Scientific Research Publishing

Journal of Computer and Communications, 2024, 12, 210-225
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Fetal MRI Artifacts: Semi-Supervised Generative Adversarial Neural Network for Motion Artifacts Reducing in Fetal Magnetic Resonance Images

Ítalo Messias Félix Santos[1], Gilson Antonio Giraldi[1], Heron Werner Junior[2], Bruno Richard Schulze[1]

[1]National Laboratory for Scientific Computing, Rio de Janeiro, Brazil
[2]Biodesign Laboratory Dasa/PUC-Rio, Rio de Janeiro, Brazil
Email: gilson@lncc.br

How to cite this paper: Santos, Í.M.F., Giraldi, G.A., Werner Junior, H. and Schulze, B.R. (2024) Fetal MRI Artifacts: Semi-Supervised Generative Adversarial Neural Network for Motion Artifacts Reducing in Fetal Magnetic Resonance Images. *Journal of Computer and Communications*, **12**, 210-225.
https://doi.org/10.4236/jcc.2024.126013

Received: January 18, 2024
Accepted: June 25, 2024
Published: June 28, 2024

Abstract

This study addresses challenges in fetal magnetic resonance imaging (MRI) related to motion artifacts, maternal respiration, and hardware limitations. To enhance MRI quality, we employ deep learning techniques, specifically utilizing Cycle GAN. Synthetic pairs of images, simulating artifacts in fetal MRI, are generated to train the model. Our primary contribution is the use of Cycle GAN for fetal MRI restoration, augmented by artificially corrupted data. We compare three approaches (supervised Cycle GAN, Pix2Pix, and Mobile Unet) for artifact removal. Experimental results demonstrate that the proposed supervised Cycle GAN effectively removes artifacts while preserving image details, as validated through Structural Similarity Index Measure (SSIM) and normalized Mean Absolute Error (MAE). The method proves comparable to alternatives but avoids the generation of spurious regions, which is crucial for medical accuracy.

Keywords

Fetal MRI, Artifacts Removal, Deep Learning, Image Processing, Generative Adversarial Networks

1. Introduction

The field of medicine has benefited significantly from technological advancements in image processing, particularly in the area of diagnosing fetal anomalies. In this area, although ultrasound remains the primary modality, in some cases, specific diagnoses during pregnancy require the application of magnetic resonance imaging (MRI), which is considered the forefront technology for imaging

the fetus due to its capability to visualize complex congenital anomalies [1].

However, technical barriers related to the speed of acquisition, spatial resolution, and signal-to-noise ratio (SNR) often impose limitations on the diagnostic utility of MRI. Improving the imaging quality through hardware resources during acquisition is challenging due to correlations between the scan parameters [2] [3]. In this context, we must also consider that fetal motion and maternal respiration are inherent factors in the examination. These movements can cause artifacts in fetal imaging, such as blurring, ghosting, contrast changes, and signal voids [1].

The mentioned issues about technical limitations and acquisition protocol in fetal imaging have led to the exploration of traditional and deep learning approaches to improve spatial resolution and SNR [4]-[9].

This paper focuses on the latter since deep neural network methods compose the state-of-the-art for noise reduction in this application [8]. Despite the advances of deep learning approaches in medical imaging [10]-[12], many algorithms require labeled or paired data, which is a fundamental issue in the case of fetal MRI due to practical limitations associated with the exam, such as fetal movements and maternal respiration, as previously mentioned. As a consequence, the mathematical formulation of matching pairs in this case is unclear.

Previous works [4] [7] [13] have usedCycle generative adversarial network (Cycle GAN) with U-net [14] as the generator network and PatchGAN [15] as the discriminator to remove artifacts from medical images. These works inspired us to choose Cycle GAN for the task of artifact removal in fetal MRI.

Our methodology employs semi-supervised and supervised Cycle GAN models, along with data augmentation techniques, to restore fetal MRI images. We begin by exploring two different types of data augmentation. The first involves standard geometric transformations and image processing steps. The second strategy for data augmentation involves simulating the fetal motion using the method proposed by Duffy *et al.* [16]. This technique facilitates the creation of matching images using synthetic data, allowing us to train a semi-supervised Cycle GAN model using both paired and unpaired fetal images. Specifically, we use the original training dataset to build the unsupervised loss term and the synthetic pairs for the supervised term in the loss function, generating the Cycle GAN V1 neural network. Moreover, a second model, named Cycle GAN V2, is produced by applying only the synthetic pairs for both supervised and unsupervised terms in the Cycle GAN loss function. The outstanding results obtained by Cycle GAN V2 indicate that the artificially generated data can accurately characterize the involved artifacts.

In this study, we compare the effectiveness of the best Cycle GAN approach with other deep learning architectures for image restoration tasks [17] [18]. We address this issue by visual analysis of the results, Structural Similarity Index Measure (SSIM), and the normalized Mean Absolute Error (MAE), which are computed over the synthetic dataset and the Fréchet Inception Score (FID) [19]

and the Jensen-Shannon Divergence (JSD) [20] computed overs the generated images in comparison to the real images without artifacts.

The main contributions of our study are: 1) the application of Cycle GAN approaches to fetal MRI motion noise removal, 2) the use of data augmentation in the fetal MRI context through the technique proposed by Duffy *et al.* [16], that simulates motion artifacts in magnetic resonance, 3) a comparison of the supervised Cycle GAN [21], Pix2Pix [18] and Mobile Unet [17] approaches for removing fetal MRI artifacts produced by the fetus movement during the image acquisition.

This paper is organized as follows. The related works are described in Section 1.1. Section 2 provides a brief description of the dataset and the protocols that are used to obtain fetal MRI besides data augmentation approaches. Section 3 describes the neural network used and the terms of its loss function. In Section 4, the results of each model are presented and analyzed. Finally, in Section 5, we give our final remarks and future works.

Related Works

In this section, we review important works that are relevant to our proposal for motion noise removal. Considering the classification for motion in MRI images given in [22], the motion in fetal MRI can be categorized as rigid, sporadic/random, inter-scan, and inter-image.

Our methodology is developed in the context of deep learning approaches for MRI processing. The use of deep learning frameworks based on convolutional neural network (CNN) approaches has shown promise for motion correction and other tasks in MRI, as noted by Chen *et al.* [23]. Methodologies based on generative adversarial networks (GANs) using CNNs for the generator architecture have been explored, including pix2pix [18], Cycle GAN [21] and its variants, which have produced good results in medical image scenarios even when paired data is not available [4] [7] [13]. Given the success of Cycle GAN methodologies for image quality improvement, we have selected it as the baseline for our work.

In the specific subject of restoration of fetal MRI, Lim *et al.* [24] proposes a Conditional GAN (CGAN) framework where the generator has an autoencoder structure formed by residual blocks and long skip connections. The discriminator architecture is a CNN composed of three 5×5 convolutional layers with instance normalization and leaky ReLU as activation functions. The generator receives as input motion-corrupted images and outputs motion-free images. The discriminator receives as input two pairs of images: the fake one, formed by the concatenation of a free-of-artifacts image produced by the generator with the original artifact image, and the real pair obtained through the concatenation of a real free-of-artifacts image with the corresponding motion-corrupted image. The discriminator loss includes a variation of the Wasserstein GAN [25] and a gradient penalty term to boost convergence. The generator loss combines the same Wasserstein GAN loss of the discriminator, the $L1$, and a perceptual loss that uses features extracted by a pre-trained CNN to compute the similarity between

the ground truth and the generator's output. Differently from this approach, our work is Cycle GAN based instead of using CGAN strategy. Our discriminator does not receive pairs of images, and the losses do not include Wasserstein GAN components or the gradient penalty term. Besides, our discriminator is based on the patch GAN proposed by Demir *et al.* [15], and the generator is a lightweight Unet where the traditional convolutional layers are substituted by depth-wise separable convolutions [17] [26].

2. Data Description and Data Augmentation

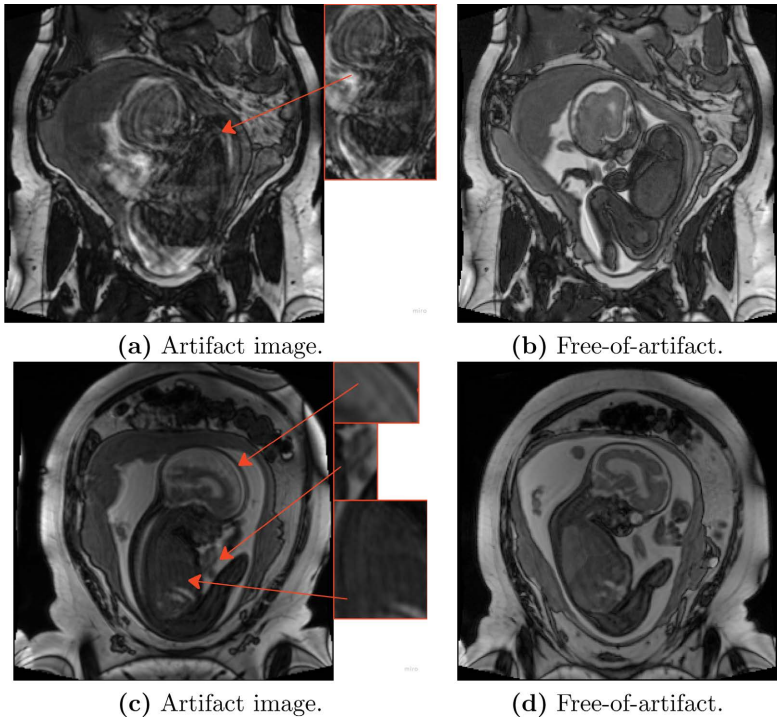
The dataset used in the present paper is composed of fetal MRI images of 38 distinct patients. MRI examination was performed by a 1.5-T scanner (Magnetom Aera, Siemens, Erlangen, Germany) with the surface coil positioned on the abdomen. We applied a 3D T2-weighted true fast imaging sequence with steady-state precession (truefisp) in sagittal plane (TR/TE = 3.02/1.43 ms); isotropic voxel 1.0 × 1.0 × 1.0 mm^3); matrix: 256 × 256 mm, 136 slices, with a total acquisition time of 26 s. Maternal sedation was not used.

To obtain the fetal MRI, the patient was positioned in dorsal or left lateral decubitus, with the feet entering the magnet first. For each patient, two volumes were acquired at the time of image acquisition. The first volume was acquired using a controlled setup with image acquisition during maternal breath-hold to produce high-quality images with low levels of artifacts. The second volume was acquired with a relaxed protocol to produce images with a high number of artifacts. In this way, we imposed that we have images with artifacts (domain A) and free-of-artifacts (domain B). Among the 72 MRI volumes obtained from 36 patients, we randomly selected 26 patients to compose the training and 10 patients for the test dataset. From the selected volumes, we could extract 2268 images with artifacts (domain A) and 1905 images free of artifacts (domain B) for the training dataset. Analogously, we separated 543 images with artifacts and 534 images free of artifacts to compose the test dataset.

In **Figures 1(a)-(d)**, two image pairs are presented, where **Figure 1(a)** and **Figure 1(c)** show images from domain A (with artifacts), and **Figure 1(b)-(d)** show images from domain B (free-of-artifacts). While these pairs are relevant from a medical perspective, they also demonstrate the difficulty of defining matching pairs for supervised training.

It is not guaranteed that images obtained in a controlled environment will be free of artifacts due to fetal movements. This can be observed in **Figure 1**, where the fetus is in a different position in **Figure 1(c)** & **Figure 1(d)**, and the hand present in **Figure 1(d)** is not present in **Figure 1(c)**. Subtle differences are also visible in **Figure 1(a)** and **Figure 1(b)**, where the legs and objects close to the head are slightly different in their positions. The indicated regions in **Figure 1(a)** and **Figure 1(c)** show examples of movement artifacts that occur in the volumes.

As a result, it is very difficult to obtain matching pairs of images in domains A and B due to these movements. Besides, the images in **Figure 1(a)** and **Figure 1(c)** show regions with artifacts (banding and motion artifacts) that we aim to remove.



(a) Artifact image. (b) Free-of-artifact.

(c) Artifact image. (d) Free-of-artifact.

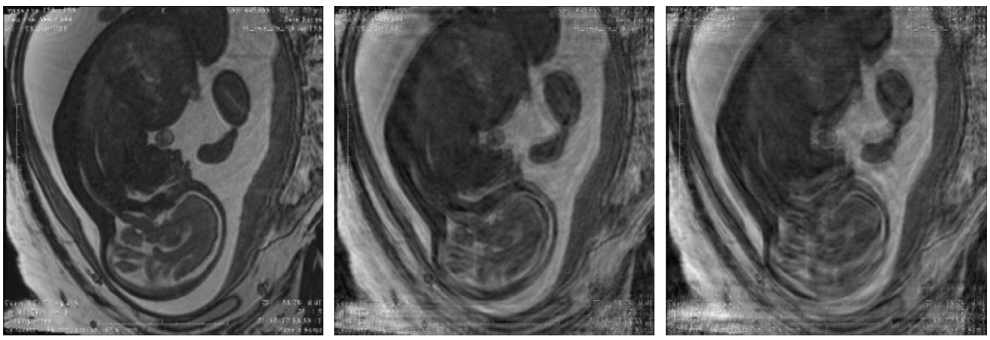
Figure 1. (a) and (c) Images in the artifacts domain (domain A). (b) and (d) Images in the free-of-artifacts domain (domain B) corresponding to images (a) and (c) from the medical perspective, respectively.

To prevent overfitting during neural network tests, we utilized data augmentation techniques. During the training step, each batch image passed to the model undergoes data augmentation as follows: Random Crop (randomly scaled between 0.65 and 1 (probability of 25%); Random brightness and contrast (probability of 20%); Random shift scale and rotation (probability of 20%); Random JPG compression (probability of 70%); Grid distortion (probability of 15%); Piece-wise affine transform (probability of 20%). All mentioned transformations were performed/implemented using the Python library Albumentations [27].

To further test our model's robustness, we simulated fetal motion noise in MR images using the method proposed by Duffy *et al.* [16] that presents an image-based approach that takes advantage of the Fourier transform translation and rotation properties where rigid motion artifacts in MR images can be simulated by rotations and translations in the frequency domain followed by the inverse Fourier transform. To create an artificial set of paired images, we applied those artificial artifacts to images in domain B, resulting in a new domain, which we named domain C. Domain B and domain C are considered paired domains, as each image in domain B has a corresponding artificially corrupted image in domain C. We included this artificial data in the unsupervised training process to improve model performance, incorporating a supervised term in the loss function.

To control the level of artifacts presented in the synthetic data we use the Structural Similarity Index Measure (SSIM). Specifically, for each image $\mathbf{x} \in B$ we

only accept a corrupted version $\mathbf{y} \in C$ if $s_0 < SSIM(\mathbf{x},\mathbf{y}) < s_1$, where (s_0, s_1) are given parameters. Knowing that $0 \le SSIM(\mathbf{x},\mathbf{y}) \le 1$, we consider the following level of artifacts for this paper: $(s_0, s_1) = (0.7, 0.8)$. We provide two examples of artificially corrupted images in **Figure 2** applying the selected level of artifacts to a clean image.



(a) Original. (b) Synthetic. (c) Synthetic.

Figure 2. (a) Original image in domain B; (b) and (c) Examples of artificial images in domain C generated from the original image (a).

3. Cycle GAN

The Cycle GAN is an unsupervised generative model that aims to learn two mapping functions, $G: X \mapsto Y$ and $F: Y \mapsto X$, between two different domains, X and Y. To handle the two distinct domains, two discriminators, $D_X: X \mapsto [0,1]$ and $D_Y: Y \mapsto [0,1]$, are used. For a given image z, D_X and D_Y discriminators provide the probability $p(z \in X)$ and $p(z \in Y)$, respectively. This is achieved by minimizing the equation below with respect to the generators G and F and maximizing it with respect to the discriminators D_X and D_Y [21]:

$$\begin{aligned}\mathcal{L}(G,F,D_X,D_Y) = {} & \mathcal{L}_{\text{GAN}}(G,D_Y,X,Y) + \mathcal{L}_{\text{GAN}}(F,D_X,Y,X) \\ & + \lambda_1 \mathcal{L}_{\text{cyc}}(G,F) + \lambda_2 \mathcal{L}_{\text{id}}(G,F),\end{aligned} \tag{1}$$

where,

$$\mathcal{L}_{\text{GAN}}(G,D_Y,X,Y) = \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\log D_Y(y)\right] + \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\log(1 - D_Y(G(x)))\right], \tag{2}$$

$$\mathcal{L}_{\text{cyc}}(G,F) = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\left\|F(G(x)) - x\right\|_1\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\left\|G(F(y)) - y\right\|_1\right], \tag{3}$$

$$\mathcal{L}_{\text{id}}(G,F) = \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\left\|G(y) - y\right\|_1\right] + \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\left\|F(x) - x\right\|_1\right], \tag{4}$$

being $p_{\text{data}}(x)$ and $p_{\text{data}}(y)$ the data distributions in X and Y, respectively, and $\mathcal{L}_{\text{GAN}}(F,D_X,Y,X)$ obtained analogously to expression (2). The objective of the generators G and F in Equation (1) is to minimize the loss against the discriminators D_X and D_Y, respectively. The generators attempt to generate images that are similar to their respective domains, while the discriminators try to differentiate between real and generated images. Equation (3) ensures that $F(G(x)) \approx x$ and $G(F(y)) \approx y$, allowing the generators to translate images be-

tween domains. Finally, Equation (4) encourages the generators to preserve image details, as the generators will be close to an identity function when input images from the target domain are provided [21].

The Cycle GAN model is constructed as an unsupervised model because its loss function (1) does not require pairs $(x, y) \in X \times Y$ to be provided. Only one set of images from each domain is necessary. However, if there are a set of images with matching pairs $(\overline{x}, \overline{y}) \in \overline{X} \times \overline{Y} \subset X \times Y$, where $\overline{X}$ and $\overline{Y}$ are subsets of X and Y respectively, we can add the following loss to Equation (1):

$$\mathcal{L}_{\text{paired}}(G, F) = \frac{1}{|\overline{X}|} \sum_{(\overline{x}_i, \overline{y}_i) \in (\overline{X}, \overline{Y})} \left(\left\| G(\overline{x}_i) - \overline{y}_i \right\|_1 + \left\| F(\overline{y}_i) - \overline{x}_i \right\|_1 \right). \tag{5}$$

Therefore, the loss function for the semi-supervised Cycle GAN is:

$$\mathcal{L}_s(G, F, D_X, D_Y) = \mathcal{L}(G, F, D_X, D_Y) + \lambda_3 \mathcal{L}_{\text{paired}}(G, F), \tag{6}$$

where we maximize the Equation (6) with respect to the discriminators D_X and D_Y, and minimize with respect to the generators G and F.

4. Results and Discussions

As shown in Section 2, we possess the original dataset and the artificially generated dataset created by the method proposed by Duffy *et al.* [16]. In this sense, we will test two different training styles for the Cycle GAN:

- Cycle GAN V1: This semi-supervised Cycle GAN model uses the original dataset for the unsupervised terms and the synthetic dataset for the supervised term in the loss function given in Equation (6).
- Cycle GAN V2: This is a supervised Cycle GAN model that uses only the synthetic pairs for all terms in the loss function given in Equation (6).

The application of Cycle GAN V2 is motivated by the fact that images in domains A and B are acquired in distinct moments. Consequently, as observed in **Figure 1**, besides motion artifacts, we also observe different structures when comparing one image with its counterpart. Simulated images are free of such problems. Hence, employing the Cycle GAN V2 approach, we expect to achieve a balance between artifact and artifact-free data during the training process. Given the inherent ability of Cycle GAN to preserve image details, we anticipate obtaining competitive results.

In addition to the visual insights, we will leverage the synthetic dataset generated through the methodology outlined by Duffy *et al.* [16], for conducting a quantitative analysis of the results. In this manner, we calculate the SSIM between the artifact-free images and the generator's outputs. We then compare this SSIM value with the SSIM between the artificially corrupted data and their corresponding artifact-free images, which is detailed in **Table 1**. This table provides information about the mean, standard deviation (STD), minimum SSIM (MIN), maximum SSIM (MAX), and SSIM percentiles. This procedure enables us to assess the quality of the restoration achieved.

Table 1. SSIM percentiles between artificially corrupted images and their corresponding free of artifacts images.

DATA	SSIM ↑						
	MEAN	STD	MIN	25%	50%	75%	MAX
TRAIN	0.75	0.028	0.694	0.726	0.752	0.774	0.797
TEST	0.751	0.028	0.693	0.727	0.754	0.776	0.796

4.1. Results over Synthetic Data

This section showcases the outcomes derived from the semi-supervised Cycle GAN models V1 and V2, which incorporate a supervised term (Equation (5)) to yield the loss described in Equation (6). With this model, our objective is to enhance artifact removal by introducing the supervised term while simultaneously preserving image structures through the influence of the unsupervised term within the loss function.

Both models were trained for 50 epochs using the Adam optimizer with a learning rate of 2×10^{-4}, momentum of 0.5, and variance momentum of 0.999. Furthermore, we simultaneously train all components, utilizing an alternating training approach that cycles between training the discriminator A, the discriminator B, and generators the A and B simultaneously. For the loss hyperparameters, we empirically select $(\lambda_1, \lambda_2) = (10, 5)$ and also evaluate the model's performance across the values of λ_3 in $(1.0, 2.5, 5.0, 7.5, 10.0)$ to show its effect in the generator's performance since this parameter scales the influence of the supervised term in the loss (6).

The results of each model over the synthetic data can be seen in **Table 2**. Additionally, **Figure 3** presents the results on four selected images in the artifacts

Table 2. SSIM statistics over each trained model on the test set where the best models are highlighted in gray.

Model	λ_3	SSIM ↑						
		MEAN	STD	MIN	25%	50%	75%	MAX
CycleGAN V1	1.0	0.751	0.032	0.670	0.727	0.751	0.775	0.848
	2.5	0.741	0.031	0.668	0.717	0.740	0.765	0.855
	5.0	0.746	0.032	0.666	0.722	0.747	0.769	0.868
	7.5	0.748	0.033	0.670	0.723	0.747	0.771	0.869
	10	0.744	0.031	0.671	0.721	0.746	0.769	0.853
CycleGAN V2	1.0	0.7724	0.036	0.677	0.7473	0.771	0.798	0.87169
	2.5	0.759	0.030	0.679	0.737	0.760	0.784	0.833
	5.0	0.769	0.036	0.666	0.742	0.769	0.796	0.866
	7.5	0.770	0.036	0.657	0.745	0.770	0.796	0.871
	10	0.7726	0.035	0.647	0.7479	0.770	0.799	0.87163

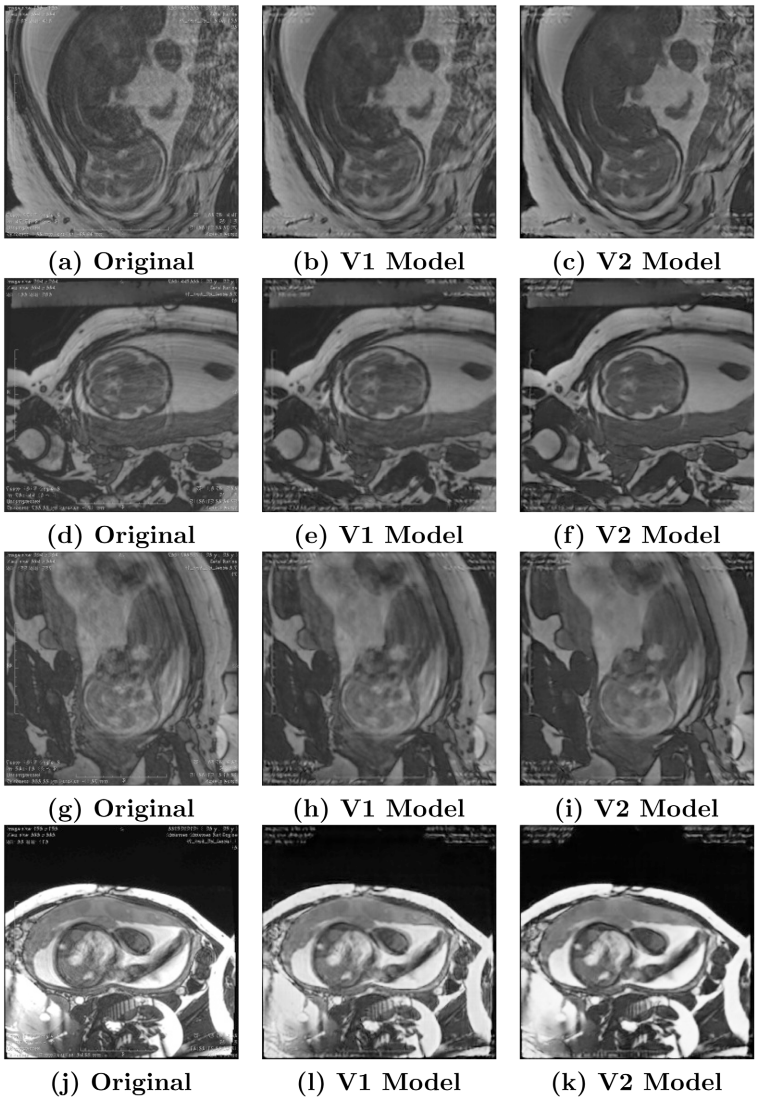


Figure 3. Results of Cycle GAN V1 and V2 over four examples of the synthetic data (test split).

domain and their corresponding free-of-artifacts versions, which were obtained by the generator G of the best models according to **Table 2**.

A fast analysis of **Table 2** shows that the best λ_3 for the cycle GAN V1 and V2 were 1.0 and 10.0, respectively. However, when we saw the SSIM mean and standard deviation in the V1 model, the SSIM stays in the statistical interval [0.751 - 0.028, 0.751 + 0.028] given by **Table 1**. On the other hand, **Table 2** shows that the Cycle GAN V2 surpassed the same statistical interval provided in **Table 1**. Additionally, for the best model (V2), at least 50% of the treated images obtained an SSIM score higher than the interval given in **Table 1** and 25% outperformed the limit value of 0.799 defined for the synthetic dataset.

In **Figure 3(b)**, **Figure 3(e)** and **Figure 3(l)** a subtle improvement can be observed using the V1 model, respectively. However, the results depicted in **Figure 3(h)** reveals that the model did not make significant improvements in the arti-

facts present.

It can be observed from **Figure 3** that the Cycle GAN V2 has improved all images, achieving better results than the V1 model in terms of artifact removal. In **Figure 3(c)** and **Figure 3(f)** the model has removed most of the artifacts present in the image. Especially in **Figure 3(c)**, the model has provided significant restoration in the fetal body with partial removal of artifacts in the entire image. In addition, in **Figure 3(i)**, a higher level of artifacts is evident, and as a result, the models were not able to eliminate all of them. However, the V2 model made the image clearer, particularly on the cyst located near the fetus's neck (the white spherical structure on the neck), which is an important result from the medical viewpoint since this structure needs special attention for diagnosis.

4.2. Comparisons with the Other Methods

In this subsection, we conduct a comparative analysis between the top-performing Cycle GAN model (V2 with $\lambda_3 = 10.0$) and two alternative supervised approaches: the original Pix2Pix [18] and the Mobile Unet [17], all of which were trained using synthetic data. We did not include other approaches for artifact removal in MRI data, because the codes are not available for training [24] [28]. Unfortunately, Liu *et al.* [28] made available only the inference code.

We trained the Pix2Pix using the default configurations given in the official repository [29] and, for the Mobile-Unet, we used the *LogCosh* as loss function with *AdamW* optimizer being the learning rate 10^{-3} with betas (0.9, 0.999).

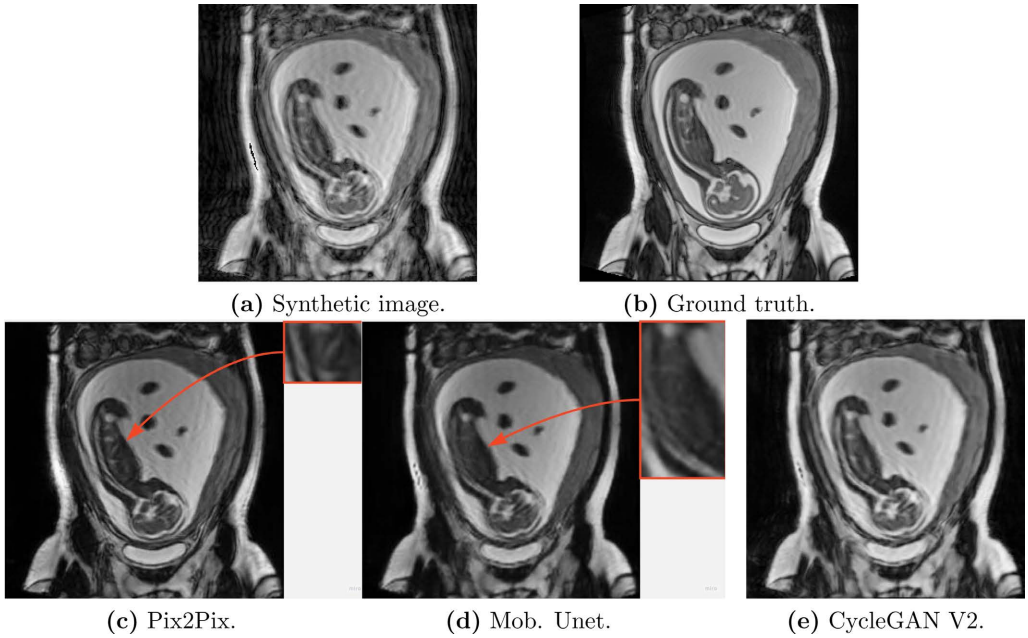
We divided the experiments into two phases. In the first phase, we employ the SSIM and MAE to measure the performance of each model when applied to synthetic data. In the second phase, we utilize Jensen-Shannon Divergence (JSD) and Fréchet Inception Distance (FID) to evaluate the performance of each method with real data. These metrics consider the disparity between the distributions and the image quality of the processed images in comparison to authentic artifact-free images. Additionally, we conduct a visual analysis of the results obtained by each method.

Table 3 summarizes the SSIM results obtained from each model when applied to synthetic data. Upon examination of this table, it becomes apparent that Cycle GAN V2 yielded the least favorable SSIM results in comparison to Mobile Unet and Pix2Pix.

In **Figure 4**, we showcase an example from the synthetic dataset, which is

Table 3. SSIM statistics of each model.

Model	SSIM ↑						
	MEAN	STD	MIN	25%	50%	75%	MAX
Cycle GAN V2	0.7726	0.0357	0.6473	0.7479	0.7704	0.7993	0.8716
Pix2Pix	0.8031	0.0327	0.6667	0.7812	0.8040	0.8271	0.8842
Mob. Unet	0.8115	0.0362	0.6704	0.7872	0.8121	0.8389	0.8943



(a) Synthetic image. (b) Ground truth.

(c) Pix2Pix. (d) Mob. Unet. (e) CycleGAN V2.

Figure 4. (a) Synthetic image in the artifacts domain. (b) Ground truth. (c) Pix2Pix result (d) MobileUnet result. (e) Cycle GAN V2 result.

worth noting that all methods exhibit significant improvements. Nevertheless, it is essential to highlight some nuances: Pix2Pix, while effective in enhancing the image, introduces certain artifacts that are absent in the results obtained with Mobile Unet and Cycle GAN V2, as evidenced in **Figure 4(c)**. (highlighted regions in Pix2Pix results). Conversely, in **Figure 4(d)** (highlighted region), Mobile Unet appears to cause the disappearance of certain objects.

From a medical standpoint, the generation of artifacts or alterations in specific regions can potentially lead to incorrect medical diagnoses. A similar concern holds for the disappearance of objects. In this regard, although Cycle GAN V2 may not enhance synthetic images to the same extent as other methods, it avoids these issues, primarily due to the incorporation of cycle consistency loss, which aids in preserving intricate details of the original images. This assertion finds support in the results presented in **Table 4**. This table summarizes the results of each model when applied to the synthetic images, evaluated using the normalized Mean Absolute Error (MAE). MAE works by a pixel-to-pixel comparison between the processed images and the ground truths, offering insights into the detail preservation capabilities of each model.

Table 4. Normalized MAE percentiles of each model.

Model	MAE (normalized) ↓						
	MEAN	STD	MIN	25%	50%	75%	MAX
Cycle GAN V2	0.4117	0.0546	0.1662	0.3831	0.4154	0.4459	0.5722
Pix2Pix	0.4873	0.0459	0.3163	0.4579	0.4887	0.5143	0.6614
Mob. Unet	0.4440	0.0397	0.2830	0.4189	0.4450	0.4692	0.6002

According to the previous analysis, we can conclude that the model that best preserves visual similarity after the treatment is the Mobile Unet. However, the model that provides artifact removal with the best detail preservation is the Cycle GAN V2.

Figure 5 presents the results of the considered techniques over three images selected by the physician of our team. These images were acquired following the standard protocol. We present the corresponding experiments in **Figure 5** where the first column displays the original images with artifacts, while the second to fourth columns exhibit the results obtained using Pix2Pix, Mobile Unet, and Cycle GAN V2, respectively. As we lack ground truth images in our dataset due to the challenges associated with obtaining them, we employed JSD and FID to evaluate the dissimilarity between the treated images and artifact-free images.

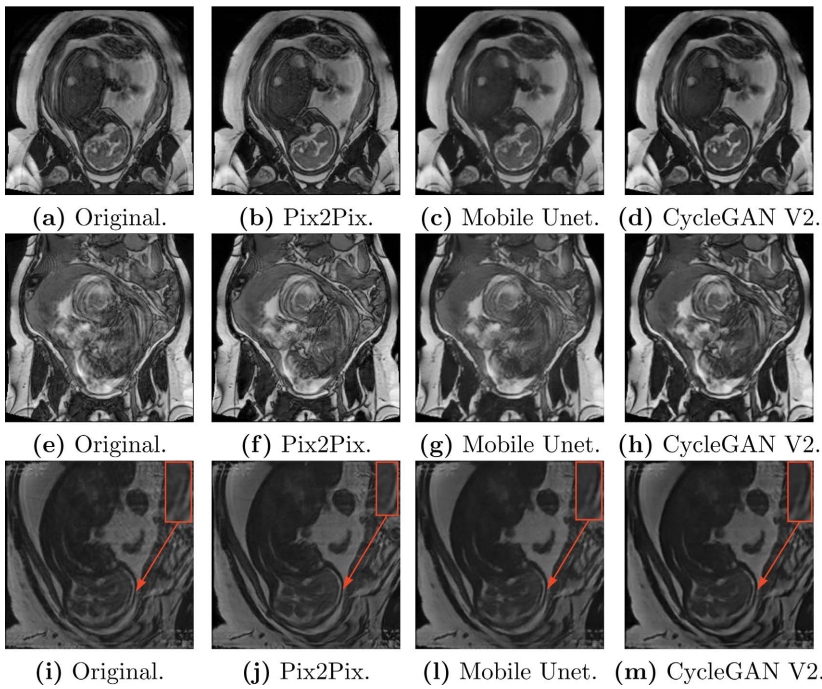


Figure 5. Results of Pix2Pix, Mobile Unet and Cycle GAN V2 over three real examples of images with artifacts in the test dataset.

An analysis of the predictions presented in **Figure 5** shows that Pix2Pix does not achieve good results in **Figure 5(f)** since it is evident that numerous artifacts persist in the treated image, and in some instances, the neural network introduces new artifacts, as was observed in the synthetic data. Besides, **Figure 5(j)** shows a subtle region that vanished (in white) on the highlighted region. Such issues are notably absent in the results obtained with Mobile Unet and Cycle GAN V2.

In both the synthetic and the real data images presented in **Figure 4** and **Figure 5**, Cycle GAN V2 provided promising results while avoiding the generation of artifacts, as observed in Pix2Pix results. Moreover, the edges of the Cycle GAN V2 results are more well-defined when compared to Mobile Unet, particularly noticeable in the entire fetus in **Figure 4(d)** & **Figure 4(e)** and **Figure 5(h)**

& **Figure 5(g)**. In some examples, as in **Figure 5(d)**, Cycle GAN V2 intensifies the contrast between black and white regions while preserving intricate details. Although this characteristic can be advantageous for distinguishing various regions, it alters the image contrast, which potentially contributes to the comparatively lower SSIM scores of Cycle GAN V2 compared to other methods since the contrast (variances) appear in the denominator of the SSIM formula [30].

The evaluation based on SSIM and MAE metrics reveals that Mobile Unet succeeded in removing a greater number of artifacts from the images. However, this came at the cost of losing some critical details or regions within the original images that could be of significance for medical diagnosis. On the other hand, Cycle GAN V2 made an interesting treatment while preserving essential details (notably well-defined edges) more efficiently compared to Mobile Unet, thanks to the inclusion of cycle consistency loss during the training process. To complete our analysis of the real data, we conducted a comparison between the set of treated images and the genuine artifact-free images. This assessment relied on JSD and FID scores to have insights into the distance between the set's distribution and the image quality of the treated images, respectively. These results are summarized in **Table 5**.

Table 5. JSD and FID scores between the images without artifacts and the treated images.

Model	JSD ↓	FID ↓
Pix2Pix	0.0220	43.58
Mobile Unet	0.0282	61.48
Cycle GAN V2	0.0229	51.02

By examining the JSD and FID scores, we can note that the images generated by Pix2Pix have the closest resemblance in terms of distribution and image quality when compared to genuine artifact-free images. We attribute the higher FID score of Cycle GAN V2 compared to Pix2Pix, as mentioned earlier, to the contrast adjustments that can occur in some images. However, this characteristic, as illustrated in **Figure 5(d)**, enhances the clarity and differentiation of various regions, which can be advantageous for visualizing and distinguishing important structures. Despite Cycle GAN V2's FID score not closely matching that of Pix2Pix, the distributions are similar according to JSD. To conclude our analysis, the Pix2Pix is the best model according to JSD and FID scores, however, the creation of artifacts and the disappearance of regions must be considered to select a good model for medical images. In this regard, Cycle GAN V2 stands out as the best model for reducing the motion artifacts of fetal MR images and facilitating medical diagnostics.

5. Final Remarks

In this study, we utilized Cycle GAN model with the addition of a supervised term to automatically remove motion noise from fetal MRI. We aimed to obtain

restored images that preserve suitable details using the Cycle GAN baseline. To create image pairs for training, we adopted the technique proposed by Duffy *et al.* [16] to simulate general magnetic resonance motion in the images of domain B. This was necessary as we lacked image pairs in the original dataset.

The proposed supervised Cycle GAN, referred to as the Cycle GAN V2 model, exclusively employed the synthetic dataset and outperformed its predecessor Cycle GAN V1, as evident from the SSIM scores computed over the synthetic pairs and visual analysis.

Furthermore, we compared the best Cycle GAN model with other supervised approaches, the Pix2Pix and Mobile Unet, by directly using the synthetic pairs. The results given the SSIM over the synthetic dataset showed that the model that best preserves the visual similarity with respect to free-of-artifacts images was the Mobile Unet.

On the other hand, the Pix2Pix model produces a distribution more similar to the images in domain B with higher visual fidelity according to JSD and FID scores, respectively. Nevertheless, our visual analysis revealed that Pix2Pix can generate additional regions not present in the original images. Additionally, both Pix2Pix and Mobile Unet exhibited the undesirable tendency to remove regions from the original images that were not artifacts caused by fetal movement. Such characteristics were verified through MAE analysis, which indicated that Cycle GAN V2 yielded the lowest error. Given that the creation or vanishing of regions can impact medical diagnostics, and considering that Cycle GAN V3's image quality is on par with other methods, Cycle GAN V2 is the best model to help with the motion artifacts reduction on fetal MRI.

In future studies, we will expand our dataset to properly generalize the fetal motion artifacts and explore other variations of Cycle GANs and loss functions to further improve our approach. A full comparison with the most recent models can also be done.

Acknowledgements

The study involving fetal magnetic resonance imaging was approved by the DASA ethics committee. All patients involved signed a consent form approving the use of the images used in this study. Due to the data being privately owned by DASA, it will not be made available by us.

This work was carried out with the support of the Coordenaão de Aperfeicoamento de Pessoal de Nível Superior-Brasil (CAPES)-Financing Code 001.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Machado-Rivas, F., Jaimes, C., Kirsch, J.E. and Gee, M.S. (2020) Image-Quality Op-

timization and Artifact Reduction in Fetal Magnetic Resonance Imaging. *Pediatric Radiology*, **50**, 1830-1838. https://doi.org/10.1007/s00247-020-04672-7

[2] Victoria, T., Jaramillo, D., Roberts, T.P.L., Zarnow, D., Johnson, A.M., Delgado, J., *et al.* (2014) Fetal Magnetic Resonance Imaging: Jumping from 1.5 to 3 Tesla (Preliminary Experience). *Pediatric Radiology*, **44**, 376-386. https://doi.org/10.1007/s00247-013-2857-0

[3] Victoria, T., Johnson, A.M., Edgar, J.C., Zarnow, D.M., Vossough, A. and Jaramillo, D. (2016) Comparison between 1.5-T and 3-T MRI for Fetal Imaging: Is There an Advantage to Imaging with a Higher Field Strength? *American Journal of Roentgenology*, **206**, 195-201. https://doi.org/10.2214/ajr.14.14205

[4] Gu, J. and Ye, J.C. (2021) AdaIN-Based Tunable CycleGAN for Efficient Unsupervised Low-Dose CT Denoising. *IEEE Transactions on Computational Imaging*, **7**, 73-85. https://doi.org/10.1109/tci.2021.3050266

[5] Jin, K.H., Um, J., Lee, D., Lee, J., Park, S. and Ye, J.C. (2016) MRI Artifact Correction Using Sparse + Low-Rank Decomposition of Annihilating Filter-Based Hankel Matrix. *Magnetic Resonance in Medicine*, **78**, 327-340. https://doi.org/10.1002/mrm.26330

[6] Landi, G. and Zama, F. (2022) A Variational Approach to Gibbs Artifacts Removal in MRI. *Annali Dell'universita' di Ferrara*, **68**, 465-481. https://doi.org/10.1007/s11565-022-00431-8

[7] Liu, S., Thung, K., Qu, L., Lin, W., Shen, D. and Yap, P. (2021) Learning MRI Artefact Removal with Unpaired Data. *Nature Machine Intelligence*, **3**, 60-67. https://doi.org/10.1038/s42256-020-00270-2

[8] Mishro, P.K., Agrawal, S., Panda, R. and Abraham, A. (2022) A Survey on State-of-the-Art Denoising Techniques for Brain Magnetic Resonance Images. *IEEE Reviews in Biomedical Engineering*, **15**, 184-199. https://doi.org/10.1109/rbme.2021.3055556

[9] Mohan, J., Krishnaveni, V. and Guo, Y. (2014) A Survey on the Magnetic Resonance Image Denoising Methods. *Biomedical Signal Processing and Control*, **9**, 56-69. https://doi.org/10.1016/j.bspc.2013.10.007

[10] Ahishakiye, E., Van Gijzen, M.B., Tumwiine, J., Wario, R. and Obungoloch, J. (2021) A Survey on Deep Learning in Medical Image Reconstruction. *Intelligent Medicine*, **1**, 118-127. https://doi.org/10.1016/j.imed.2021.03.003

[11] Apostolidis, K.D. and Papakostas, G.A. (2021) A Survey on Adversarial Deep Learning Robustness in Medical Image Analysis. *Electronics*, **10**, Article No. 2132. https://doi.org/10.3390/electronics10172132

[12] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H. and Nandi, A.K. (2022) Medical Image Segmentation Using Deep Learning: A Survey. *IET Image Processing*, **16**, 1243-1267. https://doi.org/10.1049/ipr2.12419

[13] Armanious, K., Jiang, C., Abdulatif, S., Kustner, T., Gatidis, S. and Yang, B. (2019) Unsupervised Medical Image Translation Using Cycle-MedGAN. 2019 27*th European Signal Processing Conference* (*EUSIPCO*), Volume 1, 1-5. https://doi.org/10.23919/eusipco.2019.8902799

[14] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI* 2015, Volume 9351, 234-241.

[15] Demir, U. and Unal, G.B. (2018) Patch-Based Image Inpainting with Generative Adversarial Networks.

[16] Duffy, B.A., Zhao, L., Sepehrband, F., Min, J., Wang, D.J., Shi, Y., *et al.* (2021) Retrospective Motion Artifact Correction of Structural MRI Images Using Deep Learn-

ing Improves the Quality of Cortical Surface Reconstructions. *NeuroImage*, **230**, Article ID: 117756. https://doi.org/10.1016/j.neuroimage.2021.117756

[17] Jing, J., Wang, Z., Rätsch, M. and Zhang, H. (2020) Mobile-Unet: An Efficient Convolutional Neural Network for Fabric Defect Detection. *Textile Research Journal*, **92**, 30-42. https://doi.org/10.1177/0040517520928604

[18] Isola, P., Zhu, J., Zhou, T. and Efros, A.A. (2017) Image-to-Image Translation with Conditional Adversarial Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 5967-5976. https://doi.org/10.1109/cvpr.2017.632

[19] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S. (2017) GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6629-6666.

[20] Menéndez, M.L., Pardo, J.A., Pardo, L. and Pardo, M.C. (1997) The Jensen-Shannon Divergence. *Journal of the Franklin Institute*, **334**, 307-318. https://doi.org/10.1016/s0016-0032(96)00063-4

[21] Zhu, J., Park, T., Isola, P. and Efros, A.A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2242-2251. https://doi.org/10.1109/iccv.2017.244

[22] Godenschweger, F., Kägebein, U., Stucht, D., Yarach, U., Sciarra, A., Yakupov, R., *et al*. (2016) Motion Correction in MRI of the Brain. *Physics in Medicine and Biology*, **61**, R32-R56. https://doi.org/10.1088/0031-9155/61/5/r32

[23] Chen, Z., Pawar, K., Ekanayake, M., Pain, C., Zhong, S. and Egan, G.F. (2022) Deep Learning for Image Enhancement and Correction in Magnetic Resonance Imaging—State-of-the-Art and Challenges. *Journal of Digital Imaging*, **36**, 204-230. https://doi.org/10.1007/s10278-022-00721-9

[24] Lim, A., Lo, J., Wagner, M.W., Ertl-Wagner, B. and Sussman, D. (2023) Motion Artifact Correction in Fetal MRI Based on a Generative Adversarial Network Method. *Biomedical Signal Processing and Control*, **81**, Article ID: 104484. https://doi.org/10.1016/j.bspc.2022.104484

[25] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A.C. (2017) Improved Training of Wasserstein GANs. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 5769-5779.

[26] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1800-1807. https://doi.org/10.1109/cvpr.2017.195

[27] Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M. and Kalinin, A.A. (2020) Albumentations: Fast and Flexible Image Augmentations. *Information*, **11**, Article No. 125. https://doi.org/10.3390/info11020125

[28] Liu, S., Thung, K.-H., Qu, L., Lin, W., Shen, D. and Yap, P.-T. (2020) Inference Code: Learning MRI Artefact Removal with Unpaired Data. https://zenodo.org/record/4474599

[29] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A.A. (2017) CycleGAN and pix2pix in PyTorch. https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

[30] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. https://doi.org/10.1109/tip.2003.819861