MAC: An Artifact Correction Framework for Brain MRI based on Deep Neural Networks

Alícia Oliveira, Beatriz Cepa, Cláudia Brito and António Sousa
HASLab, INESC TEC & University of Minho
Braga, Portugal
alicia.oliveira,beatriz.cepa,claudia.v.brito,antonio.l.sousa@inesctec.pt

Abstract

*The correction of artifacts in Magnetic Resonance Imaging (MRI) is crucial due to physiological phenomena and technical issues affecting diagnostic quality. Reverting from corrupted to artifact-free images is a complex task. Deep Learning (DL) models have been employed to preserve data characteristics and to identify and correct those artifacts. We propose **MAC**, a novel DL-based solution to correct artifacts in multi-contrast brain MRI scans. **MAC** offers two models: the simulation and the correction models. The simulation model introduces perturbations similar to those occurring in an exam while preserving the original image as ground truth; this is required as publicly available datasets rarely have motion-corrupted images. It allows the addition of three types of artifacts with different degrees of severity. The DL-based correction model adds a fourth contrast to state-of-the-art solutions while improving the overall performance of the models. **MAC** achieved the highest results in the FLAIR contrast, with a Structural Similarity Index Measure (SSIM) of 0.9803 and a Normalized Mutual Information (NMI) of 0.8030. Moreover, the model reduced training time by 63% compared to its predecessor. **MAC** model can correct large volumes of images faster and adapt to different levels of artifact severity than current state-of-the-art models, allowing for better diagnosis.*

1. Introduction

Magnetic Resonance Imaging (MRI) has become a crucial tool in medical diagnosis, providing relevant insights into the human body. Due to the inherent complexity associated with MRI procedures and their long examination times, susceptibility to movement increases, leading to image artifacts, particularly motion artifacts [14, 31], which result not only from patients' voluntary movements but also from involuntary physiological movements (*e.g.*, respiratory, cardiac and gastrointestinal peristaltic movements, vessel pulsations, blood and cerebrospinal fluid flow, and sudden involuntary movements such as swallowing) [31, 39]. Furthermore, other common artifacts in these exams are the chemical shift, the magnetic susceptibility, and noise artifacts [35]. Among all the artifacts, motion-related abnormalities are particularly problematic, with a prevalence ranging from 10% to 42% in brain examinations [36].

Various strategies can be used to avoid motion artifacts in MRI, such as sedating patients through moderate sedation or general anesthesia. However, these methods are not entirely effective, and motion artifacts are still present in around 12% of MRI images acquired under sedation and in around 0.7% of those acquired under general anesthesia [31, 39]. Given the limitations of these measures, correcting motion artifacts emerges as a viable solution to enhance medical imaging and improve diagnostic accuracy [11].

In this sense, Deep Learning (DL) shows great potential in correcting motion artifacts in MRI scans [13]. However, other types of artifacts are not equally addressed. DL models can learn to identify and correct artifacts, leading to clearer and more reliable images for diagnosis and artifact correction can be achieved through two approaches: prospective and retrospective. Prospective methods entail real-time correction during the MRI examination by modifying the data acquisition as it occurs [34]. Conversely, retrospective methods involve post-acquisition data enhancement after the MRI data acquisition is completed [34].

One key insight is that although real datasets present a high percentage of artifacts, publicly available ones are curated, lacking these and difficulting their usage for improving current correction models. Mimicking motion movements or any other artifact is paramount when dealing with longitudinal cases. For instance, if one wants to correct a patient's newly acquired MRI scan, a previous MRI scan without artifacts must be available to work as ground truth. Nonetheless, as these images are not publicly available, one must be able to emulate such artifacts.

Motivated by the critical need to enhance diagnostic accuracy in the presence of pervasive artifacts, this paper pro-

poses **MAC**, a novel end-to-end retrospective method for correcting artifacts in multi-contrast brain MRI scans based on DL. Unlike previous solutions [11, 26, 28, 29, 38], **MAC** is the first solution built to simulate various types of brain artifacts and correct them in four different contrasts.

In brief, **MAC** offers two models to enhance the pipeline of artifact correction. First, we offer a novel simulation model that can mimic motion, chemical shift, magnetic susceptibility, and noise artifacts. Further, the second model corrects artifacts, emulated and real, by re-implementing the DL-based MC2-Net [30] model and optimizing and adapting it to a fourth contrast. As such, our approach can be applied to four types of contrast, including T1-weighted (T1w), T1-weighted contrast-enhanced (T1CE), T2-weighted (T2w), and Fluid Attenuated Inversion Recovery (FLAIR), in brain MRI scans, and can be adapted to other imaging modalities and medical applications, expanding the model's potential in healthcare.

2. Related Work

Küstner et al. [29] proposed automated methods for detecting and quantifying motion artifacts in MRI images. Their initial work utilized a Multi-Column Neural Network (MCNN) to accurately detect motion artifacts in head and abdomen MRI images. They further developed a retrospective correction method [28] using DL techniques, including Generative Adversarial Networks (GANs) [21] and Variational Autoencoders (VAEs) [27], which resulted in improved image quality metrics. Nonetheless, this approach may alter anatomical features, potentially affecting medical diagnosis.

Johnson and Drangova [26] introduced a method for motion correction by predicting artifact-free images from motion-corrupted data using a Conditional GAN (CGAN) [32]. While this method reduced motion artifacts, it also resulted in some irregularities and degradation of image details. Armanious et al. [11] implemented a retrospective method for correcting motion artifacts using Medical GAN (MedGAN) [12], which demonstrated high correction power. However, the loss of image details remains a limitation. Lee et al. [30] proposed a method for correcting motion artifacts in multi-contrast brain MRI images, effectively reducing artifacts but showing limitations in correcting subtle motion artifacts. Finally, Tripathi et al. [38] modified the structure of CGAN [25] for motion artifact removal, achieving high accuracy but showing limitations in T1w contrast images.

MAC's correction model is a significant advancement when compared with these solutions, as it introduces the simulation of non-motion artifacts and a fourth contrast variation for correction, which extends the model's flexibility and usefulness. Concerning flexibility, **MAC** has the capacity to retain more information from larger quantities of data, improving its generalization for the four contrasts and the various types of artifacts, i.e., for each slice, the model leverages the information from each contrast and creates a more generalizable solution. This is also relevant for its usefulness, as this generalization can be exploited in real clinical scenarios, assisting in direct medical diagnosis.

3. MAC

An overview of the **MAC** model can be seen in Figure 1. The process includes data pre-processing, simulation of artifacts in the ground truth images, and correction of these artifacts using the correction model.

3.1. A day in the life of a corruption-free image

At the start of the workflow (Figure 1), input data is provided as images that are free from artifacts. These images serve as the baseline for the subsequent processing steps. The input data is pre-processed (❶) and fed into the simulation model (❷). During the simulation process, artifacts from different levels of artifact severity are introduced into the data (❸). The corruption-free and the corresponding corrupted images are divided based on contrasts and introduced in the correction model (❹) to be analyzed and corrected for any distortions or artifacts, thereby restoring them to their original state, ideally in a corruption-free condition (❺). Finally, the corrected images serve as the output of the correction model. These images are now free from the artifacts and are ready for further medical analysis (❻).

3.2. Data Pre-processing

The pre-processing of medical images in the Neuroimaging Informatics Technology Initiative (NIfTI) format [16], as illustrated in Figure 2, comprises several sequential steps aimed at enhancing data quality and usability for the developed DL model. This process starts with converting NIfTI volumes into the Portable Network Graphics (PNG) image format, which is compatible with the model. This is followed by a 90-degree rotation to optimize image orientation. Subsequently, brain slices were selected by analyzing images from multiple patients, where we determined that a cerebral brain area percentage greater than 35% is significant for evaluating distortions and identifying artifacts. Finally, the images are resized to 256×256 pixels to ensure uniformity and contribute to data normalization.

3.3. Simulation Model

The simulation model is indispensable for enhancing the correction model due to the lack of publicly available datasets with corrupted images and the difficulty of obtaining real images (*i.e.,* with artifacts) and the corresponding images without corruption (*i.e.,* ground truth images). Thus, the simulation model enables the creation of arti-

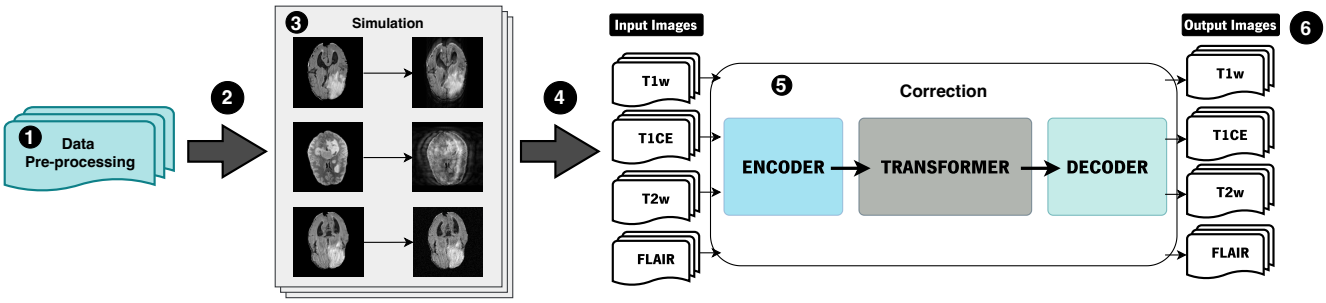

Figure 1. **MAC**'s pipeline. The pipeline is composed of the following several stages: data pre-processing, simulation, and correction.

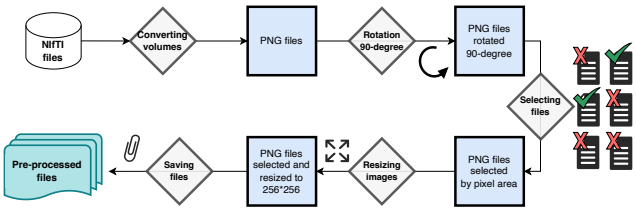

Figure 2. Pre-processing of medical images in NIfTI format involving converting to PNG, rotation, selecting slices, and resizing.

facts from images without any corruption to obtain perturbations similar to real artifacts (*i.e.,* motion, chemical shift, magnetic susceptibility, and noise [35]). Our simulation model was based on the motion simulation model of [17] and was adapted to encompass different types of artifacts. These include simulating fluctuations in resonance frequencies [24], reproducing fluctuations in magnetic susceptibility [15], and incorporating synthetic noise [19].

3.3.1 Simulating Movement

The process of simulating movement in MRI images, based on a Fourier domain motion simulation model [17], starts by transforming the original image into the frequency domain using the Fast Fourier Transform (FFT) signal processing technique [18]. Subsequently, a trajectory of random movements is generated by adjusting the parameters explained below. This involves introducing random rotations and translations into the original image to create a corrupted version mimicking movement effects. The corrupted image is then transformed into the spatial domain using the Inverse Fast Fourier Transform (IFFT) [18].

Parameters such as standard deviation of rotations and translations are configured for simulating artifacts, with different ranges of *corruption scheme*.

3.3.2 Simulating Fluctuations in Resonance Frequencies

This modification simulates variations in the resonance frequencies of different tissues or substances [24] in an MRI image. It is achieved by adding a chemical shift artifact to the image, altering the signal intensity frequency [23], as determined by a specific factor called the *shift factor*.

3.3.3 Reproducing Fluctuations in Magnetic Susceptibility

This transformation is designed to simulate variations in magnetic susceptibility [15] in MRI images, caused by " the magnetic field inhomogeneity" [5] produced by materials (*e.g.*, paramagnetic and diamagnetic materials). For this purpose, a mask is applied to the source image with a certain discrepancy, based on a specified *susceptibility factor*.

3.3.4 Incorporating Synthetic Noise

This alteration aims to introduce noise [19] into an image, controlled by a parameter termed *noise level*. To do this, a randomly generated noise matrix with the same shape as the original image is multiplied by the specified noise level and added to the original image.

The proposed simulation model considers two levels of artifact severity in medical images. At **level 1**, less severe artifacts are simulated, which include rotations and translations in the image (*i.e.,* motion artifacts). At **level 2**, in addition to rotations and translations, the model increases the modifications introduced in the MRI images, reproducing fluctuations in resonance frequencies and magnetic susceptibility and incorporating synthetic noise. The intention of the two levels is to create images with artifacts of different severity to improve the generalization of the correction model. The severity level of the simulation model depends on parameters such as the *shift factor*, the *susceptibility factor*, the *noise level*, and the *corruption scheme*. Figure 3 illustrates an example of each of these simulated artifacts.

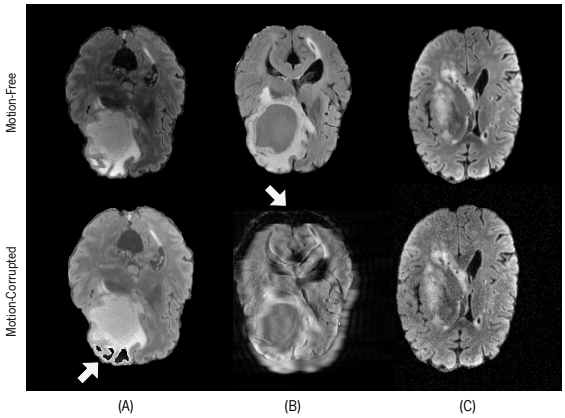

Figure 3. Examples of Simulation Model in the FLAIR contrast: (A) Simulating Fluctuations in Resonance Frequencies, (B) Reproducing Fluctuations in Magnetic Susceptibility, and (C) Incorporating Synthetic Noise.

3.4. Correction Model

The proposed model, represented in Figure 4, is based on the MC2-Net [30] architecture, which is composed of a combination of two tasks: image alignment (including registration network and NMI maximization algorithm) and a motion correction network. Our model focuses exclusively on the correction model and aims to correct motion artifacts, chemical shifts, magnetic susceptibility, and noise attenuation. We have extended the capabilities of the original correction network architecture to support four contrast images, adjusting the input and output layers accordingly. Additionally, we have improved the model's capacity by increasing the number of initial filters to 32, allowing for more feature extraction and better correction.

The DL-based model is composed of three main blocks: the *Encoder*, the *Transformer*, and the *Decoder*. The Encoder transforms the input images into unique feature vectors for each contrast image, which uses 2D convolutions (*Conv2D*). These vectors are concatenated to allow the sharing of information between contrasts. The concatenated feature vectors are then passed through the Transformer, which comprises nine residual blocks, each containing two *Conv2D* interconnected by skip connections. The Decoder uses transposed 2D convolution (*TConv2D*) layers with Rectified Linear Unit (*ReLU*) activation to reconstruct the corrected features in the corresponding contrast image.

To train the model, we use a combination of loss functions that are customized for multi-contrast MRI correction. Specifically, we used SSIM [8] to ensure data consistency and VGG [10] perceptual loss to preserve perceptual details. We have chosen Adam [3] as the optimizer with a learning rate of 0.001. Our model has been implemented using TensorFlow 2.0 [9] and the Keras [4] library.

3.4.1 Image Quality Metrics

The correction network's performance can be evaluated based on the quality of the resulting output image using quantitative metrics [37]. With this, the proposed solution was assessed using three metrics: SSIM [8], NMI [2], and Normalized Root Mean Squared Error (NRMSE) [6].

SSIM [8] measures the structural similarity between two images by analyzing the loss of structural information. SSIM is defined as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)\cdot(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)\cdot(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

where "μ_x is the average value of x, μ_y is the average value of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, σ_{xy} is the covariance of x and y, $c_1 = (0.01L)^2$ and $c_2 = (0.03L)^2$ are two variable to balance division with weak denominator, L is the dynamic range of pixel values" [38].

NMI [2] measures the agreement between the cluster labels assigned by two different clustering algorithms or methods by considering both the homogeneity within clusters and the separation between them [2]. NMI is expressed as

$$NMI(A,B) = \frac{I(A;B)}{\sqrt{H(A)\cdot H(B)}} \quad (2)$$

where A and B are the sets of clustering labels from two different clustering methods, $I(A;B)$ is the mutual information between the clusterings, and $H(A)$ and $H(B)$ are the entropies of the individual clusterings [2].

The NRMSE [6] is a metric commonly used to assess the accuracy of a predictive model, regression model, or forecast. It normalizes the RMSE (Root Mean Squared Error) [7] by the range of the observed values [6]. NRMSE referred to as

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}}{x_{\max} - x_{min}} \quad (3)$$

where N is the number of observations, x_i represents the actual observed value, $\hat{x}_i$ represents the predicted value, and $x_{\max}$ and $x_{\min}$ denote the maximum and minimum observed values, respectively [6].

4. Evaluation

We evaluate the effectiveness of **MAC** artifact correction in terms of different levels of severity with different datasets. Our key insights are as follows:

- **MAC** enhances state-of-the-art results when compared to the baseline model MC2-Net.

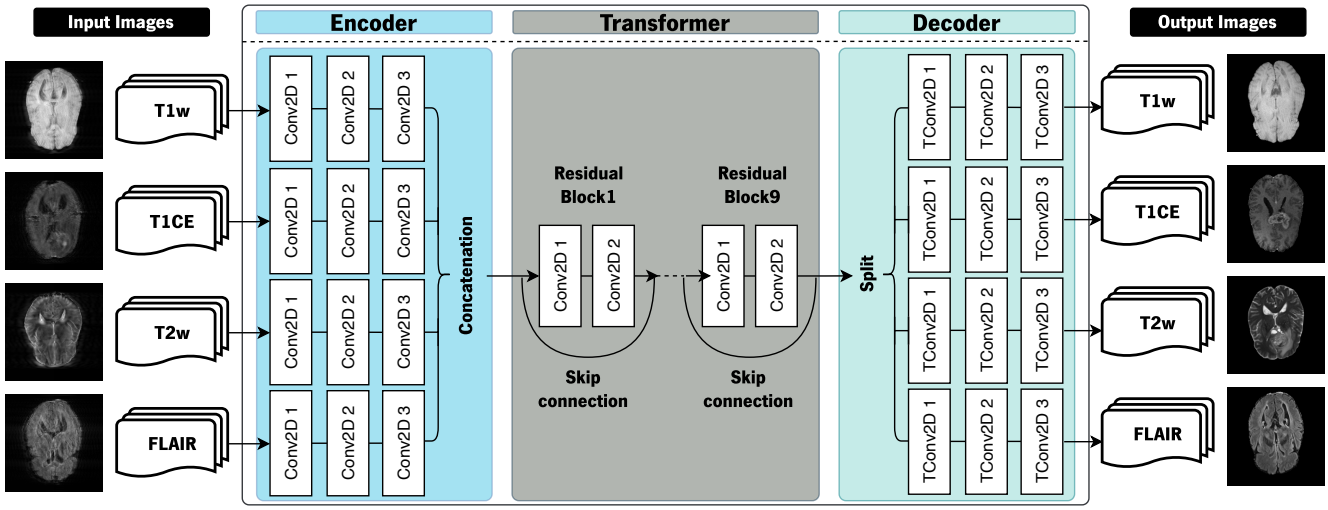

Figure 4. **MAC**'s Correction Model Architecture. The DL-based model is composed of three main blocks: the *Encoder*, the *Transformer*, and the *Decoder*. Each of these blocks is composed of interconnected convolutional layers. The convolutional layer output is concatenated in the first block (Encoder). The Transformer comprises 9 residual block layers, and its output is split to serve as input for the Decoder.

- **MAC** allows the efficient addition of a fourth contrast for correction.
- **MAC** provides different levels of severity regarding the simulation model, being able to attain results up to **0.9964** in SSIM for the highest severity (*i.e.*, heavily corrupted images).

4.1. Experimental Setup

Testbed Setup. The tests were performed in two nodes, one with an NVIDIA GeForce RTX 2080 Ti (Node 1) and a second one with an NVIDIA GeForce RTX 3090 (Node 2). The optimal filters and epochs were determined by analyzing the model's performance with varied hyperparameters. Namely, we performed tests with 8, 16, 32, and 64 filters. Moreover, we trained the model with different numbers of epochs to understand its evolution (100, 250, 500, 750, 1000, 1200). This evaluation was carefully performed to guarantee that the model did not overfit the data (which did occur when running for 1200 epochs).

Workloads and metrics. The model was trained using data from the Brain Tumor Segmentation Challenge 2021 (BraTS 2021) [1]. This dataset consists of multi-parametric MRI (mpMRI) scans available in the Neuroimaging Informatics Technology Initiative (NIfTI) format (.nii.gz) that comprise native T1w, post-contrast T1w, T2w, and FLAIR volumes [1].

Four different subsets were created from the BraTS dataset: A1, A2, B1, and C1. Dataset A1 contains 48 subjects and 786 slices (262 slices from each contrast) selected from three contrasts (T1w, T2w, and FLAIR). Of these, 193 slices were used for training, 34 for validation, and 35 for testing. Dataset A2 includes 48 subjects and 1048 slices (262 slices from each contrast) chosen from four contrasts (T1w, T1CE, T2w, and FLAIR). The distribution of slices for training, validation, and testing follows the one of dataset A1. Dataset A1 was simulated with artifact severity levels 1 and 2, while dataset A2 was simulated with level 2. Dataset B1 consists of 3600 slices from 248 subjects, where each contrast has 900 slices for a total of four contrasts (T1w, T1CE, T2w, and FLAIR). The slices were split into three sets: 630 for training, 135 for validation, and 135 for testing. Dataset B1 was simulated with artifact severity level 2. The last subset of the BraTS, dataset C1, consists of 140 slices (35 slices for each contrast) from 4 subjects. These slices contain real movement and were used to conduct real tests with artifacts (*i.e.,* for model inference), evaluating the model's capability to correct real artifacts.

Table 1 summarizes our datasets. Node 1 was used to train the model with datasets A1 and A2, and Node 2 was used with dataset B1. The model's performance is validated by resorting to SSIM, NMI, and NRMSE metrics.

Table 1. Datasets used in our evaluation. The number of slices used for each phase (i.e., train, validation, and test) is provided per contrast.

	Subjects	Slices	Contrasts	Train	Validation	Test
A1	48	786	3	193	34	35
A2	48	1048	4	193	34	35
B1	248	3600	4	630	135	135
C1	4	140	4	–	–	35

4.2. Results

The following section is split by the simulation model's severity and the real images test.

4.2.1 Low Severity - Level 1

During the initial testing phase, the **MAC** network was trained using dataset A1 and the simulation model's severity level 1. The network's performance was evaluated by altering the parameters, such as Batch Validation (BV) and the Number of Filters (NF), aiming to determine the optimal parameters for better results. All the tests for severity level were conducted with 500 epochs.

The results, summarized in Table 2, highlight significant variations in model performance based on different parameter configurations. We verified that when BV was not implemented and the NF was fixed at 16, our model's SSIM values varied from 0.9473 to 0.9608 in different contrasts. Similarly, the NMI scores ranged from 0.7616 to 0.7704, and the corresponding NRMSE values from 0.1687 to 0.1938.

Table 2. Comparison of different model parameters and **MAC** model results for SSIM, NMI, and NRMSE metrics, in dataset A1 and the simulation model's severity level 1. BV is "Yes" or "No", indicating its presence or absence. NF is 16 or 32, indicating the number of filters used by the model.

Parameters			Results			Training Time
BV	**NF**	**Contrast**	**SSIM**	**NMI**	**NRMSE**	**(HH:MM:SS)**
No	16	T1w	0.9591	0.7616	0.1730	21:58:53
		T2w	0.9608	0.7704	0.1687	
		FLAIR	0.9473	0.7638	0.1938	
Yes	16	T1w	0.9682	0.7786	0.1527	22:06:22
		T2w	0.9705	0.7811	0.1492	
		FLAIR	0.9593	0.7701	0.1639	
No	32	T1w	0.9655	0.7756	0.1457	21:58:03
		T2w	0.9654	0.7685	0.1480	
		FLAIR	0.9590	0.7818	0.1688	
Yes	32	T1w	0.9745	0.7811	0.1497	22:07:03
		T2w	0.9792	0.7821	0.1408	
		FLAIR	0.9803	0.8030	0.1580	

However, we observed significant improvement in our model's performance in all contrasts when we introduced BV. This is shown by higher SSIM values ranging from 0.9593 to 0.9705, indicating better structural similarity between predicted and ground truth images. Additionally, NMI values increased from 0.7786 to 0.7811, indicating enhanced capture of mutual information. NRMSE values decreased to a range of 0.1492 to 0.1639, signifying reduced error in model predictions.

Furthermore, increasing the NF from 16 to 32 without BV increased SSIM values in all the contrasts. A substantial boost in model performance is observed when combining an NF of 32 with BV, especially in SSIM values in different contrasts, particularly in FLAIR, which increased from 0.9590 to 0.9803. This suggests that a higher NF enables the model to extract more features and representations from the input images. At the same time, NMI values experienced a significant increase, from 0.7818 to 0.8030 in FLAIR contrast, and the NRMSE values decreased from 0.1688 to 0.1580.

When using the BV technique with **MAC**, the training time may increase slightly because the validation process is performed for each batch. However, the increase in training time is not significant (0.6% in the case of dataset A1), and it requires about 22 hours to complete 500 epochs.

In Table 3, we present a comparison between MC2-Net [1], which includes the registration and the artifact correction networks, and **MAC** models, in three different MRI contrasts: T1w, T2w, and FLAIR. All the tests were realized with 500 epochs.

Table 3. Comparison of the best SSIM and NMI metrics results in MC2-Net and **MAC** models for three MRI contrasts.

Model	Contrast	Metrics		Training Time
		SSIM	**NMI**	**(Hours)**
MC2-Net	T1w	0.9891	0.8344	37+59*
	T2w	0.9787	0.7860	
	FLAIR	0.9726	0.7698	
MAC	T1w	0.9745	0.7811	22
	T2w	0.9792	0.7821	
	FLAIR	0.9803	0.8030	

* Training time for registration and motion correction networks

When compared to MC2-Net, **MAC** outperforms in T2w and FLAIR contrasts, with an increase of SSIM value from 0.9787 to 0.9792 and from 0.9726 to 0.9803, respectively, indicating superior structural similarity and better alignment between predicted and ground truth images in these two contrasts.

The training time varied across all experiments, depending on the model architecture and parameter configurations. MC2-Net had a substantially longer training time, averaging around 37 hours for the registration network (1000 epochs) and 59 hours for the correction network (500 epochs) [30]. In contrast, the **MAC** model had significantly shorter training times, completing 500 epochs in approximately 22 hours. This reduction in training time with the **MAC** model highlights its efficiency and computational feasibility for practical deployment in real-world scenarios.

Figure 5 shows examples of the output of our correction model after the input of corrupted images with severity level 1. As can be observed, the SSIM values of motion-corrupted images are significantly lower than the SSIM values of motion-corrected ones.

[1] We compare directly with MC2-Net only, since this model has been previously compared with other state-of-the-art models, outperforming them [20, 22, 33].

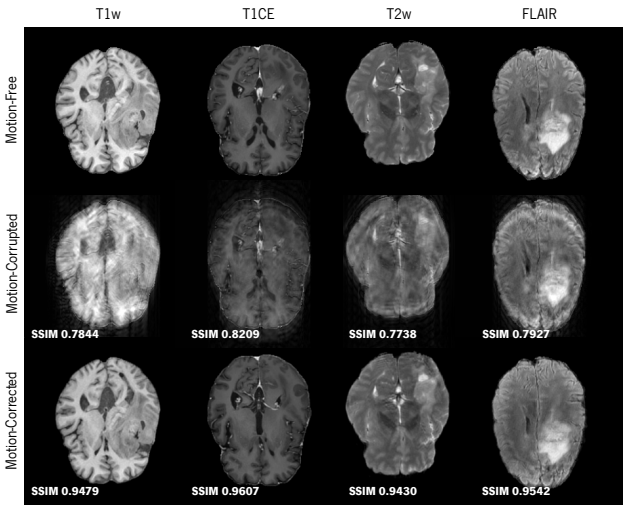

Figure 5. Example images of correction with a severity level 1 of simulation.

4.2.2 High Severity - Level 2

Table 4 compares the different model parameters and their respective performance metrics when training **MAC** using the A1 dataset with the severity level 2. The experiments were performed over 500 epochs.

Table 4. Comparison of different model parameters and **MAC** model results for SSIM, NMI, and NRMSE metrics, in dataset A1 and the simulation model's severity level 2.

Parameters			Results			Training Time
BV	**NF**	**Contrast**	**SSIM**	**NMI**	**NRMSE**	**(HH:MM:SS)**
No	16	T1w	0.9218	0.6648	0.2515	21:51:36
		T2w	0.9414	0.7376	0.2238	
		FLAIR	0.9610	0.7661	0.1393	
Yes	16	T1w	0.9243	0.6777	0.2309	22:05:07
		T2w	0.9428	0.7411	0.2105	
		FLAIR	0.9624	0.7718	0.1368	
No	32	T1w	0.9387	0.7364	0.2422	21:51:10
		T2w	0.9489	0.7506	0.1925	
		FLAIR	0.9635	0.7747	0.1323	
Yes	32	T1w	0.9425	0.7498	0.2308	22:06:42
		T2w	0.9543	0.7667	0.1536	
		FLAIR	0.9637	0.7779	0.1292	

The results showed that when BV was omitted and NF was set to 16, the model's SSIM values ranged from 0.9218 to 0.9610, NMI scores between 0.6648 to 0.7661, and NRMSE values from 0.1393 to 0.2515. However, the introduction of BV significantly improved the model's performance in all contrasts, yielding higher SSIM values (ranging from 0.9243 to 0.9624), increased NMI scores (ranging from 0.6777 to 0.7718), and reduced NRMSE values (up to 8% of reduction).

Furthermore, the increase in NF from 16 to 32 enhanced the model's performance, particularly when used alongside BV. For instance, in the T2w contrast, when NF was increased from 16 to 32 and BV was applied, the SSIM value rose from 0.9428 to 0.9543, representing a clear improvement in the structural similarity between predicted and ground truth images. Analogously, the NMI score increased from 0.7411 to 0.7667 and the NRMSE value decreased from 0.2105 to 0.1536. This illustrates how the changes in NF and BV allowed the model to capture more detailed information from the input images, resulting in clearer and more accurate predictions. It is worth noting that the model requires approximately 22 hours to complete 500 epochs.

Table 5 shows a comparison of the results obtained by using different model parameters with the **MAC** model in four image contrasts, for the A2 dataset at severity level 2 of the simulation model. The tests were conducted during 500 epochs.

Table 5. Comparison of different model parameters and **MAC** model results for SSIM, NMI, and NRMSE metrics, in dataset A2 and the simulation model's severity level 2.

Parameters			Results			Training Time
BV	**NF**	**Contrast**	**SSIM**	**NMI**	**NRMSE**	**(D-HH:MM:SS)**
No	16	T1w	0.8665	0.6349	0.3860	1-04:03:09
		T1CE	0.9367	0.7225	0.2330	
		T2w	0.8606	0.6365	0.3453	
		FLAIR	0.9604	0.7561	0.1408	
Yes	16	T1w	0.9220	0.6761	0.2483	1-04:27:09
		T1CE	0.9383	0.7209	0.2301	
		T2w	0.9421	0.7353	0.2342	
		FLAIR	0.9613	0.7643	0.1433	
No	32	T1w	0.9533	0.7690	0.2334	1-03:52:22
		T1CE	0.9344	0.7263	0.2691	
		T2w	0.9377	0.7282	0.2655	
		FLAIR	0.9545	0.7692	0.2273	
Yes	32	T1w	0.9542	0.7679	0.1517	1-04:10:43
		T1CE	0.9464	0.7329	0.2255	
		T2w	0.9482	0.7374	0.2321	
		FLAIR	0.9664	0.7784	0.1350	

By analyzing the results in Table 5, it is perceptible that when BV is not applied, and the NF is set to 16, the SSIM scores range between 0.8606 and 0.9604 for all four contrasts. With BV applied and NF set to 16, the SSIM increased to up to 0.9613. On the other hand, by increasing the NF to 32 without applying BV, the results are superior in some metrics. For example, for the FLAIR contrast, when shifting BV to the "Yes" with an NF of 32, there is a significant reduction in NRMSE to 0.1350.

In addition, the presence of BV can positively affect the metrics in other contrasts. For instance, for the T1CE contrast, activating BV with an NF of 32 increases SSIM from 0.9344 to 0.9464, which shows an improvement in the structural similarity of the images processed by the model. Regarding the training time for 500 epochs, the process was completed in 28 hours.

By tripling the number of training images, the results presented in Table 6 show that by having access to a more extensive and diverse dataset for learning, the model performs better. This increased dataset potentially enhances the model's ability to generalize and produce more accurate predictions. Due to the increased dataset size, the training endured for 1000 epochs.

Table 6. Comparison of different model parameters and **MAC** model results for SSIM, NMI, and NRMSE metrics, in dataset B1 and the simulation model's severity level 2.

Parameters			Results			Training Time
BV	**NF**	**Contrast**	**SSIM**	**NMI**	**NRMSE**	**(D-HH:MM:SS)**
No	16	T1w	0.9383	0.7387	0.1571	3-11:59:07
		T1CE	0.9451	0.7664	0.1958	
		T2w	0.9402	0.7564	0.2740	
		FLAIR	0.9313	0.7318	0.2233	
Yes	16	T1w	0.9395	0.7441	0.1577	3-12:49:35
		T1CE	0.9453	0.7666	0.1838	
		T2w	0.9452	0.7603	0.2228	
		FLAIR	0.9530	0.7876	0.1759	
No	32	T1w	0.9399	0.7402	0.1588	3-12:11:21
		T1CE	0.9482	0.7596	0.1968	
		T2w	0.9488	0.7699	0.1970	
		FLAIR	0.9542	0.7808	0.1808	
Yes	32	T1w	0.9479	0.7672	0.1290	3-12:55:43
		T1CE	0.9607	0.7929	0.1695	
		T2w	0.9530	0.7708	0.1452	
		FLAIR	0.9647	0.7966	0.1506	

The results indicate that the model consistently performs better when BV is used in all contrasts. For instance, in FLAIR images with 16 filters, SSIM increased from 0.9313 to 0.9530 with BV, while NMI rose from 0.7318 to 0.7876.

Moreover, when examining the influence of NF, we see that increasing NF generally leads to better results. This tendency is evident in T2w images, where SSIM increased from 0.9402 to 0.9488 when going from 16 to 32 filters without BV. Conversely, the effect of NF is more pronounced when BV is used. This is exemplified in T1CE images, where SSIM improved from 0.9453 to 0.9607 when BV was used with 32 filters. The model was trained for approximately 94 hours.

4.2.3 Real Images

First, the **MAC** correction model was trained with simulated images due to the difficulty of obtaining motion-free and motion-corrupted images of the same slice. Then, the model was inferred with data from four subjects (dataset C1) containing real movement. The correction results for the T1CE contrast are presented in Figure 6. The figure shows that, even when applied to real images, our correction model successfully corrected and removed movement from the images. This validates the method's effectiveness and applicability in real-world scenarios, providing a reliable solution for correcting images affected by motion.

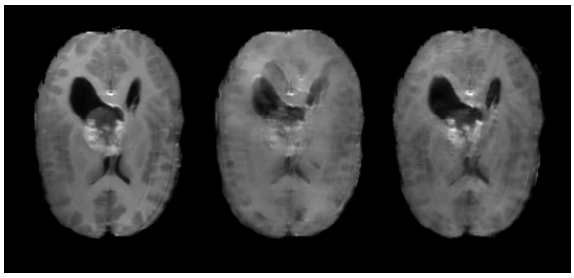

Figure 6. Example images of motion correction in the T1CE contrast of dataset C1, containing real motion-free, motion-corrupted and motion-corrected images.

4.3. Limitations

The scarcity of real images that contain artifacts and the corresponding ground truth images is bridged by **MAC**'s simulation model. As such, by following a retrospective approach, **MAC** requires both the source image and the corrupted image to apply the correction process accurately. With this, the inference of real images in such a model depends directly on the availability of corruption-free slices from the same subject. This is a general issue that should be addressed by the scientific community by gathering non-curated and artifact-corrupted scans and making them publicly available for research use.

Additionally, our correction model was able to attain an average SSIM for all contrasts over 0.93. In cases where the artifacts were subtle, the correction using 32 filters and BV showed noteworthy improvements when compared with the use of 16 filters and no BV. However, we believe that these results could be refined by resorting to distributed settings, as we defer this to future work.

5. Conclusion

We propose **MAC**, an end-to-end pipeline for artifact correction in MRI images. By offering two models, **MAC** is able to efficiently mimic common artifacts in previously curated data and further correct them. **MAC**'s capability to handle artifacts in various contrasts, combined with its efficiency in correcting different levels of distortion, particularly severe artifacts, makes it an advantageous tool in medical imaging analysis. Although there are still challenges in correcting subtle artifacts, the model's performance exhibits improvements over existing methods, indicating a promising direction for future research and development in the field of artifact correction in MRI. With continued advancements, **MAC** has the potential to assist clinicians in obtaining clearer and more trustworthy diagnoses.

References

[1] Brats 2021 task 1 dataset. 5

[2] Clustering metrics and cluster validity – building skills for data science. 4

[3] Complete guide to the adam optimization algorithm — built in. 4

[4] Keras: Deep learning for humans. 4

[5] Magnetic susceptibility artifact mri. 3

[6] Nrmse - statistics how to. 4

[7] Rmse: Root mean square error - statistics how to. 4

[8] Ssim: Structural similarity index — imatest. 4

[9] Tensorflow. 4

[10] Vgg-16 convolutional neural network - matlab vgg16. 4

[11] Karim Armanious, Sergios Gatidis, Konstantin Nikolaou, Bin Yang, and Thomas Kustner. Retrospective correction of rigid and non-rigid mr motion artifacts using gans. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1550–1554. IEEE, 2019. 1, 2

[12] Karim Armanious, Chenming Jiang, Marc Fischer, Thomas Küstner, Tobias Hepp, Konstantin Nikolaou, Sergios Gatidis, and Bin Yang. Medgan: Medical image translation using gans. *Computerized medical imaging and graphics*, 79:101684, 2020. 2

[13] Tomas Budrys, Vincentas Veikutis, Saulius Lukosevicius, Rymante Gleizniene, Egle Monastyreckiene, and Ilona Kulakiene. Artifacts in magnetic resonance imaging: how it can really affect diagnostic image quality and confuse clinical diagnosis? *Journal of Vibroengineering*, 20(2):1202–1213, 2018. 1

[14] Jerrold T Bushberg and John M Boone. *The essential physics of medical imaging*. Lippincott Williams and Wilkins, 2011. 1

[15] Stewart C Bushong and Geoffrey Clarke. *Magnetic resonance imaging: physical and biological principles*. Elsevier Health Sciences, 2003. 3

[16] Data Format Working Group (DFWG). NIfTI: Neuroimaging informatics technology initiative. 2

[17] Ben A Duffy, Lu Zhao, Farshid Sepehrband, Joyce Min, Danny JJ Wang, Yonggang Shi, Arthur W Toga, Hosung Kim, Alzheimer's Disease Neuroimaging Initiative, et al. Retrospective motion artifact correction of structural mri images using deep learning improves the quality of cortical surface reconstructions. *NeuroImage*, 230:117756, 2021. 3

[18] R Durga and Bhavani D Sudhakar. Design and implementation of inverse fast fourier transform for ofdm. *International Journal of Science and Engineering Applications*, 2:2319–7560, 2013. 3

[19] Roger Dzwonczyk, Jeffrey T Fujii, Orlando Simonetti, Ricardo Nieves-Ramos, and Sergio D Bergese. Electrical noise in the intraoperative magnetic resonance imaging setting. *Anesthesia & Analgesia*, 108(1):181–186, 2009. 3

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Advances in neural information processing systems. *Curran Associates, Inc*, 27:2672–2680, 2014. 6

[21] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[23] Maureen N Hood, Vincent B Ho, James G Smirniotopoulos, and Jerzy Szumowski. Chemical shift: the artifact and clinical tool revisited. *Radiographics*, 19(2):357–371, 1999. 3

[24] Pablo Irarrazabal and Dwight G Nishimura. Fast three dimensional magnetic resonance imaging. *Magnetic resonance in medicine*, 33(5):656–662, 1995. 3

[25] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 2

[26] Patricia M Johnson and Maria Drangova. Conditional generative adversarial network for 3d rigid-body motion correction in mri. *Magnetic resonance in medicine*, 82(3):901–910, 2019. 2

[27] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[28] Thomas Küstner, Karim Armanious, Jiahuan Yang, Bin Yang, Fritz Schick, and Sergios Gatidis. Retrospective correction of motion-affected mr images using deep learning frameworks. *Magnetic resonance in medicine*, 82(4):1527–1540, 2019. 2

[29] Thomas Küstner, Annika Liebgott, Lukas Mauch, Petros Martirosian, Fabian Bamberg, Konstantin Nikolaou, Bin Yang, Fritz Schick, and Sergios Gatidis. Automated reference-free detection of motion artifacts in magnetic resonance images. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 31:243–256, 2018. 2

[30] Jongyeon Lee, Byungjai Kim, and HyunWook Park. Mc2-net: motion correction network for multi-contrast brain mri. *Magnetic Resonance in Medicine*, 86(2):1077–1092, 2021. 2, 4, 6

[31] Seul Lee, Soozy Jung, Kyu-Jin Jung, and Dong-Hyun Kim. Deep learning in mr motion correction: a brief review and a new motion simulation tool (view2dmotion). *Investigative Magnetic Resonance Imaging*, 24(4):196–206, 2020. 1

[32] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 2

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 6

[34] Jakob M. Slipsager, Stefan L. Glimberg, Liselotte Højgaard, Rasmus R. Paulsen, Paul Wighton, M. Dylan Tisdall, Camilo Jaimes, Borjan A. Gagoski, P. Ellen Grant, André van der Kouwe, Oline V. Olesen, and Robert Frost. Comparison of prospective and retrospective motion correction in

3d-encoded neuroanatomical mri. *Magnetic Resonance in Medicine*, 87:629–645, 2 2022. 1

[35] K Somasundaram. Analysis of imaging artifacts in mr brain images (hal 136-140), department of computer science and applications gandhigram rural institute deemend university. *Tamil Nadu. India*, 2012. 1, 3

[36] K Sommer, A Saalbach, T Brosch, C Hall, NM Cross, and JB Andre. Correction of motion artifacts using a multiscale fully convolutional neural network. *American Journal of Neuroradiology*, 41(3):416–423, 2020. 1

[37] Veronika Spieker, Hannah Eichhorn, Kerstin Hammernik, Daniel Rueckert, Christine Preibisch, Dimitrios C Karampinos, and Julia A Schnabel. Deep learning for retrospective motion correction in mri: A comprehensive review. *arXiv preprint arXiv:2305.06739*, 2023. 4

[38] Vijay R Tripathi, Manish N Tibdewal, and Ravi Mishra. Denoising of motion artifacted mri scans using conditional generative adversarial network. *Multimedia Tools and Applications*, pages 1–19, 2023. 2, 4

[39] Maxim Zaitsev, Julian Maclaren, and Michael Herbst. Motion artifacts in mri: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4):887–901, 2015. 1