kim<sup>a</sup>, Sung-Min Gho<sup>b</sup>, Young Hun Choi<sup>c</sup>, Dong-Hyun Kim<sup>a,\*</sup>

<sup>a</sup> Department of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul, South Korea

<sup>b</sup> GE Healthcare, Seoul, South Korea

<sup>c</sup> Department of Radiology, Seoul National University Hospital, Seoul, South Korea

## ARTICLE INFO

### Keywords:

Deep learning  
Motion artifact correction  
MRI  
Prior-assisted  
Stacked U-Nets

## ABSTRACT

Magnetic Resonance Imaging (MRI) is sensitive to motion caused by patient movement due to the relatively long data acquisition time. This could cause severe degradation of image quality and therefore affect the overall diagnosis. In this paper, we develop an efficient retrospective 2D deep learning method called stacked U-Nets with self-assisted priors to address the problem of rigid motion artifacts in 3D brain MRI. The proposed work exploits the usage of additional knowledge priors from the corrupted images themselves without the need for additional contrast data. The proposed network learns the missed structural details through sharing auxiliary information from the contiguous slices of the same distorted subject. We further design a refinement stacked U-Nets that facilitates preserving the spatial image details and improves the pixel-to-pixel dependency. To perform network training, simulation of MRI motion artifacts is inevitable. The proposed network is optimized by minimizing the loss of structural similarity (SSIM) using the synthesized motion-corrupted images from 83 real motion-free subjects. We present an intensive analysis using various types of image priors: the proposed self-assisted priors and priors from other image contrast of the same subject. The experimental analysis proves the effectiveness and feasibility of our self-assisted priors since it does not require any further data scans. The overall image quality of the motion-corrected images via the proposed motion correction network significantly improves SSIM from 71.66% to 95.03% and declines the mean square error from 99.25 to 29.76. These results indicate the high similarity of the brain's anatomical structure in the corrected images compared to the motion-free data. The motion-corrected results of both the simulated and real motion data showed the potential of the proposed motion correction network to be feasible and applicable in clinical practices.

## 1. Introduction

Magnetic Resonance Imaging (MRI) is extremely sensitive to motion caused by patient movement. This is due to the relatively long data acquisition time required to acquire the k-space data to generate the MR image (Zaitsev et al., 2015). Motion artifacts manifest as ghosting, ringing, and blurring and may cause severe degradation of image quality. As a result, it becomes a challenge for radiologists to accurately interpret and diagnose patients with motion artifacts, leading to increased cost if the level of artifacts is too severe and the acquisition has to be repeated (Andre et al., 2015). Physiological motions, such as respiratory and cardiac movement, blood flow, and vessel pulsation, and voluntary and involuntary head motions, known as rigid or bulk motion, are the main sources of MRI motion artifacts (Godenschweger et al., 2016). Both the involuntary motions in pediatrics or neuro-degenerative patients and the sudden conscious motions due to discomfort or mental conditions

are unavoidable during data acquisition. Thus, the correction of motion artifacts is being more of interest for the MRI society.

The type of MRI motion artifacts relies on data acquisition strategies, such as k-space trajectory and pulse sequence diagram, with regards to the time and amount of the subject movement (Zaitsev et al., 2015; Shaw et al., 2020; Liu et al., 2020). Motion artifacts usually happen in the phase-encoding direction of the k-space, where patient motion during filling the center of k-space is relatively corresponding to the low image frequencies, resulting in motion with high severity. In contrast, patient movement during data acquisition of the edges of k-space tends to result in ringing artifacts. In general, motion artifacts reduction or correction methods can be categorized into two main groups: prospective (Tisdall et al., 2012; Herbst et al., 2012) and retrospective (Atkinson et al., 1997; Vaillant et al., 2014; Loktyushin et al., 2015; Haskell et al., 2018; Cordero-Grande et al., 2016). The former method is a real-time approach that attempts to compensate the motion ar-

\* Corresponding author.

E-mail address: [donghyunkim@yonsei.ac.kr](mailto:donghyunkim@yonsei.ac.kr) (D.-H. Kim).

tifacts during scanning by updating the pulse sequence using motion tracking devices, such as navigators or sensors. In opposite, the retrospective approach estimates the motion parameters after completing the scanning (i.e., during image reconstruction) using prior information derived from external sensors or applying iterative or autofocus-ing algorithms. However, such approaches require additional hardware and time, which imply a concern of calibrations with the MRI scanner and intensive computations to estimate motion parameters, respectively (Godenschweger et al., 2016). Therefore, automated correction of MRI motion artifacts without the need of external navigators is of great clinical significance and has become an active research area.

Recently, the advances of deep learning convolutional neural networks (CNNs) are gaining a lot of attention and have been widely utilized in the field of medical image analysis, specifically the MRI domain. For instance, the CNNs have been succeeded in developing several clinical MRI applications, including reconstruction of images (Hyun et al., 2018; Ryu et al., 2019; Schlemper et al., 2018; Aggarwal et al., 2019) and correction of artifacts (Terpstra et al., 2020; Chatterjee et al., 2020; Haskell et al., 2019; Ko et al., 2021; Lee et al., 2021; Liu et al., 2020; Oh et al., 2021; Oksuz et al., 2019; Wang et al., 2020; Zhang et al., 2020; Shaw et al., 2020). The last few years have witnessed numerous attempts to solving the task of motion artifacts correction using the deep learning paradigm. (Meding et al., 2017) introduced a fully automated data-driven CNN approach to determine whether the recorded MR image is affected by motion artifacts during image acquisition. They utilized a binary classification model to detect 'motion' and 'no motion' categories within intra-scan motion rather than inter-scan. The works in Oksuz et al. (2019), Zhang et al. (2020) proposed automatic multi-stream CNNs to detect the presence of motion artifacts in cardiac MR images. A curriculum learning, a part of active learning, was employed to efficiently determine the level of motion severity by training a deep learning network with samples of gradually increasing the difficulty (i.e., from good to poor image quality) (Oksuz et al., 2019). The attention supervision was used in Zhang et al. (2020) to guide the network to focus on the target region. This particular attention was achieved by computing the importance weights of the gradients with respect to the activation maps. Also, Ko et al. presented the significance of using a self-spatial attention module within the deep residual network for rigid and non-rigid motion artifacts reduction in computed tomography images (Ko et al., 2021). Kustner et al. (2019) compared two image-to-image translation methods, Generative Adversarial Network(GAN) and Variational Auto Encoder (VAE), for retrospective correction of rigid and non-rigid motion artifacts. The results presented that the GAN method has better feasibility to produce near-realistic motion-free images. For 3D rigid-body motion correction in brain MRI, Johnson and Drangova (2019) developed a 3D version of a 2D pixel2pixel network based on conditional GAN (cGAN). The 3D cGAN's generator was composed of 3D U-Net, and its discriminator contained six 3D CNN layers. Their method resulted in a qualitative image improvement for motion-corrupted test-set volumes. An investigation of the applicability of 3D CNNs methods for retrospective motion artifact correction was presented in Duffy et al. (2018). Both High-resolution 3D Network (HR3DNet) and HR3DNetGAN successfully diminished the severity of artifacts compared to the conventional Gaussian smoothing method. As an extension to this work, the same research group demonstrated the possibility of 3D CNNs to improve cortical surface reconstruction and their ability to better identify brain morphological changes in subjects with Parkinson's disease (Duffy et al., 2021). In Haskell et al. (2019) developed a separable motion correction model called Network Accelerated Motion Estimation and Reduction (NAMER) (Haskell et al., 2019), which integrated a deep learning motion artifacts detection with a motion estimation model. Although the NAMER has achieved promising motion mitigation, it required a long processing time of around 7.13 min for a single slice. Hence, the need for an end-to-end method with less execution time is still demanded. Similar to the NAMER, Wang et al. utilized the CNN prediction as an initial guess (i.e., data fidelity)

into the optimization problem to correct the out-of-field-of view motion (Wang et al., 2020). Their results show that the combined motion model-based data fidelity outperformed the CNN prediction for the 2D motion correction. However, in the case of 3D motion correction, better image quality was obtained using the CNN prediction compared to the data fidelity model. In Liu et al. (2020), the authors developed a deep residual U-Net with densely connected multi-resolution blocks (DRN-DCMB) to correct the motion artifacts in MR images. The proposed network was designed to learn the motion artifacts, where a single shortcut connection that connected the corrupted input with the residual map prediction could reproduce the output with reduced motion artifacts. The performance of motion artifacts reduction was improved from 0.867 (the input simulated motion) to 0.965 in term of the Structural Similarity index (SSIM). Compared to the original U-Net, their network achieved a slight improvement rate of 0.8% in term of SSIM.

Different from the above studies, Lee et al. (2021), Chatterjee et al. (2020) attempted to solve the motion problem by the inclusion of multi-contrast images. Lee et al. built a framework consisting of two parts: multi-contrast image registration and motion correction networks for three MRI sequences, namely T<sub>1</sub>-weighted (T<sub>1</sub>w), T<sub>2</sub>-weighted (T<sub>2</sub>w), and Fluid-Attenuated Inversion Recovery (FLAIR) (Lee et al., 2021). After the multi-contrast image alignment has been performed, the encoder of the motion correction network separately extracted feature vectors for each contrast image. Then, all these features were combined to share information from different contrasts. In the end, the decoder path was able to learn these features and generate multi-output corrected predictions. Even though this work presented different scenarios of using input images with motion-free or motion-corrupted, it seems that the authors did not intend to exploit multi-contrast as priors as clearly mentioned in Chatterjee et al. (2020). The best motion correction scenario for any corrupted target was through feeding the rest pure contrast images without any motion. More specifically, for better motion artifact reduction in T<sub>1</sub>w, it was recommended to pass the motion-free T<sub>2</sub>w and FLAIR images into the network along with the corrupted T<sub>1</sub>w image. Chatterjee et al. introduced a retrospective deep learning network to remove motion artifacts with the assistance of additional information presented as image priors (Chatterjee et al., 2020). Supplying additional prior knowledge was performed using two different types of image priors: similar slices of the image contrast from other subjects and different contrasts of the same subjects. The second prior type is similar to the work presented in Lee et al. (2021), but it used different image contrasts, including T<sub>1</sub>w, T<sub>2</sub>w, and Proton Density (PD) images. During the correction process of the corrupted T<sub>2</sub>w image, they utilized the motion-free T<sub>1</sub>w and PD images as priors. Their findings concluded that the correction performance was not improved when the priors of similar contrasts but different subjects were utilized. In contrast, promising results were obtained in the case of using priors of different contrasts but of the same subjects.

Since it is challenging to acquire pairs of motion-corrupted images and ground-truth clean images, most research in the field of motion artifact correction trained and tested their methods using synthesized corrupted images from motion-free data. The motion artifact simulation could be achieved by applying some rotation and translation transformations in the spatial domain or by adding phase shift to the k-space in the frequency domain. The amount of motion could be controlled to derive various levels of motion severity. For example, the average SSIM for the corrupted motion data was between 40% and 74% in Wang et al. (2020), 86.70% in Liu et al. (2020), and around 77.0% in Chatterjee et al. (2020). However, the corrupted input images in Lee et al. (2021) seem to be not severe since the average SSIM scores of 98.11%, 91.25%, and 92.79% were reported for T1w, T2w, and FLAIR, respectively.

It is observed that most of the aforementioned works endeavored to reduce the MRI motion artifacts through developing new deep learning architectures. However, the task of motion correction is quite different

from other tasks such as segmentation and detection. It is of note that deep learning networks in segmentation and detection tasks usually try to learn some representative features from input data that reflect the target Region of Interest (ROI) (i.e., target ROI is visible to the network). Differently, deep learning motion correction networks struggle to find or correct missing structural details from the distorted images. Only a few works (Lee et al., 2021; Chatterjee et al., 2020) have utilized additional prior knowledge from different image contrasts of the same subjects. Nevertheless, these works required additional MRI scans, which could not be available in all routine clinical exams and does not seem to be feasible for future medical practices. Therefore, sufficient room exists for improving the feasibility and effectiveness of the motion artifacts correction.

In this paper, we address the above-mentioned issues by designing efficient 2D stacked U-Nets with self-assisted priors to solve the problem of motion artifacts in 3D brain MRI. The proposed work aims to exploit the usage of additional knowledge priors from the corrupted images themselves without the need for additional contrast data. Since the brain structure is homogenous among the relatively neighboring regions (i.e., has high correlations in successive slices), it seems possible to transfer these anatomical details from the adjacent slices that might have less motion corruption. Eventually, each specified motion-corrupted image can be corrected significantly via enabling the proposed stacked networks to learn extra knowledge information using the self-assisted priors (i.e., from contiguous slices). More specifically, the proposed network initiates by concatenating multi-inputs (i.e., the corrupted image and its adjacent slices) and eventually yields a single corrected image. In this case, the network could reveal some missed structural details throughout the assistance of the information that exists in the adjacent slices, especially in the case of 3D imaging (as the case of this work). Recent retrospective deep learning studies included multi-contrast images or similar slices as image priors for better motion correction performances. In contrast, this work aims to develop a reliable motion correction method without multi-contrast MR images that require additional scans and high computational costs. Alternatively, we include the contiguous slices in 3D MR images with high spatial similarities as image priors into the proposed network to fill in missing structures. The adjacent slices might be utilized as a useful contextual information fusion strategy in some medical image analysis applications. However, to the best of our knowledge, this is the first time to employ the idea of self-assisted priors to address the problem of motion artifacts correction. Furthermore, we develop stacked U-Nets that enable to re-evaluate the initial estimates by capturing some spatial relationships between predictions. Here, the prediction reuse with spatial attention preserves the spatial location and leads to better refinement of the corrected images. Finally, our proposed stacked U-Nets with the inclusion of self-assisted priors and spatial attention result in better performance of the MRI motion artifact correction.

Even though 3D motion correction methods can inherently leverage context information on 3D MRI data, they require high computational complexity (i.e., computationally expensive). Additionally, the more extensive training parameters of 3D methods compared to 2D networks may require larger training samples. In contrast, our proposed 2D motion correction approach is quite similar to 3D techniques since it can share the context from the adjacent slices with much fewer computations.

The main contributions of this paper are outlined as follows.

- (1) We propose a novel motion artifact correction method that learns additional knowledge priors from the adjacent slices of the same corrupted subject. The main idea of the self-assisted priors is to capture some missed structural information from image priors and enable the network to learn these unreachable patterns to achieve more accurate correction of motion artifacts.
- (2) With the inclusion of self-assisted priors and spatial attention, we develop end-to-end 2D stacked refinement networks. This can further

improve the pixel-to-pixel dependency through pixel-wise matching and spatial location preservation between predictions.

- (3) We provide an intensive analysis of using different types of image priors: the proposed self-assisted priors and priors from other image contrast of the same subject. This analysis proves the effectiveness and feasibility of our self-assisted priors since it does not require any further data scans.
- (4) The source code of our proposed stacked U-Nets with self-assisted priors is available at [\(13:italic \)https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Correction-Self-Assisted-Priors/\(13:italic \)](https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Correction-Self-Assisted-Priors/). We also make our MRI motion artifact simulation tool publicly available at [\(13:italic \)https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Simulation-Tool/\(13:italic \)](https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Simulation-Tool/).

The rest of this article is organized as follows. Section 2 presents the motion artifacts simulation and provides theory and details of the proposed stacked networks. Section 3 demonstrates the ablation analysis of using different types of priors and presents all experimental results of both the simulated and *in-vivo* data. Eventually, Sections 4 and 5 discuss and conclude the paper, respectively.

## 2. Materials and methods

### 2.1. Original dataset

To accomplish the motion artifact correction task, we have collected a set of clinical brain MRI data at Seoul National University Hospital (SNUH). Written informed consent was obtained from all participants and the study was approved by the Institutional Review Board of SNUH. Our dataset contains 83 motion-free clinical subjects, which was utilized to generate the simulated motion data to conduct network training. Additionally, separate 24 clinical subjects with motion distorted were acquired to perform the *in-vivo* assessment. These real clinical subjects were not instructed to induce motion during data acquisition. 3D T1-weighted gradient echoes, called BRAin VOLUME (BRAVO), were acquired on a 3.0 Tesla MRI scanner (SIGNA Premier, GE Healthcare, United States) with the following imaging parameters: echo time (TE) of 2.77 ms, repetition time (TR) of 6.86 ms, flip angle of 10°, and pixel bandwidth of 244 Hz/pixel. The sagittal image matrix varies from 256 × 256 to 512 × 512 pixels with an in-plane resolution of 0.90 × 0.90 mm<sup>2</sup> and slice spacing of 1 mm, while the third dimension is in the range of 144 to 380.

Also, 38 subjects out of the 83 motion-free clinical patients have additional Contrast-Enhanced (CE) scans. This CE data has the same imaging parameters and FOV as the original non CE data. In this work, we utilized the real-valued magnitude data as input to the proposed network.

### 2.2. Simulation of motion artifacts

Since there is a limitation to acquire various pairs of reference motion-free and target motion-corrupted datasets, simulation of MRI motion artifacts is inevitable to fulfill the network training. This was achieved by synthesizing the motion-corrupted images from the 83 original motion-free subjects. This work included the simulation of both 3D sudden and continuous brain motions to generate the motion-corrupted data. To achieve the feasibility and reliability of the proposed motion correction method, it is important to generate motion data that highly imitates the real motion artifacts. To do so, we assumed that the bulk rigid motion is a combination of both the rotation and translation motions. Motion artifacts were generated by applying sporadic rotational motions in the range of [-7°, +7°] on all three axes and applying translational motions between -7 and +7 mm on all three planes. Motion simulation adopted both the in-plane and through-plane motions. This was performed in a 3D sequence, and each axis of translation and rotation (anterior-posterior (AP), right-left (RL), superior-inferior (SI), yaw,

pitch, and roll) was considered. Both the rotation and translation motions could be achieved by controlling their parameters ( $\rho_x, \rho_y, \rho_z$ ) and ( $t_x, t_y, t_z$ ). To generate various motion severity (i.e., various strengths of motion artifacts), similar to the real motion cases, the motion parameters were acquired by generating random peaks of various numbers and ranges as follows. For mild and moderate motions, we set the range of motion for each peak: 2~4% and 2~4%, number of motions: 2~4 and 4~6, and maximum motion strength: 3 and 5, respectively. Whereas, we set the range of motion for each peak: 3~5%, number of motions: 5~7, and maximum motion strength: 7 for severe motion. Also, we set the rotational and translational motions in range of [-3°, +3°] and -3 to +3 mm for mild, [-5°, +5°] and -5 to +5 mm for moderate, and [-7°, +7°] and -7 to +7 mm for sever motions. It is noteworthy that this motion simulation could be performed in the image domain or frequency domain. In this work, we performed a complicated motion simulation process simultaneously in both image and frequency domains in order to derive the simulated motion data with very high similarity to the real motion cases. More specifically, the rotation motion was performed in the image domain (image-based simulation) by applying some rotational operations on a motion-free image and then sampling the k-space lines of the Fourier transform of that rotated image. Meanwhile, the translation motion (k-space-based simulation) was performed by adding linear phase shifts directly to the k-space in the phase-encoding direction. Since the phases in the frequency domain are only affected by translational motion, mathematically, this is equivalent to multiplying each line in k-space by a phase ramp whose slope equals the spatial displacement at the relevant time point. The Fourier Transform identity asserts that circular shifting of the signal in spatial domain equates to multiplication with a linear phase ramp in Fourier domain. In other words, translational motion is considered a convolution in the image domain. Therefore, when the translation motion is simulated in the k-space domain, linear phase shifts are applied in certain k-space lines. Computation-wise, this can be done in both Fourier and spatial domains; however, the former is faster to compute, especially when dealing with 3D data. We repeated this motion simulation process for N times, where N refers to the number of distorted lines in k-space. Since the motion simulation was performed in three different motion severity levels, the N varies according to motion level. For the diversity of motion simulation, each motion level has a range of k-space lines N affected by motion: 10~40, 40~60, and 50~90 lines for mild, moderate, and severe, respectively. At the time of motion simulation, the range of k-space lines was selected randomly. The deformed k-space lines of both rotation and translation motions were accumulated to eventually generate the corrupted k-space. Then, mapping this result using the DFT produces a motion-corrupted image. The computation time for sporadic motion is different for each patient since they got a different number of distorted k-space lines. When we performed the simulation with GPU, it took about 250 to 350 s per patient. Fig. 1 demonstrates the entire motion simulation process. More details about the motion simulation tool are available in our previously published paper (Lee et al., 2020). We make our motion simulation tool ‘view2Dmotion’ publicly available at <https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Simulation-Tool>.

To accomplish the network training and evaluation, we divided the synthesized motion data based on the subject level into two main sets, namely the training and testing with portions of 80% and 20%, respectively. In term of images, a total of 9,996 images were utilized for training, while 3,390 images were used for testing. Among the training data, we randomly selected 5% of the images and used them for validation. The utilized data were normalized based on the subject level using 3D data intensities.

### 2.3. Theory

The goal of the correction methods of MRI motion artifacts is to recover the motion-free image  $x$  from the motion-corrupted image  $x_m$  that was caused due to the patient movement during data acquisition. This

motion causes a degradation at the time of acquiring the k-space points. To simplify the problem formulation, we generally defined the image degradation in the spatial domain as follows:

$$x_m = \text{Tr}[x], \quad (1)$$

where  $\text{Tr}$  is the rigid motion transformation defined as  $\text{Tr} = F^{-1} \mathcal{M} T_\theta \mathcal{R}_\theta$ .  $T_\theta$  and  $\mathcal{R}_\theta$  are the translation and rotation operators in the image domain, while  $\theta$  indicates their motion parameters ( $t_x, t_y, t_z$ ) and ( $\rho_x, \rho_y, \rho_z$ ), respectively.  $F$ ,  $\mathcal{M}$ , and  $F^{-1}$  are the Discrete Fourier Transform (DFT), sampling operator in k-space, and inverse DFT, respectively. Basically, it is hard to solve this equation linearly since it could contain many unknowns (i.e., motion parameters). However, this task can be formed as an optimization problem using non-linear deep learning networks to derive the inverse transformation map that leads to obtaining the corrected image  $\hat{x}$ . We defined  $\text{net}_W$  as an optimization of network weights that inversely maps the motion transformation  $\text{Tr}$  as follows:

$$\hat{x} = \underset{\text{net}_W}{\operatorname{argmin}} \| \text{net}_W[x_m] - x \|_2^2. \quad (2)$$

It is assumed that the structural details of the white matter and gray matter in the brain are similar within the contiguous slices, and in particular, during the 3D imaging. Thus, the inclusion of additional prior knowledge from the adjacent slices of the same corrupted subject could assist to significantly solve this problem. The above equation can be re-formulated as:

$$\hat{x}^{[i]} = \underset{\text{net}_W}{\operatorname{argmin}} \| \text{net}_W[\text{Concat } FE_{w1}(x_m^{[i-1]}), FE_{w2}(x_m^{[i]}), FE_{w3}(x_m^{[i+1]})] \\ - x^{[i]} \|_2^2 \quad (3)$$

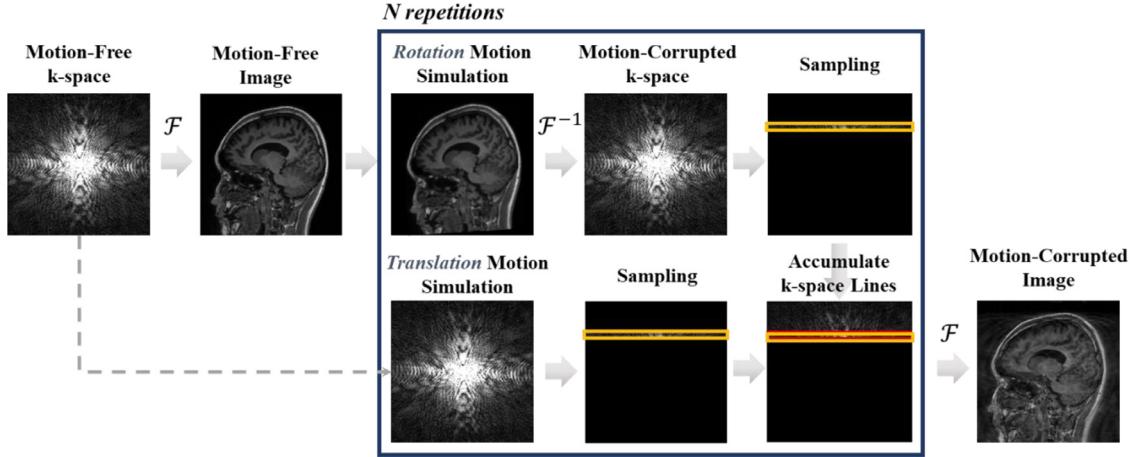
where  $x_m^{[i-1]}$  and  $x_m^{[i+1]}$  are the image priors derived from the adjacent slices of the current motion-corrupted image  $x_m^{[i]}$ .  $\text{Concat}(\cdot)$  refers to the concatenation function of the extracted features of all slices.  $w1$ ,  $w2$ , and  $w3$  are the weights of feature extractors  $FE_w(\cdot)$ , which are part of all network weights  $\text{net}_W$ . In spite of the fact that the utilized image priors are corrupted data, they can share some details that may be missed in the current  $i$ th image. It is of note that this is a multi-input single-output process, which corrects the motion artifacts for a single motion image at every time. Therefore, the requirement of learning additional structural information can be addressed through applying the concept of the proposed self-assisted priors.

### 2.4. Proposed motion artifact correction network

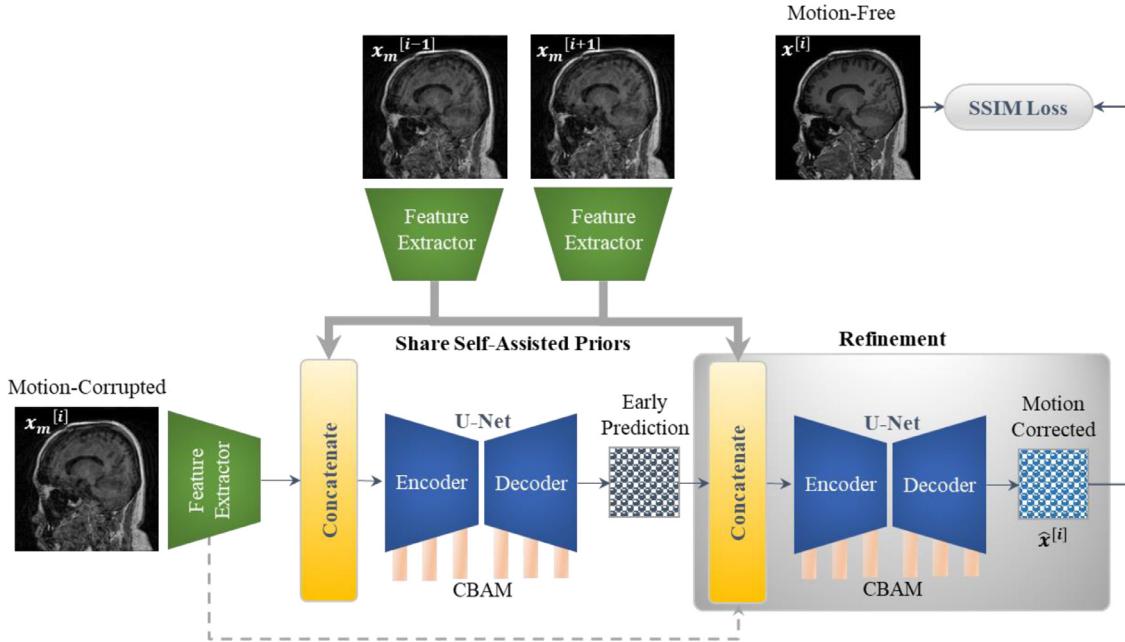
Inspired by U-Net (Ronneberger et al., 2015), which was first designed for image segmentation and then successfully applied to the denoising and reconstruction tasks, we propose 2D stacked U-Nets with self-assisted priors for image correction of the MRI motion artifacts. An overview of the proposed motion correction network is graphically illustrated in Fig. 2. The primary goal of designing such a deep learning network is to retrieve the motion-free image and increase the readers' ability to correctly diagnose patients from the structure of interest that does not include any obstacles such as artifacts. The essential contributions of the proposed architecture have both the self-assisted priors and stacked refinement network. The following subsections describe these concepts in detail.

### 2.5. Self-assisted priors (SAP)

The proposed network initiates by integrating the image priors from the adjacent slices of the target corrupted image. This is achieved by re-designing the original U-Net to be capable of importing and passing multiple inputs instead of a single input. A grayscale motion-corrupted image  $x_m^{[i]}$  along with its self-assisted priors  $x_m^{[i-1]}$  and  $x_m^{[i+1]}$  with a fixed image size of 256 × 256 pixels are passed independently to feature extractor encoders  $FE_w(\cdot)$ , generating 32 separate feature maps for



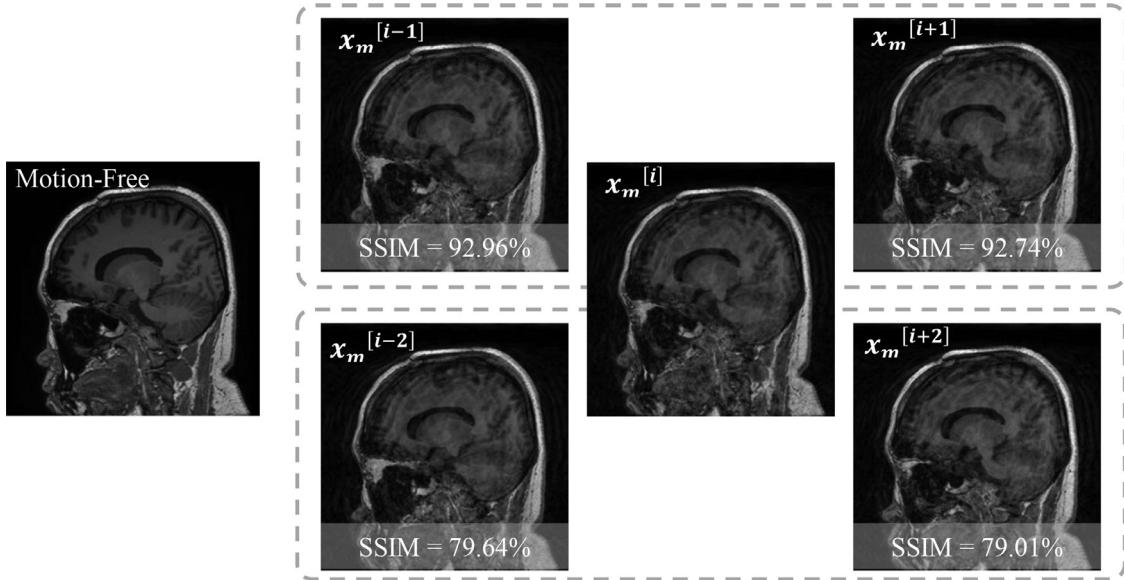
**Fig. 1.** Whole process of the motion artifacts simulation. Both the rotational and translational operators were utilized and proceeded in the image domain and k-space domain, respectively.



**Fig. 2.** Schematic diagram of the proposed stacked U-Nets with self-assisted priors for motion artifact correction. The attention module (CBAM) is applied to each resolution level of the encoder-decoder U-Nets.

each input. All these extracted features are then concatenated to serve as an input to the proposed network, as shown in Fig. 2. This process is applicable to all image slices within the training and testing subjects, except for the first and last slices, in which they were repeated due to the absence of the other prior. The inclusion of these image priors assists in sharing some structural details that may be missed in the target motion-corrupted image  $x_m^{[i]}$ , resulting in improvements in recovering the motion-free image. This is due to the fact that the brain structure is of high homogeneity in the contiguous locations, especially if the data acquisition were performed in 3D imaging strategy as the case of this work (i.e., no impact of the slice gap). In other words, the 3D imaging produces higher spatial resolution, which makes the usage of additional prior knowledge from the adjacent slices of great importance to retrieve the missed parts and, hence, achieve better motion correction. It should be noted that this is a multi-input single-output process, which implies that the network learns how to correct only the motion-corrupted image  $x_m^{[i]}$  in a supervised manner by computing the loss with the corresponding ground-truth motion-free image  $x^{[i]}$ . In each training

iteration, the role of the priors  $x_m^{[i-1]}$  and  $x_m^{[i+1]}$  is to feed additional useful patterns from the same corrupted subject into the network. However, correction of priors does not take place during the correction of the current motion-corrupted image  $x_m^{[i]}$ . Compared to Lee et al. (2021), Chatterjee et al. (2020), the proposed self-assisted priors approach has the following advantages: (i) it eliminates the need for additional MRI scans to be used as image priors, and (ii) it also reduces the computational cost since it does not require any further image pre-processing such as image registration and alignment. Nevertheless, if another contrast data is available, it could be used as an additional image prior besides the self-assisted priors of the same contrast. In this paper, we test different types of image priors, including both the self-assisted priors and the priors from other contrast data. Fig. 3 shows some examples of the adjacent slices with their SSIM scores related to the current corrupted slice. The relatively high SSIM scores of the closer adjacent slices ( $x_m^{[i-1]}$  and  $x_m^{[i+1]}$ ) in this figure exemplifies the benefits of our assumption of using self-assisted priors for better motion correction by sharing some missed structural details. It is observed that the



**Fig. 3.** Examples of using the adjacent slices ( $x_m^{[i-1]}, x_m^{[i+1]}$ ) and ( $x_m^{[i-2]}, x_m^{[i+2]}$ ) of a certain corrupted image  $x_m^{[i]}$  as additional self-assisted priors with their SSIM scores. The computed SSIM score of  $x_m^{[i]}$  with its ground-truth motion-free image is 68.86%.

farther slices ( $x_m^{[i-2]}$  and  $x_m^{[i+2]}$ ) have fewer similarities with the current motion-corrupted image. Therefore, we conducted this work using the self-assisted priors from only the closer adjacent slices since they contain higher similarities.

## 2.6. Stacked refinement U-Nets

The idea of stacked networks has been utilized in semantic segmentation (Shah et al., 2018; Jha et al., 2020) to refine the segmented target and achieve better performance. In this work, we proposed a new architecture consisting of two cascaded encoder-decoder U-Nets. We developed the 2D stacked U-Nets with the reuse of image priors throughout all networks. More specifically, the early prediction of the first encoder-decoder U-Net was concatenated with the image priors and their corresponding motion-corrupted image and passed into the second encoder-decoder network. This connection enables the second network to capture more spatial differences among the two predictions and hence, better refine of the corrected images. The proposed stacked refinement network was trained in an end-to-end manner. Thus, the initial estimates were re-evaluated during the training process. In the end, the proposed stacked U-Nets with the benefits of sharing additional knowledge could improve the pixel-to-pixel dependency and lead to recovering motion-free images with significant preservation of spatial information.

## 2.7. Attention module

We also adopted the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) into all resolution levels of the proposed stacked networks. The CBAM module was achieved by integrating two sequential attention maps along the channel and spatial directions. The channel attention module ( $Att_C$ ) benefits from the inter-channel relationship of feature maps and attempts to tell ‘what’ is the most meaningful map (i.e., feature detector). This attention module is derived by firstly squeezing a given feature map  $F$  into two independent vectors using max-pooling and average-pooling operations. Then, these two descriptors are passed independently into a shared multi-layer perceptron (MLP) of a single hidden layer. Both the generated feature vectors from the MLP are aggregated using element-wise summation to produce a channel attention map.

Complementary to channel attention, the spatial attention module ( $Att_S$ ) exploits the inter-spatial relationship of a given feature map and

highlights ‘where’ is the most effective informative part. This attention module begins by applying max-pooling and average-pooling along the feature channel direction. Then, a convolutional layer is applied to the concatenated feature descriptors to produce a spatial attention map. Eventually, both the  $Att_C$  and  $Att_S$  get multiplied sequentially to their input feature maps as  $F' = Att_C(F) \otimes F$  and  $F'' = Att_S(F') \otimes F'$  for adaptive feature refinement, respectively. Supplementary Fig. S1 shows a diagram of integrating both channel attention with spatial attention within CBAM. This process can improve the representation of interests and focus more on regions that may need extra correction.

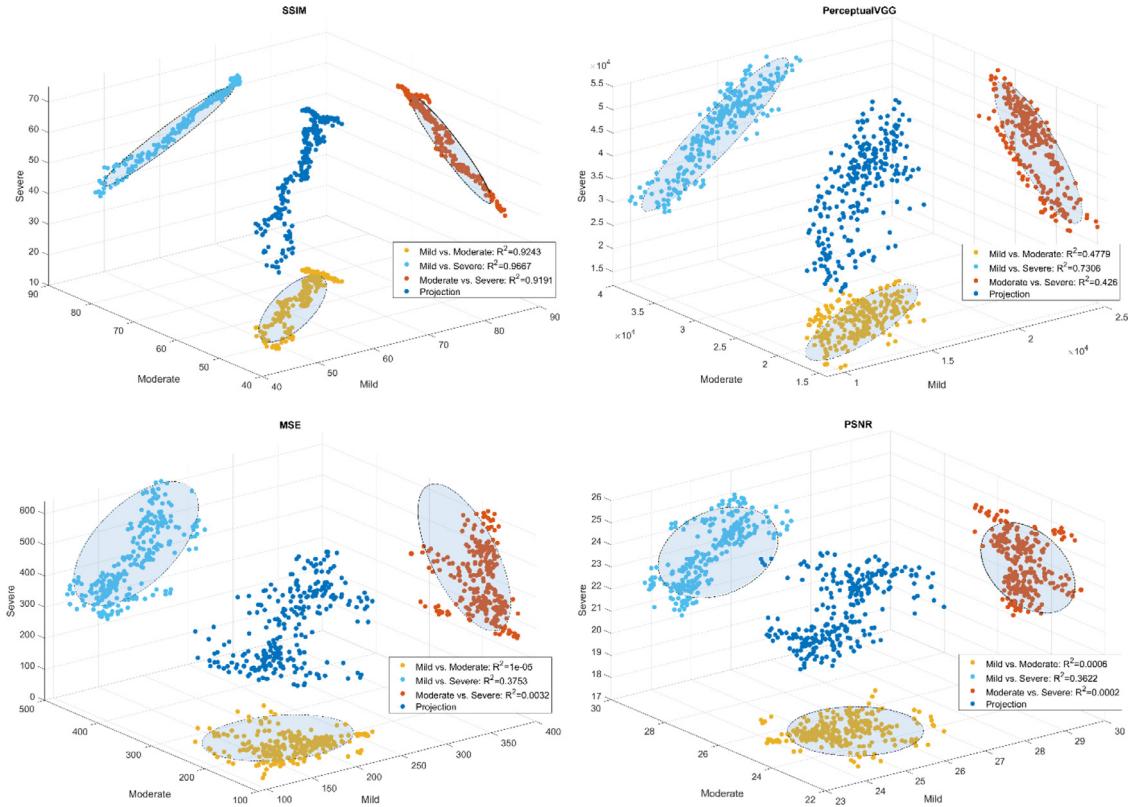
## 2.8. Implementation details

In this work, the proposed network was trained in a supervised manner using the simulated motion-corrupted images with their corresponding ground-truth motion-free data. The exploited U-Net contains four encoder-decoder levels. The convolution kernels throughout all networks have the size of  $3 \times 3$  with feature maps of 32, 64, 128, and 256 in the encoder path, while reverse feature maps were utilized in the decoder path followed by a final output prediction map. We sequentially applied the batch normalization and activation function of Rectified Linear Unit (ReLU) over all the convolutional layers. We used the average pooling with a filter size of  $2 \times 2$  and a stride of 2. It is noteworthy that the feature maps in each resolution level of the encoder path were concatenated with the corresponding maps of the same resolution level at the decoder path. This allows fusing both the global and local representations among early and late convolution layers. During network training, we utilized the Adam optimizer with a batch size of 10 samples for gradient descent. We initially set the learning rate to 0.001, and it is decayed exponentially by a factor of 10 throughout 50 epochs. The proposed network took around 5.8 h to accomplish its training over 4.01 million parameters.

The implementation of this paper was performed on a PC equipped with the following specifications: a Cuda-enabled GPU of NVIDIA GeForce RTX 3080 with 128 GB RAM. This work was implemented using Python 3.7.10 and Tensorflow with Keras library on the Ubuntu 18.04 operating system.

## 2.9. Loss measures

Previous studies for image translation used loss functions such as the SSIM, Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), or



**Fig. 4.** Relationship between various motion severity levels in terms of SSIM, MSE, PSNR, and PerceptualVGG. The correlations of various motion pairs are reported using the R-squared values. Each data point represents the loss measurement of a single image of a certain motion type, while the axes of each plot refer to the same index of the motion artifacts pairs.

perceptual losses; however, it is not certain as to which measure would be most sensitive for training motion correction methods. While MSE has been utilized as a loss function in many image denoising and reconstruction applications, for motion correction tasks, a mixture of loss functions including SSIM has been used (Lee et al., 2021). This is due to that motion artifact significantly degrades the image structure and the SSIM loss basically computes the relation of the structural similarities among the reference and target images.

The primary motivation of this experiment is to exhibit which loss is more sensitive to different types of motion artifacts. Thereby, the most effective measure can represent the better loss function to be utilized for network optimization. Ideally, the loss function value should be proportionate to the level of motion. To perform this experiment, we generated three different types of motions, namely mild, moderate, and severe, for the same patient data (Kecskemeti and Alexander, 2020). Our analysis concentrated on the SSIM, MSE, PSNR, and Perceptual, from the pre-trained Visual Geometry Group Network (VGGNet) on ImageNet, since they are usually utilized in the denoising tasks. We then computed all measurements for each motion severity level corresponding to the original non-motion data. Fig. 4 presents the relationships among various types of motion artifacts for each loss index. Each data point in the sub-figure represents the loss measurement of a single image of a certain motion type, while the projection plot shows the relationship between three motion artifacts. We also show the R-squared value of the regression line for the motion artifacts pair (i.e., mild vs. moderate, mild vs. severe, and moderate vs. severe). The 3D projection plot for each loss measurement provides an intuition of the data points distribution and could simply show the trend of three merged motions within the same measure. This experiment shows that the SSIM index has higher R-squared values for all types of motion artifact pairs compared to the other evaluation measures. Thus, the SSIM index provided the best in-

dication on distinguishing various motion artifact strengths (i.e., mild, moderate, and severe motions) compared to the MSE, PSNR, and PerceptualVGG. It is of note that the PerceptualVGG was computed by passing all the motion-corrupted images and their reference motion-free data into a fixed pre-trained model such as the VGGNet and then quantifying the feature differences among them over a certain layer within the network. In this work, the perceptual loss was computed on the 7th convolutional layer of VGGNet, named ‘block3-conv3’. In terms of correlations between different motion pairs, the SSIM index shows the best correlations with very high R-squared values of 0.9243 for mild vs. moderate, 0.9667 for mild vs. severe, and 0.9191 for moderate vs. severe. It is also observed that the PerceptualVGG shows a better correlation of different motion types compared to the MSE and PSNR. Hence, this investigation provides a justification for using the SSIM as a loss function. It is of note that although there have been many works using the VGG loss, we quantitatively investigate the performance using the SSIM, MSE, and PSNR since they have been widely utilized in motion artifact correction studies.

### 3. Experimental results

#### 3.1. Ablation studies

The goal of this section is to experimentally investigate different ablation studies that make the proposed network powerful enough to correct the motion artifacts from MR images, such as the inclusion of image priors from the same corrupted subjects and the incorporation of additional data from different image contrast. Throughout all these investigations, we have monitored the motion correction performance using the SSIM and MSE indices.

**Table 1**

Experimental results of the proposed stacked U-Nets with the self-assisted priors (SAP) compared to the corrupted simulation results.

ID	Experiment	SSIM (%)	MSE	PSNR
SAP-0	Corrupted Simulated Data	71.66	99.25	28.83
SAP-1	Single U-Net with Single Corrupted Input (No Priors)	94.20	37.06	32.87
SAP-2	Single U-Net with Inclusion of Self-Assisted Priors	94.44	33.87	33.27
SAP-3	Single U-Net with Inclusion of Self-Assisted Priors and Attention Module	94.64	33.85	33.25
SAP-4	Stacked U-Nets with Inclusion of Self-Assisted Priors and Attention Module	95.03	29.76	33.81

### 3.2. Image priors from same corrupted subjects

This section presents the ablation experiments for the main contributions of this paper. As illustrated in **Table 1**, the motion simulation tool produced motion-corrupted images with a degradation in the image quality of 71.66%, 99.25, and 28.83 in terms of SSIM, MSE, and PSNR, respectively. The first experiment was straightforward by passing a single corrupted image into the original supervised U-Net. The network learned how to reduce the motion artifacts by measuring the SSIM loss between the input corrupted images  $x_m^{[i]}$  and its reference motion-free images  $x^{[i]}$ . This experiment showed an improvement in the motion correction compared to the corrupted simulated data with SSIM of 94.20% and MSE of 37.06%. In the second experiment, we included additional knowledge from the adjacent slices ( $x_m^{[i-1]}$  and  $x_m^{[i+1]}$ ) of the same corrupted image into the network. We called this process self-assisted priors (SAP) since all the incorporated information was derived from the same patient. This yields a network with a multi-input single-output structure, which implies that at each iteration the network takes the corrupted image and its adjacent slices as inputs and be able to construct and correct the motion of only the current slice. This experiment could achieve a slight improvement rate of 0.22% in term of SSIM, while it significantly reduced MSE from 37.06 to 33.87. Next, we examined the effect of adding the CBAM attention module into the network. We could gain some improvements in the overall performance of the motion correction with an SSIM score of 94.64%. In the last experiment, we trained and tested all the above components via the stacked U-Net (see **Fig. 2**). The main purpose of this stacked architecture was to get better refinement of the motion correction image. The self-assisted priors were also shared with the second stacked network. As presented in **Table 1**, the proposed stacked U-Nets with the inclusion of the self-assisted priors achieved significant improvements with SSIM of 95.03%, MSE of 29.76, and PSNR of 33.81.

### 3.3. Additional prior from different contrast-enhanced (CE) data

Despite the cost, time, and patient inconvenience, the acquisition of additional MRI scans could be crucial in some clinical applications, especially the CE images. Fortunately, there were 38 patients among our dataset that had additional CE data with the same imaging parameters. The CE data has the exact same FOV and resolution as the non-contrast data. This section exhibits the importance of using the CE data as an additional image prior besides the self-assisted priors. To accomplish this experiment, we have split this new dataset based on the subject level into 80% for training (3,684 images) and 20% for testing (1,378 images). Here, we only used the 38 simulated motion-corrupted data from the T1-weighted BRAVO subjects with their corresponding CE data. Thus, this is a smaller dataset compared to the previous experiments. It is of note that in this experiment, only the non-CE T1-weighted BRAVO data contains motion artifacts, while the CE data itself is motion-free. **Table 2** presents the effect of using additional CE data as an image prior besides the self-assisted priors. The image quality of this simulated motion-corrupted dataset was reduced to 68.54%, 123.23, and 27.48 in terms of SSIM, MSE, and PSNR, respectively. All the investigations in this section were performed utilizing the stacked U-Net architecture. To examine the effect of using different types of image priors, we started

training the stacked networks without incorporating any image priors (i.e., neither the self-assisted priors nor the CE data prior). This stacked network without any image priors reduced the motion artifacts and obtained the SSIM score of 91.77% and MSE of 62.96. However, significant improvement was achieved by the inclusion of only the self-assisted priors. The performance was improved in term of SSIM from 91.77% to 92.10%, while the MSE was dramatically reduced in term of MSE from 62.96 to 53.55. It is noteworthy that for the CE prior, we utilized the same image slice  $CE^{[i]}$  of the target motion-corrupted image  $x_m^{[i]}$ . Also, promising achievement was obtained in the case of using only the CE data as image prior. It improved the performance with overall increments in term of the SSIM from 91.77% to 93.03% and decrements in term of MSE from 62.96 to 59.62. It is observed that the inclusion of self-assisted priors achieved a better MSE of 53.55 compared to the use of CE prior, which obtained 59.62. In opposite, the performance of the proposed network with the CE prior outperformed the network with the inclusion of the self-assisted priors with scores of 93.03% and 92.10% in terms of SSIM, respectively. Next, we integrated both priors together. Thus, to correct the target motion-corrupted image  $x_m^{[i]}$ , we used four inputs into the network. Three of them are the image priors ( $x_m^{[i-1]}$ ,  $x_m^{[i+1]}$ , and  $CE^{[i]}$ ) and last one is the target image itself ( $x_m^{[i]}$ ). This approach results in maintaining a SSIM score of 93.04% and keeping a MSE of 54.06. In order to address the issue of unregistered CE data to the non-contrast data, which may occur due to motion between scans, we applied more priors from CE adjacent slices (anatomical priors) besides the self-assisted priors of the same corrupted data. This implies that five priors of  $x_m^{[i-1]}$ ,  $x_m^{[i+1]}$ ,  $CE^{[i]}$ ,  $CE^{[i-1]}$ , and  $CE^{[i+1]}$  can be utilized for further improvement of motion artifact correction in the case of not perfectly aligned anatomies in CE data. This experiment achieves the highest SSIM score of 93.25%. We conclude that if additional MRI scans are available, they can be used as image priors besides the self-assisted priors to further enhance the performance of motion artifacts correction. Qualitative experimental results for each presented method in **Tables 1** and **2** are shown in Supplementary Figs. S2 and S3.

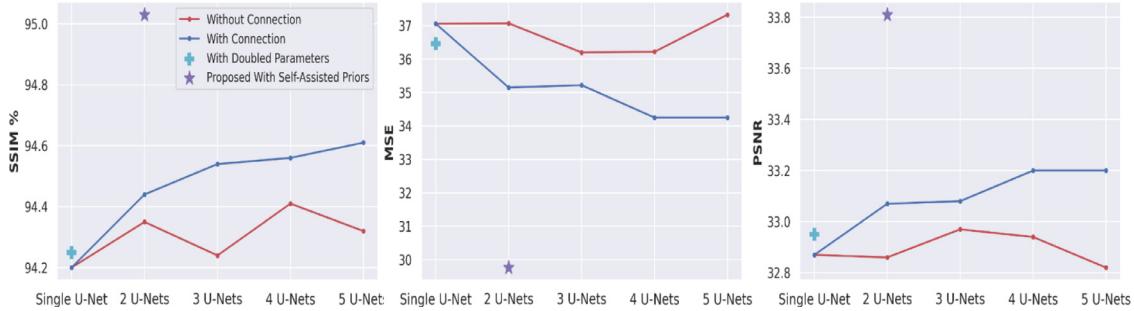
### 3.4. Impact of stacking multiple U-Nets

This experiment aims to demonstrate the effect of sequentially stacking several U-Nets on the performance of the motion artifacts correction. We performed this investigation using two different scenarios as follows. The first scenario was conducted by only passing the output of each U-Net to the subsequent networks without any connections of the preceding outputs of the early U-Nets as well as neither passing the original input. In comparison, the second scenario proceeded throughout connecting all the preceding outputs together with the original input into the subsequent networks. **Fig. 5** illustrates the impact of stacking up to five U-Nets in terms of SSIM, MSE, and PSNR. The network sizes from single U-Net to the stacked five U-Nets shown in this figure are 1.95, 3.90, 5.85, 7.80, and 9.75 million parameters. Obviously, in the case of no connections, the fluctuation of overall performances has been recorded among different stacked U-Nets. However, the motion correction performance was dramatically improved when the number of stacked networks increased when using connections from the preceding networks. This experiment demonstrates the significant role of choosing an appropriate input into every stacked architecture. In other

**Table 2**

The effect of using additional CE data as an image prior besides the self-assisted priors (SAP-CE).

ID	Experiment	SSIM (%)	MSE	PSNR
SAP-CE-0	Corrupted Simulated Data	68.54	123.23	27.48
SAP-CE-1	Stacked U-Nets without any Image Prior	91.77	62.96	30.43
SAP-CE-2	Stacked U-Nets with only Self-Assisted Priors	92.10	53.55	31.06
SAP-CE-3	Stacked U-Nets with only CE Data as Prior	93.03	59.62	30.68
SAP-CE-4	Stacked U-Nets with Integrated Self-Assisted Priors and CE Data Prior	93.04	54.06	31.04
SAP-CE-5	Stacked U-Nets with Integrated Self-Assisted Priors and CE Data Prior Along with CE Adjacent Slices	93.25	54.51	30.96



**Fig. 5.** Impact of stacking multiple U-Nets with and without using the connections from early predictions of the preceding networks. Also, the case of using single U-Net with doubled parameters is shown as a plus sign. The overall motion correction performance of the proposed self-assisted priors is presented as a star sign.

words, bigger networks do not always reflect better performance. The MSE and PSNR plots in Fig. 5 show that large improvement rates were obtained in the case of using two stacked U-Nets. Also, the MSE and PSNR results of the two stacked U-Nets are almost the same as those of three stacked U-Nets with smaller network parameters. Thus, we considered using two stacked U-Nets in this work since it benefits from learning early predictions along with the original inputs and also due to the reasonable network size. Furthermore, we performed an additional experiment to clarify whether the stacked two U-Nets really helped the motion artifacts correction with the early prediction and the reutilized of the original input or the bigger network showed the better performance because it had doubled parameters. This experiment is shown as a plus marker in Fig. 5 and called a single U-Net with doubled parameters. Intelligibly, the results showed that even though both networks (i.e., stacked two U-Nets and single U-Net with doubled parameters) were trained using almost the same number of parameters, the stacked network obtained significantly better performance. These results indicate the great importance of using early prediction for better improvement of motion correction.

In addition, we presented the significance of using the proposed self-assisted priors with the stacked two U-Nets as a star sign in the same figure. The performance showed too much improvement compared to all stacking U-Nets.

### 3.5. Effect of increased slice spacing on priors from adjacent slices

In Fig. 3, we showed the impact if the further apart slices, not the direct neighboring slices, are considered priors. It is observed that the farther slices ( $x_m^{[i-2]}$  and  $x_m^{[i+2]}$ ) have fewer similarities with the current motion-corrupted image. This section demonstrates the impact of increased slice spacing on the overall motion correction performance. We retrained and tested the proposed stacked U-Nets using self-assisted priors from different neighboring slices. We started with direct adjacent slices ( $x_m^{[i-1]}$  and  $x_m^{[i+1]}$ ) that contain similar anatomical information with the current motion-corrupted image ( $x_m^{[i]}$ ). This investigation continues until the increased slice spacing reaches  $\pm 10$ , which presents the most extreme case with almost no overlapping anatomical details. Fig. 6 shows the results of this investigation in terms of SSIM, MSE, and PSNR. The results confirmed our assumption of the self-assisted priors, which could share valuable knowledge about the missing information. The per-

formance of motion artifact correction was degraded if self-assisted priors were derived from the farther slices.

Interestingly, these adjacent slices serve as auxiliary information to the network and are connected with their weights in the feature extractor block  $FE_{w1}(x_m^{[i-1]})$  and  $FE_{w3}(x_m^{[i+1]})$  (see Fig. 2 and Eq. (3)). Since the proposed network is trained in a supervised manner, these weights ( $w1$  and  $w3$ ) get optimized to determine the contributions derived from these priors, if any. Therefore, this work was conducted using the self-assisted priors from only the closer adjacent slices  $\pm 1$  since they contain higher similarities.

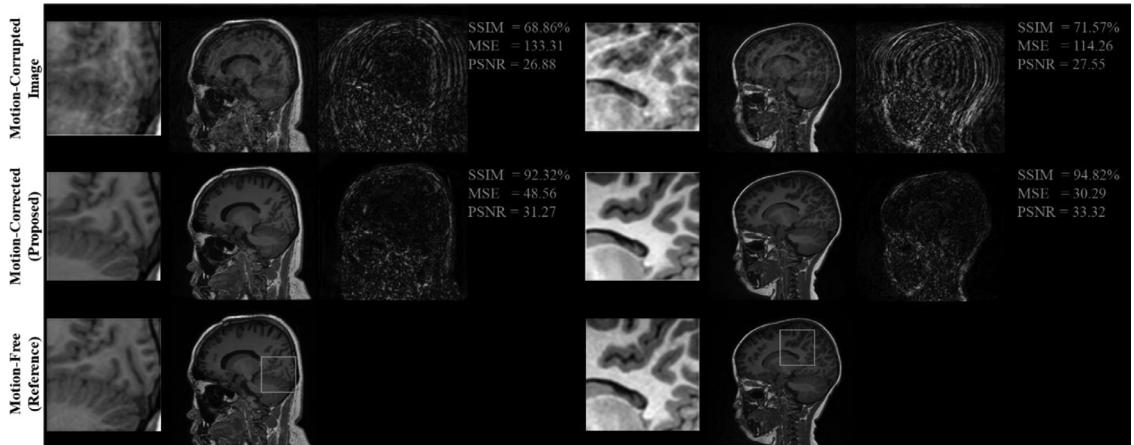
### 3.6. Motion correction results on simulated dataset

This section shows the motion artifacts correction performance of the proposed stacked U-Nets using two testing situations. The first testing set (SAP) includes only the image prior from the same corrupted subjects (i.e., the proposed self-assisted priors from the adjacent slices). We evaluated this category using 3,390 corrupted simulated images. The second examination involves the self-assisted priors with the incorporation of additional prior from different CE data (SAP-CE), which has a total of 1378 corrupted images. As presented in Tables 1 and 2, the proposed deep learning network was able to learn additional knowledge of the motion patterns from various image priors, leading to achieve promising results in the motion artifacts correction task. For the SAP experiments, the proposed network provides promising results in motion artifacts correction with significant improvement rates of 23.37%, 70.02%, and 17.27% in terms of SSIM, MSE, and PSNR compared to the simulated motion-corrupted data, respectively. Furthermore, the proposed stacked network with the integration of both the self-assisted priors and CE prior (SAP-CE) achieves promising results in motion artifacts correction with significant improvement rates of 24.71%, 55.77%, and 12.66% in terms of SSIM, MSE, and PSNR compared to the simulated motion-corrupted data, respectively.

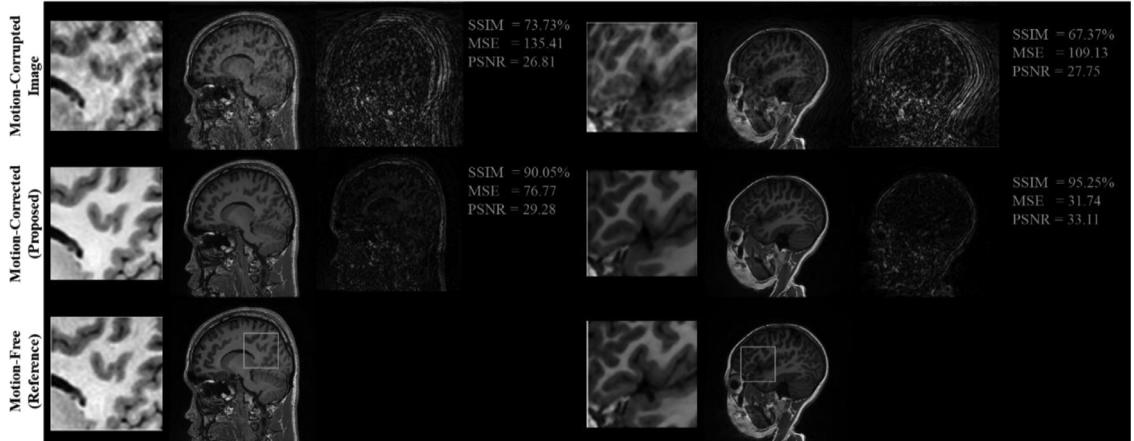
Fig. 7 illustrates some exemplar motion correction results of the proposed network compared to the ground-truth motion-free images from SAP and SAP-CE. This figure clearly shows how the proposed method can significantly improve the image quality and reduce motion artifacts. The zoomed regions of the grey rectangle inscribed on the reference image intelligibly revealed the capability of the proposed network to retrieve the corrupted structural details and provide comparable visual



**Fig. 6.** Effect of increased slice spacing on motion artifact correction. The results show a degradation of the overall performance in the case of self-assisted priors were derived from further apart slices.



(a)

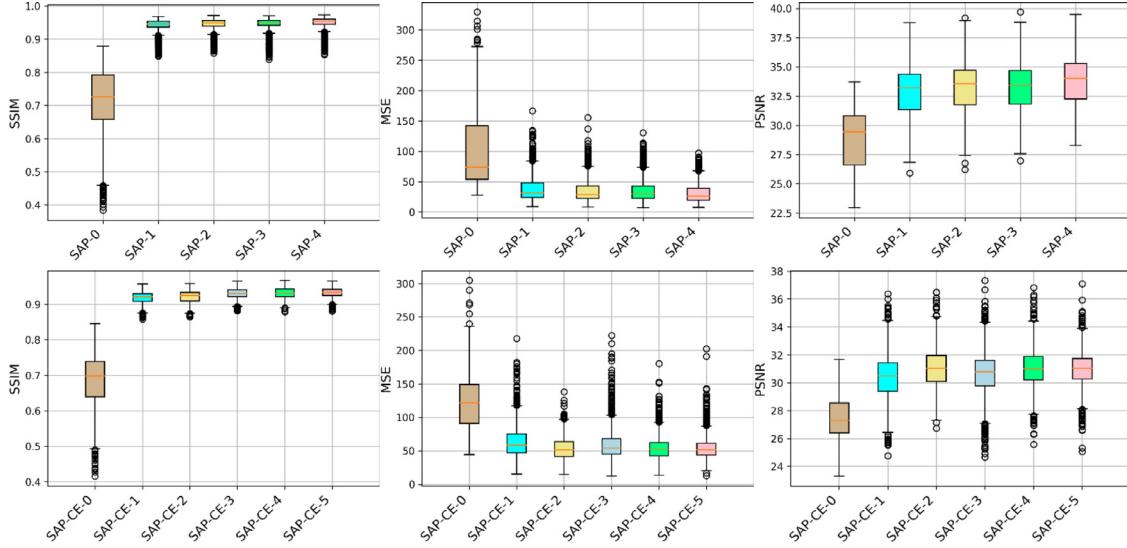


(b)

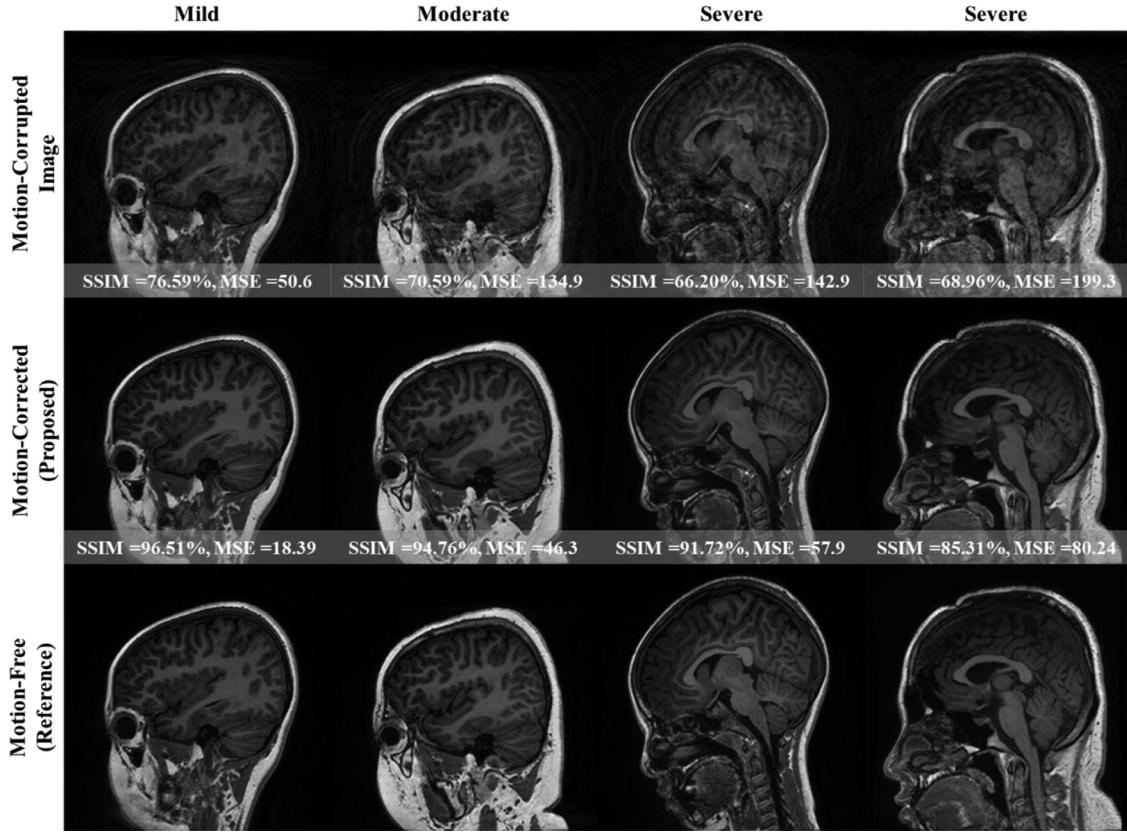
**Fig. 7.** Simulated motion artifacts correction results of the proposed stacked U-Nets with (a) self-assisted priors and (b) integration of self-assisted priors with CE prior. First column of each example shows the zoomed region of the grey rectangle that was inscribed on the reference image. Second column presents the motion-corrupted image, motion-corrected via proposed work, and the reference motion-free image. In the last column of each sample, we show the pixel-wise difference maps compared to the reference image. For each case, we report the evaluation indices before and after the correction of motion artifacts.

performances with the reference motion-free image. Thus, our findings seem to be feasible in increasing the reader's ability to examine and identify the presence of pathologies. One interesting observation from the difference maps in Fig. 7 is that the background regions of the corrected images were near totally recovered, while the motions over the brain parenchyma regions were notably corrected better than the skeletal structure of the head. Moreover, we illustrate the motion correction performances for each testing image throughout all the ablation experiments of SAP and SAP-CE in Fig. 8. These boxplots show the SSIM, MSE, and PSNR indices before and after correction of motion artifacts. As presented in this figure, the SSIM scores of different correction meth-

ods are relatively close to each other in both SAP and SAP-CE. It is shown that the proposed methods in SAP-04 and SAP-CE-05 obtained SSIM scores greater than 0.9 in the majority of testing images. The great achievements of the SAP-04 compared to other experiments were clearly demonstrated in the MSE and PSNR plots. In contrast, the MSE and PSNR for all testing images in SAP-CE-2 were the best compared to other experiments that included CE priors. Additionally, we present the corrected images for each type of motion artifacts (i.e., mild, moderate, and severe) in Fig. 9. It is shown how efficient the proposed method is in the correction of mild and moderate motions. However, there is still a limitation in correcting the images with a high motion severity level.



**Fig. 8.** Boxplots of the motion correction performances for each testing data in SAP (top row) and SAP-CE (bottom row). SAP refers to the results of the proposed network using the self-assisted priors, while SAP-CE indicates the results using additional prior from different CE data. Plots from left to right represent the SSIM, MSE, and PSNR scores, respectively. The abbreviations of ‘SAP-0~4’ and ‘SAP-CE-0~5’ represent various ablation experiments that were stated in Tables 1 and 2.



**Fig. 9.** Simulated examples of some corrected images for each type of motion artifacts (mild, moderate, and severe) using the proposed stacked U-Nets with self-assisted priors. It seems that both the mild and moderate motion cases can be corrected with very high SSIM scores. There is still a severity limit with that the proposed method cannot correct the motion artifacts efficiently as illustrated in the last example.

For instance, the last example in this figure shows a severe case where some inherent structures have deteriorated and become a burden to the network.

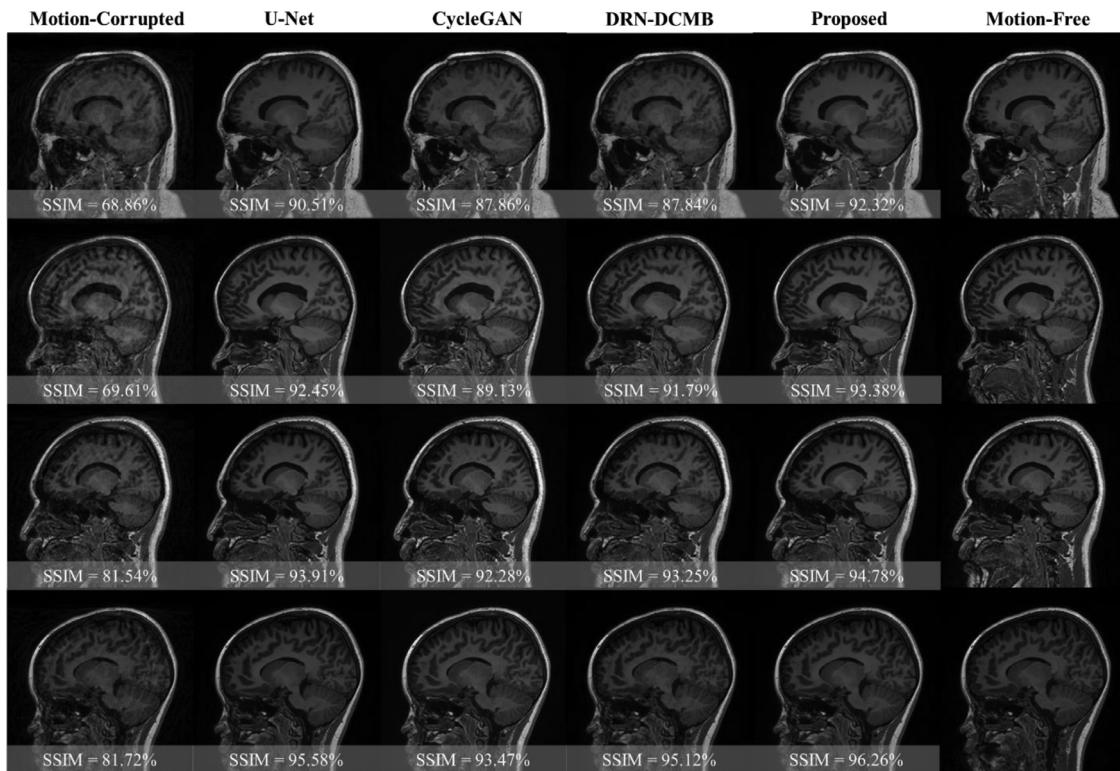
In Table 3, we compared the proposed method against the original U-Net (Ronneberger et al., 2015), Cycle Generative Adversarial Networks (CycleGAN) (Zhu et al., 2017), and DRN-DCMB (Liu et al., 2020). This

table presents a direct comparison with the re-implemented methods from the literature using the same simulated training and testing sets. It is of note that the CycleGAN was recently developed in some studies to correct MRI motion artifact (Khalili et al., 2019; Armanious et al., 2020; Chung et al., 2021). The results showed that the proposed network outperformed the U-Net, CycleGAN, and DRN-DCMB with overall

**Table 3**

Direct comparison of the proposed stacked U-Nets with self-assisted priors against recent deep learning works using the same simulated training and testing sets. The measures in parenthesis indicate the improvement rates.

Ref.	Subjects	Corrupted			Corrected		
		SSIM	MSE	PSNR	SSIM	MSE	PSNR
(Ronneberger et al., 2015)	83 (Sagittal)	71.66	99.25	28.83	94.20	37.06	32.87
U-Net (Implemented)					(22.54%)	(62.66%)	(14.01%)
(Zhu et al., 2017)					91.79	53.48	30.88
CycleGAN (Implemented)					(20.13%)	(46.12%)	(7.11%)
(Liu et al., 2020) DRN-DCMB (Implemented)					93.56	37.08	33.00
<b>Proposed (Self-assisted Priors)</b>					(21.90%)	(62.64%)	(14.46%)
					95.03	29.76	33.81
					(23.37%)	(70.02%)	(17.27%)



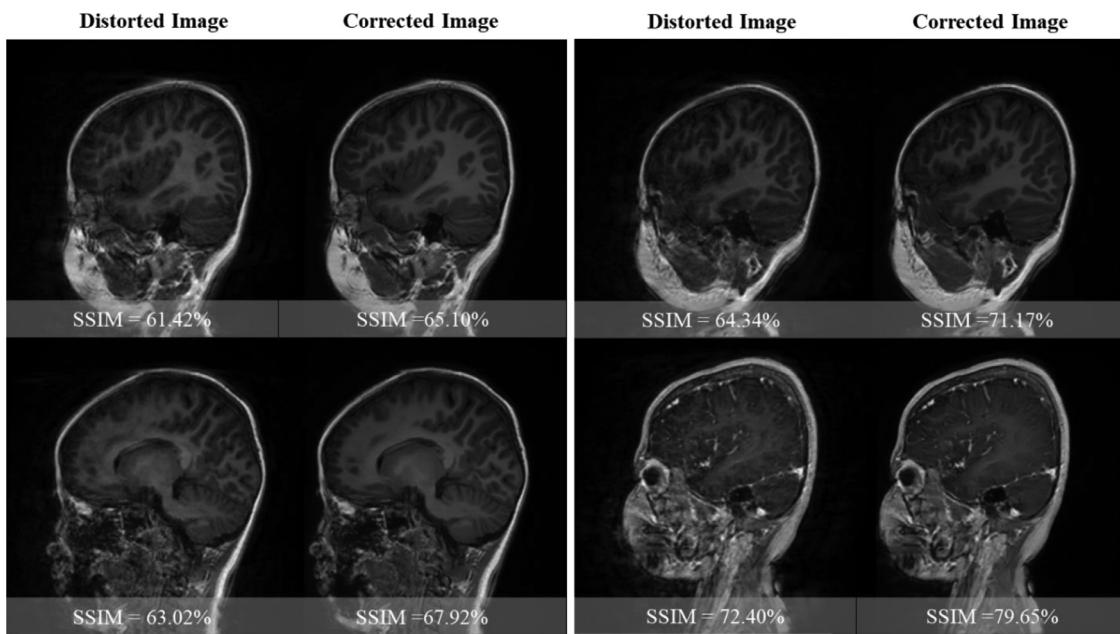
**Fig. 10.** Qualitative comparison of the proposed stacked U-Nets with self-assisted priors against the U-Net, CycleGAN, and DRN-DCMB methods on simulated data. All the SSIM values inscribed on the images are computed corresponding to the reference motion-free images.

improvement rates of 23.37% vs. 22.54%, 20.13%, and 21.90%, 70.02% vs. 62.66%, 46.12%, and 62.64%, and 17.27% vs. 14.01%, 7.11%, and 14.46% in terms of SSIM, MSE, and PSNR, respectively. Although the CycleGAN obtained remarkable success in the image-to-image translation tasks, it showed limited achievements to address the problem of motion artifact correction, especially if the level of motion in the corrupted image is high. A qualitative comparison is illustrated in Fig. 10. It is shown how efficient the proposed network is in producing high image quality that matches reference motion-free data. In contrast, CycleGAN struggled to reduce motion artifacts in the severe motions, as presented in the first two examples of Fig. 10. In these instances, CycleGAN generated results with some structural deterioration. However, CycleGAN obtained better motion correction performances in the cases of moderate motions compared to severe ones, as presented in the last two examples. Nevertheless, the proposed method still overcomes the CycleGAN, DRN-DCMB, and U-Net on different motion levels. An interesting observation is that the motion correction via U-Net achieved better performance compared to the CycleGAN method. The implemented DRN-DCMB obtained a comparable performance with U-Net. Unlike the original DRN-DCMB work that trained the network using around  $3 \times 10^6$

paired image patches, we trained the same network architecture using our few thousand images.

### 3.7. In vivo experiments: motion correction results on real patient data

To examine the efficiency and feasibility of the proposed motion artifacts correction method, we have tested the network on real clinical motion data. We show some real motion cases in Fig. 11. For these cases, clinicians recommended performing another data acquisition due to the poor image quality derived in the first scans that were affected by motion artifacts. The second scans were more clear data but with a slight misalignment with the motion distorted images. However, if we assumed that the second scans are the references (or semi-reference) of the distorted images, then the proposed network with self-assisted priors was able to reduce the motion artifacts and obtain improvement rates of 3.68%, 4.90%, 6.83%, and 7.25% in term of SSIM for all presented samples in this figure, respectively. It is noteworthy that the last example in this figure is CE data, while our network was only trained using non-CE data. Here, we tested real CE data that contained some motion and checked the possibility of our network to reduce the motion arti-



**Fig. 11.** Real motion cases of some subjects that have semi-reference from additional scan. Left images are the distorted real clinical motion images, while right images are the motion-corrected images via the proposed stacked network with the self-assisted priors. The bottom-right example shows motion correction improvement of a real CE case despite that the network was only trained using non-CE data. The SSIM values inscribed on the images are computed corresponding to the semi-reference images from additional scans.

fact on CE data. The proposed method still achieved promising results in motion artifact correction.

Further, Fig. 12 presents the motion correction results for some examples with various strengths of motion artifacts. For this kind of assessment, it is impossible to compute the quantitative indices due to the absence of reference motion-free images. Nevertheless, the motion artifacts seem significantly reduced and can be observed visually in the motion-corrected images. The first two examples in Fig. 12 show how efficient the proposed network is in retrieving clear brain structures from the real blurred images. In the third example, despite that the real distorted subject has a severe motion artifact, the proposed network could improve the image quality and reduce the motion artifacts. In the last example, it is observed that the acquired real data has extremely severe motion, which results in degrading the image quality to the level of destroying all the brain structural details. In spite of that, the motion-corrected image seems improved in some regions, and some structural patterns have appeared.

We also inspected the data consistency of the motion-corrected in other axial and coronal planes, as illustrated in Fig. 13 from an actual clinical case. It is clear that the reconstructed motion-corrected volumes maintain the uniformity of the brain structure while reducing the motion artifacts. Note that viewing from different planes shows interslice dependencies in the distorted images. The corrected images do alleviate this, but there are still some interslice differences. This could be a limitation of the study. It is of note that the sequence was based on a sagittal prescription, and radial fan-beam acquisition was performed in the ky-kz direction (AP, RL direction). The acquisition of our data was basically derived on sagittal 3D T1 BRAVO with in-plane resolution of  $0.90 \times 0.90 \text{ mm}^2$  and slice thickness of 1 mm. This implies that the sagittal view has a better image resolution compared to the other axial and coronal planes. Due to this, we have conducted this work utilizing the sagittal slices. After finishing the correction of all sagittal slices for each patient, the other planes get corrected simultaneously. Thus, there is still clinical feasibility since the clinicians can examine the patient brain at any view as presented in Fig. 13.

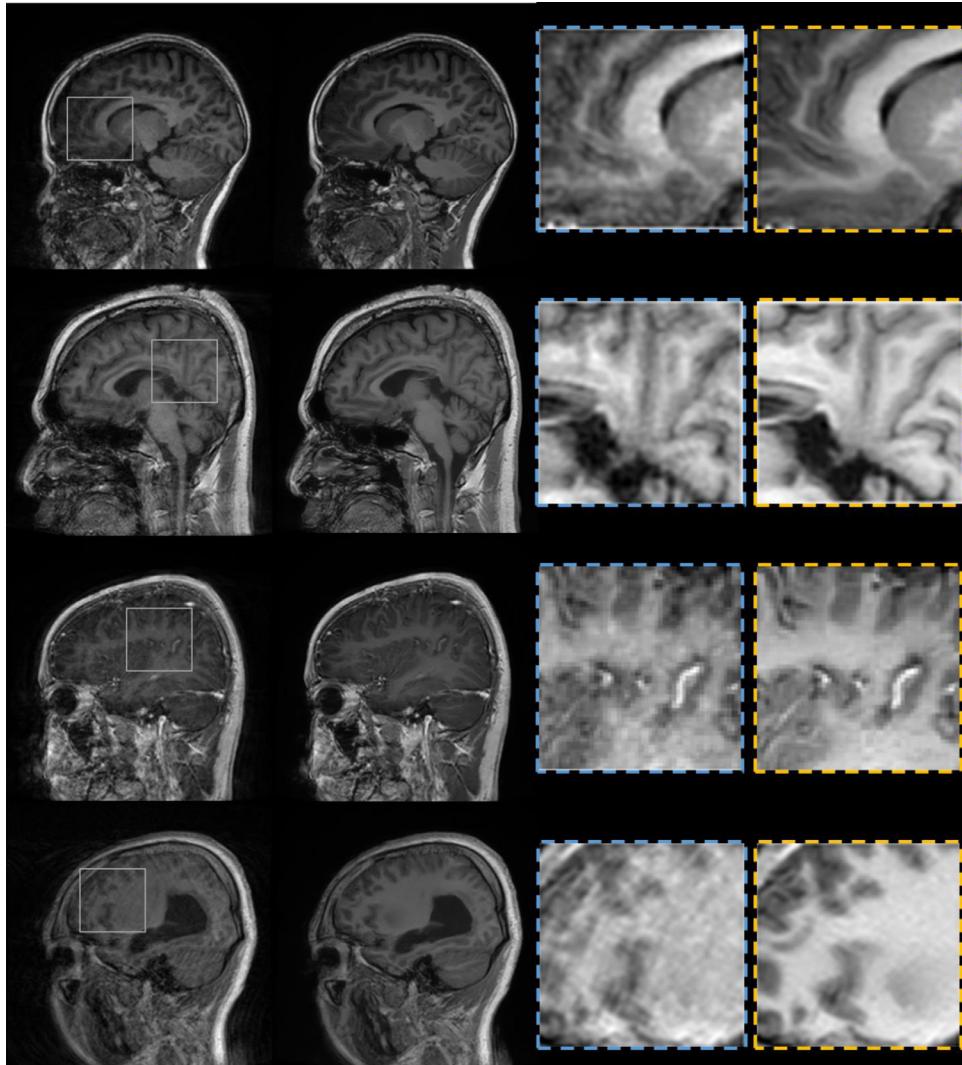
To demonstrate the feasibility of an *in-vivo* experiment with more severe cases, we tested the proposed method with additional 3D T1 brain

data acquired using a different 3.0 Tesla Skyra MRI scanner (Siemens Healthineers, Germany). The motion artifacts in these data manifest as a severe ringing. Fig. 14 shows the ability of the proposed stacked U-Nets with self-assisted priors to remarkably eliminate the ringing artifacts and provide better image quality over three planes (sagittal, axial, and coronal). These results reveal how efficient and generalized the proposed method is in correcting motion artifacts even with data from the different scanners.

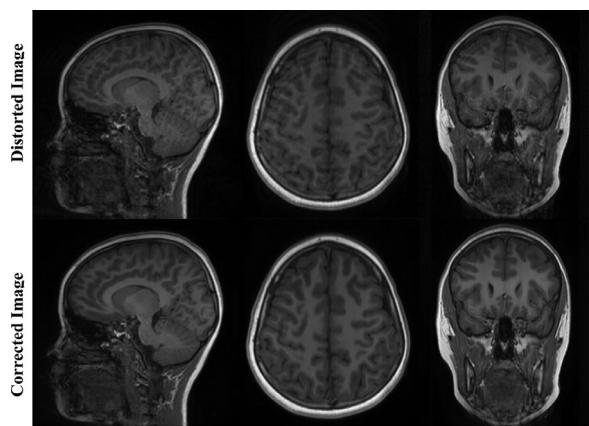
#### 4. Discussion

In this work, we developed a new supervised deep learning network called stacked U-Nets with self-assisted priors towards rigid motion artifacts correction. The proposed network secured two main contributions. The adjacent slices of the corrupted image enable the employment of prior knowledge. These self-assisted priors contain useful information that leads to enhancing the image quality. Additionally, the stacked U-Nets provide a refinement stage, offering prediction re-estimation and eventually preserving the spatial details. To accomplish the network training, it was inevitable to simulate MRI motion artifacts from motion-free data. This is due to the difficulty of acquiring different pairs of both motion-free and motion-corrupted data from the same patients. In this paper, we synthesized the motion-corrupted data using both rotational and translational motions simultaneously in the image and k-space domains, respectively. This procedure can produce various levels of motion severity similar to the real motion data by controlling the motion parameters.

The experiments in this work were conducted to investigate the robustness of the proposed network to enhance the motion artifacts correction from brain MRI data. The experimental results shown in Tables 1 and 2 demonstrated the significance of incorporating the additional priors information to improve the overall performance of motion correction. These inclusions of image priors from the adjacent slices of the same corrupted subjects or from other image contrasts (i.e., CE data in our work) enable the network to share some missing structural patterns such as borders of the white and grey matters in the brain. This idea offers a potential solution to correct the MRI motion artifacts with a



**Fig. 12.** Results of motion artifacts correction for four clinical subjects. First column indicates the real images with motion artifacts, while second column represents the motion-corrected images via the proposed stacked U-Nets with self-assisted priors. The blue and yellow boxes in the right highlight the zoomed regions of both real motion-distorted and motion-corrected via proposed network, respectively.

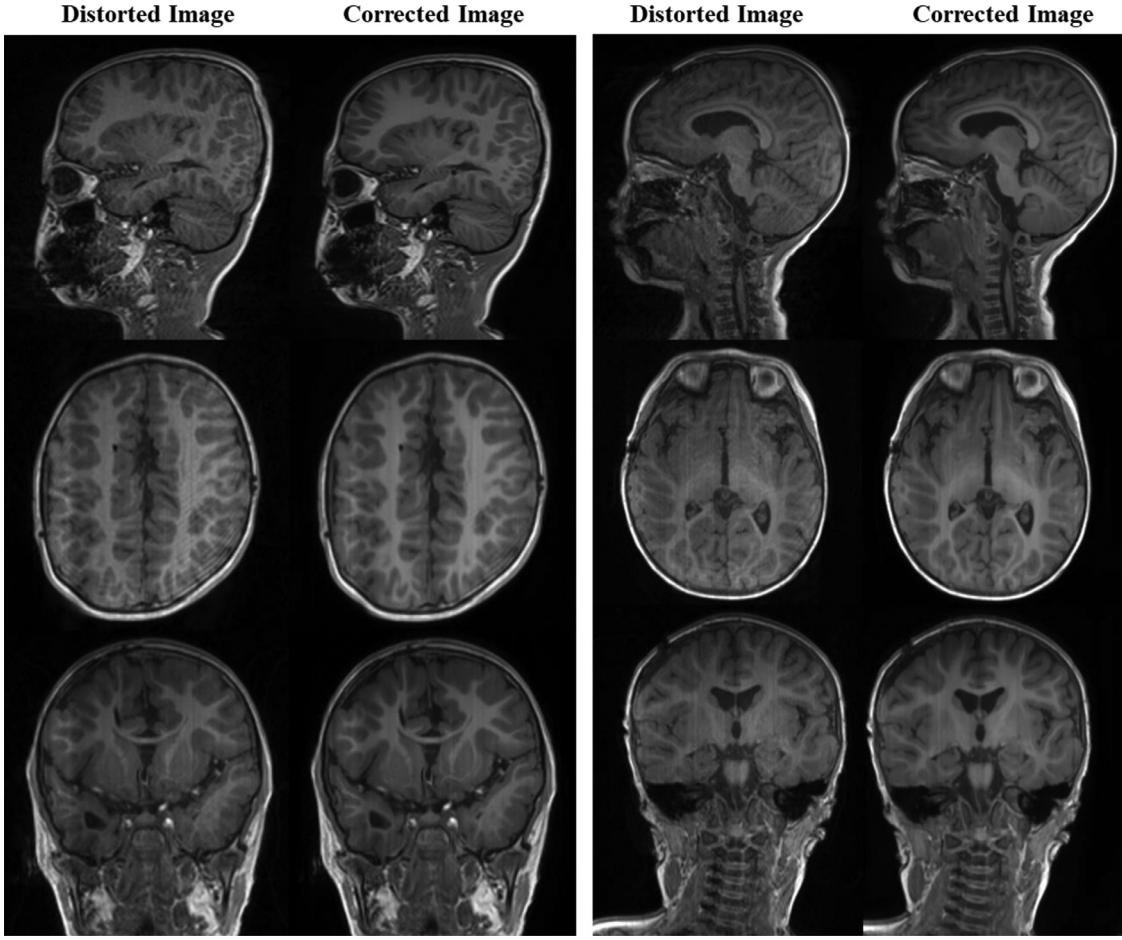


**Fig. 13.** Performance of motion correction results on the other axial and coronal planes of a real clinical patient. It is of note that the network was trained and tested using only the sagittal images.

significant improvement compared to the simulated corrupted data. For the case of using only the self-assisted priors from the corrupted data itself, the overall performance was significantly enhanced. We also improved the SSIM score in the correction of motion artifacts in the case

of incorporating both self-assisted priors with the additional CE prior data, but with a bit degradation of the MSE compared to the case of only using the adjacent slices. These results indicate the high similarity of the brain's anatomical structure in the corrected image compared to its motion-clean image. In both cases, the proposed motion correction method quantitatively outperformed the case without utilizing any priors, as presented in [Table 2](#). Generally, the proposed self-assisted priors have the ability to share valuable patterns from the contiguous slices and assist in retrieving better image quality. Other image contrasts such as CE data are of high accessibility in realistic clinical exams. Thus, employing both image priors (i.e., self-assisted priors and extra image contrast priors) of the same subject could improve the performance of artifact motion correction.

A comparison against the latest studies in the literature has been reported in [Table 4](#). As presented, each work had generated simulated input data with various levels of motion severity and utilized different amounts of training and testing MRI subjects. Nevertheless, all methods achieved improvements in reducing motion artifacts from MRI images. It is observed that the work in [Lee et al. \(2021\)](#) generated simulated motion data with mild motion and hence obtained lower improvement rates throughout three different image contrasts compared to other works listed in this table. The limitations of this work are as follows. First, even though the proposed self-assisted priors' strategy was beneficial to improve the motion correction by sharing some missing structural details, the utilized adjacent slices were derived from the same corrupted



**Fig. 14.** Motion artifacts correction of real clinical examples over three planes with severe ringing artifacts acquired using a different scanner (Siemens Healthineers, Germany).

**Table 4**

Comparison of the proposed stacked U-Nets with self-assisted priors against the latest works in the literature on MRI motion artifact correction. The measures in parenthesis indicate the improvement rates.

Ref.	Subjects	Corrupted			Corrected		
		SSIM	MSE	PSNR	SSIM	MSE	PSNR
(Wang et al., 2020)	1113 (Sagittal)	40.00	-	18.50	89.00	-	29.60
2D Motion Correction		74.00	-	27.20	89.00	-	(60.00%)
(Wang et al., 2020)	3D Motion Correction	86.70	-	-	96.50	-	32.40
(Liu et al., 2020)		85 (Multi-Planes)	-	-	(15.00%)	-	(19.12%)
(Chatterjee et al., 2020)	100 (Axial)	77.00	-	-	97.00	-	-
(Lee et al., 2021) (T <sub>1w</sub> )		98.11	-	-	(20%)	-	-
(Lee et al., 2021) (T <sub>2w</sub> )	41 (Axial)	91.25	-	-	99.05	-	-
(Lee et al., 2021) (FLAIR)		92.79	-	-	(0.94%)	-	-
Proposed (Self-assisted Priors)	83 (Sagittal)	71.66	99.25	28.83	95.03	29.76	33.81
Proposed (Additional CE Prior)	38 (Sagittal)	68.54	123.23	27.48	(23.37%)	(70.02%)	(17.27%)
					93.25	54.51	30.96
					(24.71%)	(55.77%)	(12.66%)

data. That implies there is still a loss of complete information. Due to this, most motion correction works cannot resolve this problem completely. Thus, we attempted to learn our proposed network using more additional knowledge throughout including the CE image prior. It may be interesting if multiple images of the same subject can be used as additional image priors besides the self-assisted priors for further im-

provement of motion artifacts correction. However, there is a tradeoff since this requires additional cost by acquiring new MRI scans. Second, it seems there is a difficulty to recover the very severe motion cases with severe blurring or ghosting. A complete loss of the borders of the white and grey matters can be clearly noticed, making it a challenging case for any motion correction method. Such cases may lead to an inaccurate di-

agnosis, and therefore the MRI scan must be repeated. Third, there is an absence of paired real motion-free and repeated motion-corrupted scans. All the motion artifacts correction works have trained their networks using simulated motion data. An interesting work could be to collect an open-access database to address this research topic. Although the MR data is inherently complex-valued, this work conducted data processing using only the real-valued magnitude images. Since the magnitude images do not have complete information of the acquired k-space data, the inclusion of phase information or processing data on raw k-space complexed-valued might assist in learning better knowledge of the missing details.

## 5. Conclusion

This paper presents a deep learning motion artifact correction method called stacked U-Nets with self-assisted priors. This work aims to develop a reliable motion correction method without multi-contrast MR images that require additional scans and high computational costs. The proposed network exploits the employing of the adjacent slices of each corrupted image to learn additional knowledge that may contain useful structural details. Specifically, the 3D imaging enables the construction of higher spatial resolution, which makes the usage of additional prior knowledge from the adjacent slices significant in retrieving the missed parts. Further, we design a refinement stage via developing the stacked U-Nets, which facilitates the generation of better motion-corrected images with superior maintaining of the image details and contrast. Generally, the proposed method increases the efficiency of the clinicians to correctly diagnose MR images by improving the overall clinical image quality. We conclude that if additional MRI scans are available, they can be used as image priors besides the self-assisted priors to further enhance the performance of motion artifacts correction.

## CRediT authorship contribution statement

**Mohammed A. Al-masni:** Conceptualization, Methodology, Software, Writing - original draft, Visualization. **Seul Lee:** Methodology. **Jaeuk Yi:** Formal analysis. **Sewook Kim:** Formal analysis. **Sung-Min Gho:** Data acquisition!, Funding acquisition. **Young Hun Choi:** Data acquisition!, Funding acquisition. **Dong-Hyun Kim:** Supervision, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work was supported in part by GE Healthcare research funds and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) ([NRF-2019R1A2C1090635](https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Correction-Self-Assisted-Priors/)).

## Data

The utilized MRI data in this paper is not able to be made openly available due to the privacy issues of clinical data.

## Code

The source code of our proposed stacked U-Nets with self-assisted priors is available at [\(13:italic \)https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Correction-Self-Assisted-Priors/\(13:italic\)](https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Correction-Self-Assisted-Priors/). We also make our MRI motion artifact simulation tool publicly available at [\(13:italic \)https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Simulation-Tool/\(13:italic\)](https://github.com/Yonsei-MILab/MRI-Motion-Artifact-Simulation-Tool/).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2022.119411](https://doi.org/10.1016/j.neuroimage.2022.119411).

## References

- Aggarwal, H.K., Mani, M.P., Jacob, M., 2019. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging* 38, 394–405.
- Andre, J.B., Bresnahan, B.W., Mossa-Basha, M., Hoff, M.N., Smith, C.P., Anzai, Y., Cohen, W.A., 2015. Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical mr examinations. *J. Am. Coll. Radiol.* 12, 689–695.
- Armanious, K., Tanwar, A., Abdulatif, S., Küstner, T., Gatidis, S., Yang, B., 2020. Unsupervised adversarial correction of rigid mr motion artifacts. In: Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa, USA. IEEE Xplore.
- Atkinson, D., Derek Lg, H., Peter, N.S., Paul, E.S., Steven, F.K., 1997. Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Trans. Med. Imaging* 16, 903–910.
- Chatterjee, S., Sciarra, A., Dünnwald, M., Oeltze-Jafra, S., Nürnberger, A., Speck, O., 2020. Retrospective motion correction of MR images using prior-assisted deep learning. In: Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Vancouver, Canada arXiv preprint arXiv:2011.14134.
- H. Chung, J. Kim, J.H. Yoon, J.M. Lee & J.C. Ye 2021. Simultaneous super-resolution and motion artifact removal in diffusion-weighted MRI using unsupervised deep learning. arXiv preprint arXiv:2105.00240.
- Cordero-Grande, L., Teixeira, R.P.A.G., Hughes, E.J., Hutter, J., Price, A.N., Hajnal, J.V., 2016. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Trans. Comput. Imaging* 2, 266–280.
- Duffy, B., Zhao, L., Sepehrband, F., Min, J., Wang, D.J.J., Shi, Y.G., Toga, A.W., Kim, H., Initia, A.D.N., 2021. Retrospective motion artifact correction of structural MRI images using deep learning improves the quality of cortical surface reconstructions. *Neuroimage* 230, 117756.
- Duffy, B.A., Zhang, W., Tang, H., Zhao, L., Law, M., Toga, A.W., Kim, H., 2018. Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion. In: Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL). Amsterdam, The Netherlands.
- Godenschweger, F., Kagebein, U., Stucht, D., Yarach, U., Sciarra, A., Yakupov, R., Lusebrink, F., Schulze, P., Speck, O., 2016. Motion correction in MRI of the brain. *Phys. Med. Biol.* 61, R32–R56.
- Haskell, M.W., Cauley, S.F., Bilgic, B., Hossbach, J., Splitthoff, D.N., Pfeuffer, J., Setsompop, K., Wald, L.L., 2019. Network accelerated motion estimation and reduction (NAMER): Convolutional neural network guided retrospective motion correction using a separable motion model. *Magn. Reson. Med.* 82, 1452–1461.
- Haskell, M.W., Cauley, S.F., Wald, L.L., 2018. TArgeted motion estimation and reduction (TAMER): data consistency based motion mitigation for MRI using a reduced model joint optimization. *IEEE Trans. Med. Imaging* 37, 1253–1265.
- Herbst, M., Maclarens, J., Weigel, M., Korvink, J., Hennig, J., Zaitsev, M., 2012. Prospective motion correction with continuous gradient updates in diffusion weighted imaging. *Magn. Reson. Med.* 67, 326–338.
- Hyun, C.M., Kim, H.P., Lee, S.M., Lee, S., Seo, J.K., 2018. Deep learning for undersampled MRI reconstruction. *Phys. Med. Biol.* 63, 135007.
- Jha, D., Riegler, M.A., Johansen, D., Halvorsen, P., Johansen, H.D., 2020. DoubleU-Net: a deep convolutional neural network for medical image segmentation. In: Proceedings of the IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). Rochester, Minnesota, USA. IEEE.
- Johnson, P.M., Drangova, M., 2019. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. *Magn. Reson. Med.* 82, 901–910.
- Kecskemeti, S.R., Alexander, A.L., 2020. Test-retest of automated segmentation with different motion correction strategies: a comparison of prospective versus retrospective methods. *Neuroimage* 209, 116494.
- Khalili, N., Turk, E., Zreik, M., Viergever, M.A., Benders, M.J.N.L., Işgum, I., Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A., 2019. Generative adversarial network for segmentation of motion affected neonatal brain MRI. *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, Cham.
- Ko, Y., Moon, S., Baek, J., Shim, H., 2021. Rigid and non-rigid motion artifact reduction in X-ray CT using attention module. *Med. Image Anal.* 67, 101883.
- Küstner, T., Armanious, K., Yang, J.H., Yang, B., Schick, F., Gatidis, S., 2019. Retrospective correction of motion-affected MR images using deep learning frameworks. *Magn. Reson. Med.* 82, 1527–1540.
- Lee, J., Kim, B., Park, H., 2021. MC(2)-net: motion correction network for multi-contrast brain MRI. *Magn. Reson. Med.* 86, 1077–1092.
- Lee, S., Jung, S., Jung, K.J., Kim, D.H., 2020. Deep learning in mr motion correction: a brief review and a new motion simulation tool (view2Dmotion). *Investig. Magn. Reson. Imaging* 24, 196–206.
- Liu, J., Kocak, M., Supanich, M., Deng, J., 2020. Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB). *Magn. Reson. Imaging* 71, 69–79.
- Loktyushin, A., Nickisch, H., Pohmann, R., Scholkopf, B., 2015. Blind multirigid retrospective motion correction of MR images. *Magn. Reson. Med.* 73, 1457–1468.
- Meding, K., Loktyushin, A., Hirsch, M., 2017. Automatic detection of motion artifacts in MR images using CNNs. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA. IEEE Xplore.

- Oh, G., Lee, J.E., Ye, J.C., Ye, J.C., 2021. Unpaired MR Motion artifact deep learning using outlier-rejecting bootstrap aggregation. *IEEE Trans. Med. Imaging* 40 (11), 3125–3139 1-1.
- Oksuz, I., Ruijsink, B., Puyol-Anton, E., Clough, J.R., Cruz, G., Bustin, A., Prieto, C., Botnar, R., Rueckert, D., Schnabel, J.A., King, A.P., 2019. Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning. *Med. Image Anal.* 55, 136–147.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich, Germany. Springer Cham.
- Ryu, K., Shin, N.Y., Kim, D.H., Nam, Y., 2019. Synthesizing T1 weighted MPRAGE image from multi echo GRE images via deep neural network. *Magn. Reson. Imaging* 64, 13–20.
- Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D., 2018. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* 37, 491–503.
- S. Shah, P. Ghosh, L.S. Davis & T. Goldstein 2018. Stacked U-Nets: a no-frills approach to natural image segmentation. arXiv preprint arXiv:1804.10343.
- Shaw, R., Sudre, C.H., Varsavsky, T., Ourselin, S., Cardoso, M.J., 2020. A k-space model of movement artefacts: application to segmentation augmentation and artefact removal. *IEEE Trans. Med. Imaging* 39, 2881–2892.
- Terpstra, M.L., Maspero, M., D'Agata, F., Stemkens, B., Intven, M.P.W., Lagendijk, J.J.W., Van Den Berg, C.A.T., Tijssen, R.H.N., 2020. Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy. *Phys. Med. Biol.* 65, 155015.
- Tisdall, M.D., Hess, A.T., Reuter, M., Meintjes, E.M., Fischl, B., Van Der Kouwe, A.J., 2012. Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn. Reson. Med.* 68, 389–399.
- Vaillant, G., Prieto, C., Kolbitsch, C., Penney, G., Schaeffter, T., 2014. Retrospective rigid motion correction in k-space for segmented radial MRI. *IEEE Trans. Med. Imaging* 33, 1–10.
- Wang, C., Liang, Y., Wu, Y., Zhao, S., Du, Y.P., 2020. Correction of out-of-FOV motion artifacts using convolutional neural network. *Magn. Reson. Imaging* 71, 93–102.
- Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. CBAM: convolutional block attention module. In: Proceedings of the 15th Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany.
- Zaitsev, M., Maclarens, J., Herbst, M., 2015. Motion artifacts in MRI: a complex problem with many partial solutions. *Magn. Reson. Imaging* 42, 887–901.
- Zhang, Q., Hann, E., Werys, K., Wu, C., Popescu, I., Lukaschuk, E., Barutcu, A., Ferreira, V.M., Piechnik, S.K., 2020. Deep learning with attention supervision for automated motion artefact detection in quality control of cardiac T1-mapping. *Artif. Intell. Med.* 110, 101955.
- Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE Xplore.