
Translation of Fetal Brain Ultrasound Images into Pseudo-MRI Images using Artificial Intelligence

Naomi Silverstein

Technion Institute of Technology
Haifa

naomish@campus.technion.ac.il

Efrat Leibowitz

Rambam Health Care Campus
Haifa

e_leibowitz@rambam.health.gov.il
1

Ron Beloosesky

Rambam Health Care Campus
Haifa

ronbel13@gmail.com

Haim Azhari

Technion Institute of Technology
Haifa

haim@bm.technion.ac.il

Abstract

Ultrasound is a widely accessible and cost-effective medical imaging tool commonly used for prenatal evaluation of the fetal brain. However, it has limitations, particularly in the third trimester, where the complexity of the fetal brain requires high image quality for extracting quantitative data. In contrast, magnetic resonance imaging (MRI) offers superior image quality and tissue differentiation but is less available, expensive, and requires time-consuming acquisition. Thus, transforming ultrasonic images into an MRI-mimicking display may be advantageous and allow better tissue anatomy presentation. To address this goal, we have examined the use of artificial intelligence, implementing a diffusion model renowned for generating high-quality images. The proposed method, termed "Dual Diffusion Imposed Correlation" (DDIC), leverages a diffusion-based translation methodology, assuming a shared latent space between ultrasound and MRI domains. Model training was obtained utilizing the "HC18" dataset for ultrasound and the "CRL fetal brain atlas" along with the "FeTA" datasets for MRI. The generated pseudo-MRI images provide notable improvements in visual discrimination of brain tissue, especially in the lateral ventricles and the Sylvian fissure, characterized by enhanced contrast clarity. Improvement was demonstrated in Mutual information, Peak signal-to-noise ratio, Fréchet Inception Distance, and Contrast-to-noise ratio. Findings from these evaluations indicate statistically significant superior performance of the DDIC compared to other translation methodologies. In addition, a Medical Opinion Test was obtained from 5 gynecologists. The results demonstrated display improvement in 81% of the tested images. In conclusion, the presented pseudo-MRI images hold the potential for streamlining diagnosis and enhancing clinical outcomes through improved representation.

Keywords Diffusion Model, Deep-Learning, Image Translation, Ultrasound, MRI, Fetal, Lateral-Ventricles, Sylvian-Fissure, Correlation.

1 Introduction

Ultrasound (US) stands out as a highly cost-effective modality within contemporary medical imaging. Its non-invasive attributes, prevalent availability, and safety make it applicable across diverse medical domains. Despite its pronounced merits, the US has notable disadvantages. Particularly, ultrasound-generated images exhibit a **suboptimal signal-to-noise ratio (SNR)**. Moreover, the images are characterized by considerable **speckle noise**, resulting in limited anatomical visualization. As a result, clinical insights are compromised and necessitate a substantial reliance on the interpretative skills of the radiologist (1).

Conversely, magnetic resonance imaging (**MRI**) offers **enhanced soft tissue differentiation**, a significantly **higher Signal-to-Noise Ratio (SNR)**, and **better anatomical display**. However, the trade-off lies in the **protracted duration** required for MRI image acquisition, the need for substantial infrastructure, high cost, and limited accessibility compared to ultrasound (2). Consequently, the potential transformation of ultrasonic images into a display akin to that of MRI promises potentially improved diagnostic capabilities.

In recent years, neural networks have made significant progress, especially in the field of image generation and image translation. One of the first architectures to achieve the image translation goal was the **Cycle Generative Adversarial Network (CycleGAN)** (3), followed by more GAN-based architectures (4–6). Within the domain of US-MRI translation, several studies have been conducted employing GAN-based methodologies for **US-to-MRI conversion** (7–9). Similarly, there have been significant advances in **iterative generative models** such as **score-based models** (10) and **diffusion models** (11,12), which have exhibited the capacity to generate synthetic images of **superior quality** comparable to those produced by **contemporary GAN methodologies** (13). These **diffusion models** operate by systematically **adding Gaussian noise** to an image until the **image reaches a state of complete white noise**. Subsequently, a model is trained to execute the **inverse process**, gradually **diminishing Gaussian noise** from the white noise until attaining a coherent, high-quality image. In the field of **diffusion-based image translation**, contrary to **GAN-based** architectures, relevant studies are absent (to the best of our knowledge), addressing the synthesis of MRI images from US data.

An important component of routine prenatal ultrasound anomaly scans is the assessment of the posterior horns of the fetal lateral cerebral ventricles (14). The International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) guidelines note that: "due to artifacts in the near field of the image, caused by shadowing from the proximal parietal bone, in the standard trans ventricular plane, only the hemisphere and the lateral ventricle on the far side of the transducer are usually visualized clearly" (15). However, evaluating both ventricles is clinically significant, as fetal cerebral lateral ventriculomegaly is among the most common findings in routine prenatal ultrasounds (16). Additionally, the ability to identify the shape of the Sylvian fissure is crucial during routine ultrasound examinations, as it can serve as an indicator of cortical dysplasia. The **Sylvian fissure** is typically more distinctly visible on both sides on MRI images (17).

It is postulated here that there may be a potential clinical benefit in transforming US images into pseudo-MRI images, thereby enabling the incorporation of the high-quality visibility characteristic in MRI images into a cost-effective and expeditious ultrasound machine. The hypothesis advanced here suggests that such a transformation can be realized through the implementation of a modified **diffusion-based translation model**. The modified model suggested here **imposes correlation between the source and outcome images** in **every step** during the **reconstruction process** to better **preserve the anatomical information** and is therefore titled Dual Diffusion Imposed Correlation (DDIC).

2 Methods

2.1 Diffusion models

Initially, we describe the training methodology of diffusion models as introduced by Ho *et al.* (11) "Denoising Diffusion Probabilistic Models" (DDPM). Given an **image x_0** drawn from a **source data distribution $q(x_0)$** , the forward process is established by the **addition of a small amount of Gaussian noise** to x_0 over T steps, with **step sizes governed by the variance schedule $\{\beta_t \in (0, 1)\}_{t=1}^T$** . At each step, given the **image x_{t-1}** , the resulting next image x_t is derived from the incorporation of Gaussian noise. The definition of a **forward step $q(x_t | x_{t-1})$** is articulated as follows:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t} \cdot x_{t-1}, \beta_t I).$$

Eq. (1)

Defining $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, the final x_T can be calculated in a single step:

$$x_T = \sqrt{\bar{\alpha}_T} \cdot x_0 + \sqrt{1 - \bar{\alpha}_T} \cdot \epsilon, \quad \text{where } \epsilon \sim \mathcal{N}(0, I).$$

Eq. (2)

To generate an image from the initial condition $p(x_T) \sim \mathcal{N}(x_T; 0, I)$, a necessity arises for the reverse diffusion process. Since the elimination of the noise depends on the distribution of the source data, it becomes necessary to train a model to predict noise removal. The backward process defined by the function $p(x_{t-1} | x_t)$, given the noised image x_t , the function p calculates the resultant denoised image x_{t-1} :

$$p(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I),$$

Eq. (3)

where $\sigma_t^2 = \beta_t$, and μ_θ is the average of the added Gaussian noise which the model learns to predict. μ_θ is given by:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\underbrace{x_t}_{\substack{\text{mean of previous} \\ \text{noiseless image}}} - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right), \quad \substack{\text{Expectation} \\ (\text{average})} \rightarrow \substack{\text{Image with noise}} \quad \text{Eq. (4)}$$

where $\epsilon_\theta(x_t, t)$ is the trained model by minimizing the loss function L_t :

$$L_t = \mathbb{E}_{t \sim [1, T], x_0, \epsilon_t} [\|\epsilon_t - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon_t, t)\|^2]. \quad \substack{\text{Expectation} \\ (\text{average})} \quad \text{Eq. (5)}$$

The DDPM sampling method is not deterministic, meaning that different images can result from the same latent noise sample. Thus, we utilize the deterministic sampling method outlined in Denoising Diffusion Implicit Models (DDIM) (18):

$$x_{t+\Delta t} = \sqrt{\frac{\alpha_{t+\Delta t}}{\alpha_t}} \cdot x_t + \left(\sqrt{1 - \alpha_{t+\Delta t}} - \sqrt{1 - \alpha_t} \cdot \sqrt{\frac{\alpha_{t+\Delta t}}{\alpha_t}} \right) \cdot \epsilon_\theta(x_t, t),$$

Eq. (6)

where x_t is a noised image at timestep t , and $x_{t+\Delta t}$ is the obtained noised or denoised image when Δt is positive or negative accordingly. \rightarrow If Δt is positive it's noised, and negative is denoised

For conciseness, we define the symbol ForwardODE^{US} for the noising direction, meaning $\Delta t = 1$ and $\epsilon_\theta(x_t, t)$ is the learned denoiser from the US dataset.

$$x_{t+1} = \text{ForwardODE}^{US}(x_t, t),$$

Eq. (7)

It has been added noise or for t time

where x_t designates here the generated US image at timepoint t . In addition, we define the symbols BackwardODE^{MRI} and BackwardODE^{US} for the denoising process, meaning $\Delta t = -1$, and $\epsilon_\theta(x_t, t)$ is the learned MRI dataset denoiser for the former and the learned US dataset denoiser for the later.

$$x_{t-1} = \text{BackwardODE}^{US}(x_t, t),$$

Eq. (8)

$$y_{t-1} = \text{BackwardODE}^{MRI}(y_t, t),$$

Eq. (9)

where y_t designates the generated MRI image at timepoint t .

Dual Diffusion Implicit Bridge (DDIB) (19) is a method suggested for translating between unpaired images. DDIB comprises two separate diffusion processes, one for each domain. The algorithm involves

it is important that they are unpaired

1. Equation (1): The Single Step Noise Addition

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \underbrace{\sqrt{1 - \beta_t} \cdot x_{t-1}}_{\text{mean}}, \underbrace{\beta_t I}_{\text{variance}})$$

This equation describes a **Markov chain** that gradually adds Gaussian noise to an image. ⚡

- x_{t-1} : The image at the previous time step.
- x_t : The image at the current time step (which is slightly noisier).
- β_t : A hyperparameter known as the **variance schedule** (where $\beta_t \in (0, 1)$). It controls how much noise is added at each step.
- **The Mechanism:** The equation says that to get the next image x_t , you take the previous image x_{t-1} , scale it down slightly by $\sqrt{1 - \beta_t}$ (making the original signal slightly fainter), and add Gaussian noise with variance $\beta_t I$.

In Equation 2:

- **Components:**

- $\sqrt{\bar{\alpha}_T} \cdot x_0$: This represents the fraction of the original signal (the Ultrasound image) that remains. As T gets large, $\bar{\alpha}_T$ approaches 0, meaning the original signal fades away.
- $\sqrt{1 - \bar{\alpha}_T} \cdot \epsilon$: This represents the accumulated noise. As T gets large, this term dominates, and the image becomes pure Gaussian noise ($\mathcal{N}(0, I)$).

1. What is a Markov Chain?

In the context of this paper, the **Markov Chain** refers to the specific way the sequence of images is generated.

A process is considered a "Markov Chain" if the future state depends **only** on the current state, not on the history of how you got there. This is often called the "memoryless" property.

- **In this specific algorithm:** The process starts with a clean Ultrasound image (x_0). To get the slightly noisy image at step 2 (x_2), the model **only** needs to know what step 1 (x_1) looks like. It does not need to look back at the original image (x_0). ⚡
- **The Chain:** It creates a chain of events:

$$x_0 \rightarrow x_1 \rightarrow x_2 \cdots \rightarrow x_t$$

At any point t , the transition to the next step relies solely on the image at x_{t-1} .

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} \cdot x_{t-1}, \beta_t I)$$

The symbol \mathcal{N} represents a **Normal (Gaussian) Distribution**. In mathematics, a multidimensional Normal distribution is usually written as $\mathcal{N}(x; \mu, \Sigma)$, where μ is the **Mean** and Σ is the **Covariance Matrix**.

A. The Mean (μ): $\sqrt{1 - \beta_t} \cdot x_{t-1}$

This parameter represents the **center** of the distribution. It dictates what the new image (x_t) will mostly look like before the randomness is added.

- x_{t-1} : This is the image from the previous step.
- $\sqrt{1 - \beta_t}$: This is a scaling factor (a number slightly less than 1).
- **What it does:** This term takes the previous image and "dims" or shrinks the pixel values slightly. It represents the **signal** that is being preserved from the previous step. ☀️

B. The Covariance (Σ): $\beta_t I$

This parameter represents the **spread** or the **randomness** (noise) of the distribution.

- β_t : This is the "variance schedule". It is a small number that determines **how much noise is added** at this specific step. ☀️
- I : This is the **Identity Matrix**.
- **What it does:** This term dictates the shape of the Gaussian noise bubbles added to the image. The use of the **Identity Matrix (I)** means that the **noise added to each pixel is independent** (what happens to pixel A doesn't change the noise added to pixel B) and **isotropic** (the spread is the same in all directions).

Speckle noise is a granular, "salt-and-pepper" texture that degrades the quality of images produced by waves, such as **ultrasound** and **radar**.

Unlike the static "hiss" you might hear on a radio (which is additive noise), speckle noise is **multiplicative**, meaning it gets **worse as the signal gets stronger**. In an ultrasound, this means the **brighter the tissue appears, the more intense the speckled texture becomes**, making it difficult to see fine details like the edges of organs or small cysts.

Speckle noise reduces the **contrast resolution** of the image. It blurs the boundaries between different tissues, which is exactly the problem the research paper you uploaded tries to solve.

- **In your PDF:** The authors mention that ultrasound images "exhibit a suboptimal signal-to-noise ratio" and are "characterized by considerable speckle noise".
- **The Fix:** This is why the **DDIC** model in your paper uses a **median filter** during its reconstruction steps (see Section 2.2 of your PDF). The median filter smoothes out this high-frequency granular noise so the model can calculate a clean correlation between the Ultrasound and the MRI.

Summary Table: Additive vs. Speckle Noise

Feature	Additive Noise (e.g., Gaussian)	Speckle Noise
Nature	Added <i>on top</i> of the signal.	Multipled <i>by</i> the signal.
Brightness	Visible even in dark areas.	Mostly visible in bright areas; dark areas remain dark.
Cause	Electronic heat, sensor errors.	Wave interference (physics-based).
Removal	Easy (simple averaging).	Hard (requires complex filters like Lee or Median).

This is a great question. It comes down to the difference between describing a distribution (the shape) versus calculating a probability (a specific number).

The short answer is: The semicolon (;) is the key.

The notation $\mathcal{N}(x_t; \mu, \Sigma)$ is a shorthand way of writing the actual mathematical formula for the Gaussian curve evaluated at the specific point x_t .

1. The "Sampling" Notation (What you are thinking of)

If you just want to say "Variable X comes from this distribution," you use the tilde symbol (\sim) and omit the variable inside the function:

$$x_t \sim \mathcal{N}(\sqrt{1 - \beta_t} \cdot x_{t-1}, \beta_t I)$$

Meaning: " x_t is sampled from a Normal distribution with this mean and this variance."

↳ why it has written $x_{t-1}, \sqrt{1 - \beta_t}$ not $\sqrt{1 - \beta_t}?$
because for each pixel of x_t , it take
correspondent pixel in x_{t-1} and use scaled
down value of that pixel as mean of
distribution that it sampled of

2. The "Density Function" Notation (What the paper uses)

The equation in the paper starts with $q(x_t|x_{t-1}) = \dots$ This signifies that q is a function.

Specifically, it is a Probability Density Function (PDF). A function needs an input and an output.

- Input: A possible image x_t .
- Output: The probability (density) of that specific image occurring.

To define this function mathematically, they write:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \mu, \Sigma)$$

Meaning: "The probability density of observing the specific value x_t , given a Normal distribution defined by parameters μ and Σ ."

Why it matters for the math

If you wrote $q(x_t|x_{t-1}) = \mathcal{N}(\mu, \Sigma)$, strictly speaking, the Right Hand Side would be a constant object (a distribution), while the Left Hand Side is a function of x_t .

By putting x_t inside with a semicolon, they are telling you explicitly where to plug x_t into the Gaussian formula:

$$\mathcal{N}(x_t; \mu, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{(x_t - \mu)^2}{2\Sigma}}$$

So, the x_t is there to show exactly where the variable sits in the exponent.

1. Equation (4): The Recipe for Removing Noise

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right)$$

This equation calculates the **Mean (μ)** of the previous, cleaner image step (x_{t-1}). In simple terms, it is the formula the computer uses to decide: "Given this noisy image x_t , what should the slightly cleaner version look like?". ☺

- $\mu_\theta(x_t, t)$: This is the output. It is the estimated "center" or "average" pixel values for the cleaner image we want to build.
- x_t : The current noisy image you have.
- $\epsilon_\theta(x_t, t)$: **This is the most important part.** This represents the **Artificial Intelligence (Neural Network)**. ↗ θ means Neural network Parameters
 - The function ϵ_θ looks at the noisy image and predicts: "I think this specific part of the image is noise."
- **The Constants ($\alpha_t, \bar{\alpha}_t$)**: These are just the scaling numbers from the schedule we defined earlier.

The Intuition: The equation essentially says: To get the clean image mean (μ_θ), take the noisy image (x_t) and subtract the noise predicted by the AI (ϵ_θ), scaled by some math constants.

→ we train a model to predict the noise, and we subtract scaled version of it from noisy image
* function of ϵ_θ which get 2 variable of actual image and t

2. Equation (5): The Training (Loss) Function

$$L_t = \mathbb{E}_{t \sim [1, T], x_0, \epsilon_t} \left[\left\| \epsilon_t - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon_t, t) \right\|^2 \right]$$

This equation describes how the AI **learns**. It is the **"Loss Function"** that the model tries to minimize during training. ☺

- **The Goal:** We want the AI to be perfect at guessing what the noise is so it can subtract it in Equation (4).
- ϵ_t (**The Truth**): This is the **actual** random noise we added to the image. Since we are training, we created the noisy image ourselves, so we know exactly what noise was used.

we are training model
so we generate noise ϵ_t , and predict noise ϵ_θ

- $\epsilon_\theta(\dots)$ (**The Guess**): This is the noise the AI predicts is in the image.
 - Note: The complex term inside the parentheses— $\sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon_t$ —is just the mathematical formula for "the noisy image x_t " (from Eq 2 in the previous turn). It is the input given to the AI.
- $\|\dots\|^2$ (**The Score**): This subtracts the **Guess** from the **Truth** and squares the result.
 - If the AI guesses the noise perfectly, the result is 0 (Perfect score).
 - If the AI guesses wrong, the result is a large number (High error).

In equation 5:

The symbol \mathbb{E} stands for **Expectation** (or Expected Value).

In the context of training a neural network, you can think of this simply as the "**Average**". Since it is impossible to calculate the error for every possible image and every possible noise level in the universe, we calculate the "Expected" loss by averaging the error over a batch of random samples.

- x_0 (**The Image**):
 - **Meaning:** Sample a real data point.
 - **Action:** The computer reaches into your dataset (the "HC18" ultrasound dataset mentioned in the paper) and picks one real image at random. ↗
- ϵ_t (**The Noise**):
 - **Meaning:** Sample Gaussian noise.
 - **Action:** The computer generates a block of random static (Standard Normal Distribution $\mathcal{N}(0, I)$) with the same size as the image.
- $t \sim [1, T]$ (**The Time Step**):
 - **Meaning:** Sample a time step uniformly. ↗? *we don't pick specific noisy image them until first noisy image of one specific step and denoise to the previous one*
 - **Action:** The computer picks a random integer between 1 and T (where $T = 1000$ in this paper). ↗

3. In what order is t selected?

This is a critical point: During training, t is selected Randomly, not in order.

- **NOT Sequential:** The model does *not* go 1, 2, 3 ... 1000.
- **Random Sampling:** In one training step, the computer might pick $t = 500$ (medium noise). In the very next step, it might pick $t = 2$ (very little noise), and then $t = 999$ (pure noise).

Why random? The model needs to learn how to remove noise at **every** stage of the process. By picking t randomly, the neural network learns to recognize and remove "faint noise" (low t) and "heavy noise" (high t) simultaneously.

When the computer calculates this Loss (L_t) for one training step, it effectively follows these instructions:

1. **Roll a die** to pick a random noise level t (e.g., 450).
2. **Pick a card** to get a random image x_0 (a fetal brain).
3. **Generate static** ϵ_t .
4. **Combine them** to make a noisy image (using the formula inside the parenthesis).
5. **Ask the AI:** "Look at this noisy image. Guess what static I added?"
6. **Calculate Score:** Compare the AI's guess (ϵ_θ) to the actual static (ϵ_t) and square the difference.

* all Process
is described
here

DDLB explanation

utilizing $\text{ForwardODE}^{\text{domain1}}$ to noise the first domain image into a latent code and then denoise the latent into the second domain image using $\text{BackwardODE}^{\text{domain2}}$.

This method proves efficacious in scenarios where translation is required between objects sharing fundamental structural characteristics and global features, but fails in the translation of objects with finer details. However, in the context of medical imaging translation, wherein even minor details hold significance for clinical diagnosis, the preservation of the precise composition of the source image becomes paramount. Hence, we suggest the Dual Diffusion Imposed Correlation method (DDIC), in which the backward process of the DDIB method mandates correlation between the restored two domains as schematically shown in Figure 1.

ddm is deterministic
version of ddpm

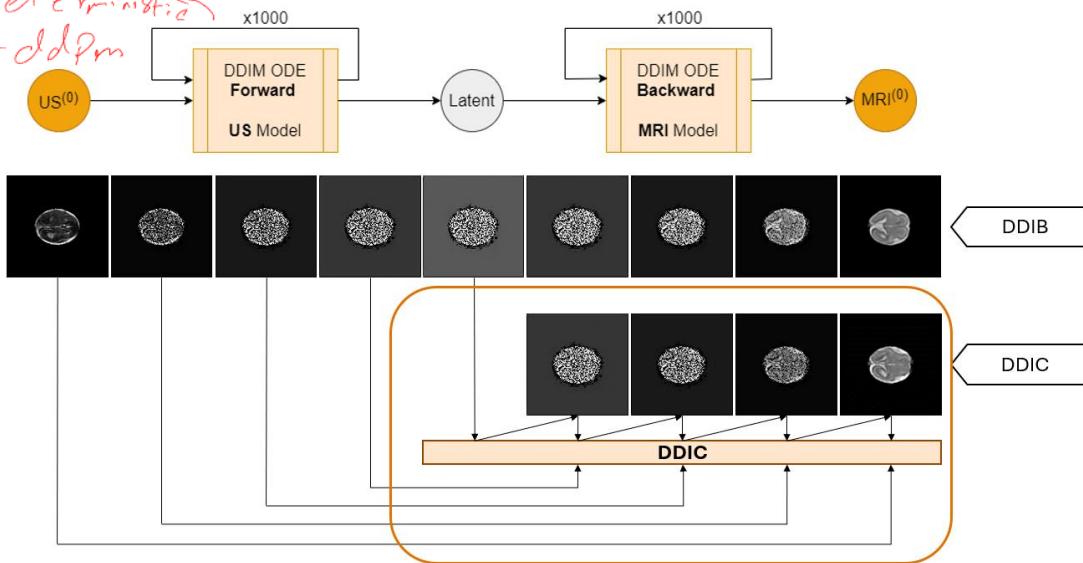


Figure 1: **Top** schematic depiction of the DDIB algorithm diagram and exemplary test sample. **Bottom** schematic depiction of the DDIC. The DDIC diagram is shown in detail in Figure 2.

DDLB is not proposed by this paper, but DDIC is

2.2 The DDIC model proposed which include line

Initially, the US image X undergoes denoising through T steps with $\text{ForwardODE}^{\text{US}}$ utilization, resulting in the extraction of the US latent code. The reconstruction of the pseudo-MRI image from the US latent code is achieved gradually. In each timestep t , the model uses $\text{BackwardODE}^{\text{MRI}}$ to get primary denoised MRI image and use $\text{BackwardODE}^{\text{US}}$ from the same latent code to get the corresponding parallel US reconstruction. Subsequently, both images undergo a median filtering process, in order to overcome the speckle noise. Next, the correlation between the two median filtered images is calculated, receiving the loss function:

$$\text{loss} = -1 \cdot \text{CorrCoef}(\tilde{X}_{t-1}, \tilde{Y}_{t-1}), \quad \text{Eq. (10)}$$

where $\tilde{X}_{t-1}, \tilde{Y}_{t-1}$ are the reconstructed US and MRI images after passing the median filter respectively, and the correlation CorrCoef is given by:

$$\text{CorrCoef}(a, b) = \frac{\text{cov}(a, b)}{\sqrt{\text{cov}(a, a) \cdot \text{cov}(b, b)}}, \quad \text{Eq. (11)}$$

where $\text{cov}(a, b)$ is the covariance between images a and b .

Our conjecture posits that the efficacy of this approach arises from the median filter's inherent smoothing effect, which enables correlation analysis to emphasize crucial image features rather than being influenced by the intrinsic noise characteristic of the ultrasound imagery.

Following, optimization of Y_t , the initial MRI image for the $\text{BackwardODE}^{\text{MRI}}$, is conducted by calculating the derivative of the loss function according to Y_t , adopting the methodology outlined in detail in Parmar et al. (20), which has demonstrated efficacy with maximum cross-attention optimization. This derivation is accomplished via a single-step gradient descent employing an optimal step size selected based on the most favorable outcomes observed during testing. The reconstruction overall consists of $T = 1000$ time steps, the algorithm for each step is illustrated in Figure 2.

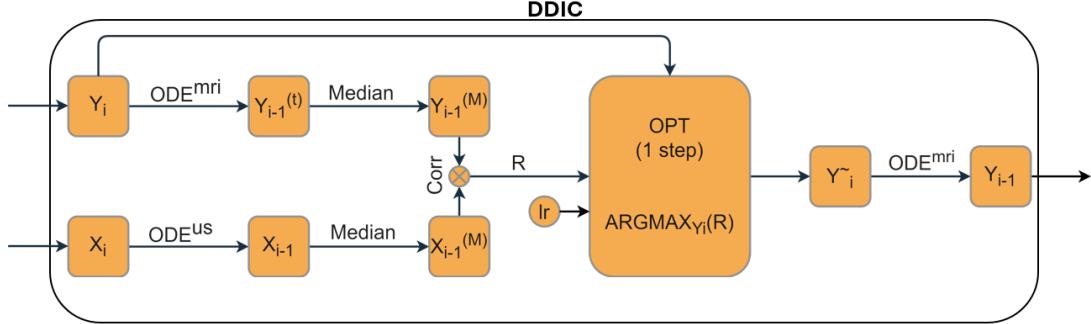


Figure 2: A block diagram of a single DDIC step.

2.3 Datasets

The chosen two-dimensional (2D) anatomical cross-section for the US-MRI translation in this study was the "transthalamic" axial plane. This plane serves as a key monitoring tool for assessing fetal development through pregnancy, hence, it is of high clinical importance (21). The gestational age range was chosen to be within 21-38 weeks due to the common overlap between US and MRI examinations during these weeks.

The US images were sourced from the "Automated measurement of fetal Head Circumference" (HC18) dataset (22), encompassing 1000 training images with dimensions of 800×540 pixels, with a pixel size range of $0.049\text{-}0.393 \text{ mm}^2$. The training set also includes an image with the manual annotation of the head circumference (HC) and its measurement in millimeters for each HC, which was outlined by a trained sonographer. To adhere to the designated gestational age range, only images with HC falling within the range of 170 mm to 350 mm were selected from the training set (23). After excluding images that were either outside the above range or unclear, the final refined US dataset comprised 365 images.

The MRI data employed in this study were downloaded from two distinct datasets: the "CRL fetal brain atlas" database (24) and the "FeTA" challenge dataset (25). The former consists of 18 three-dimensional (3D) T2-weighted scans, with dimensions of $135 \times 189 \times 155$ pixels, pixel size of 0.8 mm^3 , and covering a gestational age range of 21 to 38 weeks. The latter encompassed 80 3D T2-weighted scans, with dimensions of $256 \times 256 \times 256$ pixels, pixel size of $\sim 0.5 \text{ mm}^3$, and gestational age spanning from 20 to 35 weeks. In both datasets, several slices around and including the transthalamic axial plane were extracted, resulting in a combined MRI dataset composed of 251 images.

Due to limitations in computing power, as a proof of concept, all images (US and MRI) used in this study were preprocessed to yield a size of 128×128 pixels with a pixel size of 1.094 mm^2 . Fetal heads were segmented based on the skull circumference in the US scans and brain tissue in the MRI scans. All images were centered and rotated so that the third ventricle is horizontal in the frame and the lateral ventricles are on the left side.

2.4 Evaluation protocols

Importantly, it should be noted that ground truth fetal US-MR exactly paired data are unavailable, as the images from those modalities are not perfectly aligned. This is due to the inability to perform ultrasound and MRI scans simultaneously in the womb. Even when scans are conducted within a short timeframe, fetal movement often results in positional changes, leading to misalignment between slices from the two imaging modalities (26). Therefore, we compared the US images to the pseudo-MRI images using different approaches.

The comparison of image interpretability is inherently subjective, and conventional evaluation metrics, such as Structural Similarity (SSIM), commonly employed in image reconstruction quality analysis,

become irrelevant in the context of unpaired US-MRI data. Given this circumstance, we chose the adoption of alternative metrics, specifically the Mutual Information (MI) metric, the Peak Signal-to-Noise Ratio (PSNR) metric, and the Fréchet Inception Distance (FID) for the quantitative evaluation in our study.

MI serves as a metric for image alignment, obviating the necessity for identical signal characteristics in the compared images. It quantifies the predictive capacity regarding the signal in the second image based on the signal intensity observed in the first, thereby offering an assessment of the degree of concordance between the two images. The MI is defined as(27):

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} P_{(X,Y)}(x, y) \log \left(\frac{P_{(X,Y)}(x, y)}{P_X(x)P_Y(y)} \right),$$

Eq. (12)

where P_X and P_Y are the marginal probability mass function of X and Y respectively, and $P_{(X,Y)}$ is the joint probability mass function of X and Y.

The PSNR is defined as (28):

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right),$$

Eq. (13)

where MAX_I is the maximum possible pixel value of the image, 255 in our case, and MSE is the mean square error between the two images.

The FID is a metric employed to evaluate the quality of images generated by a generative model. It compares the distribution of generated synthetic images to that of a set of real images. A lower FID score indicates a greater similarity between the two distributions, with a score of zero signifying identical distributions. In this work, we employed the FID between the set of the generated pseudo-MRI to the set of the source US images (29).

Another evaluation performed is the Contrast-to-Noise Ratio (CNR). This metric was taken to assess the improved contrast of the generated MRI images relative to the original US images. The CNR is defined as:(1)

$$CNR = \frac{|\bar{I}_{ROI} - \bar{I}_{background}|}{\sqrt{\sigma_{ROI}^2 + \sigma_{background}^2}}$$

Eq. (14)

Where \bar{I}_{ROI} and $\bar{I}_{background}$ are the average intensity in a region of interest (ROI) and its background respectively, and σ_{ROI}^2 and $\sigma_{background}^2$ are their standard deviation. The ROI selected for this evaluation is the distal lateral ventricle.

In addition to these quantitative methods, semi-qualitative evaluation was applied. For that aim, we have utilized the open-source "Segment Anything" algorithm developed by Meta AI (30). The algorithm was applied to the test set images, where the lateral ventricles and the Sylvian fissure were segmented. The segmentation process aims to demonstrate the practical benefits of converting US images into MRI images. It was done on both full-resolution and low-resolution US images and the DDIC-generated pseudo-MRI. The quality of segmentation was visually examined.

Finally, we assessed the clinical benefits of translating images from US to pseudo-MRI qualitatively using the Medical Opinion Test (MOT). A user study was conducted to evaluate MOT performance. Five physicians were each given a set of 40 pairs of US images and their corresponding translated pseudo-MRI images. For each pair, the physicians were asked to indicate whether the translated image provided an advantage in terms of clinical diagnosis based on several parameters. These parameters included: the proximal lateral ventricle, the distal lateral ventricle, the proximal Sylvian fissure, and the distal Sylvian fissure.

2.5 Training parameters

The US dataset was divided into a training set, comprising 90% of the data, and a test set, comprising of the remaining 10%. The training set for the MRI model was the complete MRI dataset. The network architecture was based on the framework provided by DDIB with the addition of the DDIC module. The diffusion model hyperparameters were set to timestep 1000, cosine beta scheduler, and ϵ_t -prediction. For the DDIC reconstruction, the gradient step size was set to $lr = 3$ (see Figure 2). The model was trained on an NVIDIA RTX 2080 GPU running under Linux. Training computation time was approximately 48 hours for each model.

3 Results

To evaluate the performance of the DDIC model, we have used the 10% test images taken from the US dataset to synthesize pseudo-MRI images. Additionally, DDIB and CycleGAN, were also applied to generate pseudo-MRI images, providing reference outputs for comparative analysis. Figure 3 illustrates a collection of representative images transformed from US to pseudo-MRI by the three methods. Notably, the pseudo-MRI exhibits enhanced visual clarity regarding brain structures, characterized by sharper delineation of borders and improved contrast. It is noteworthy that the algorithm effectively mitigates acoustic shadow artifacts, commonly appearing in ultrasonic images, especially when evaluating the size of the proximal lateral ventricle in cases suspected of dilated cerebral ventricles in late gestation.

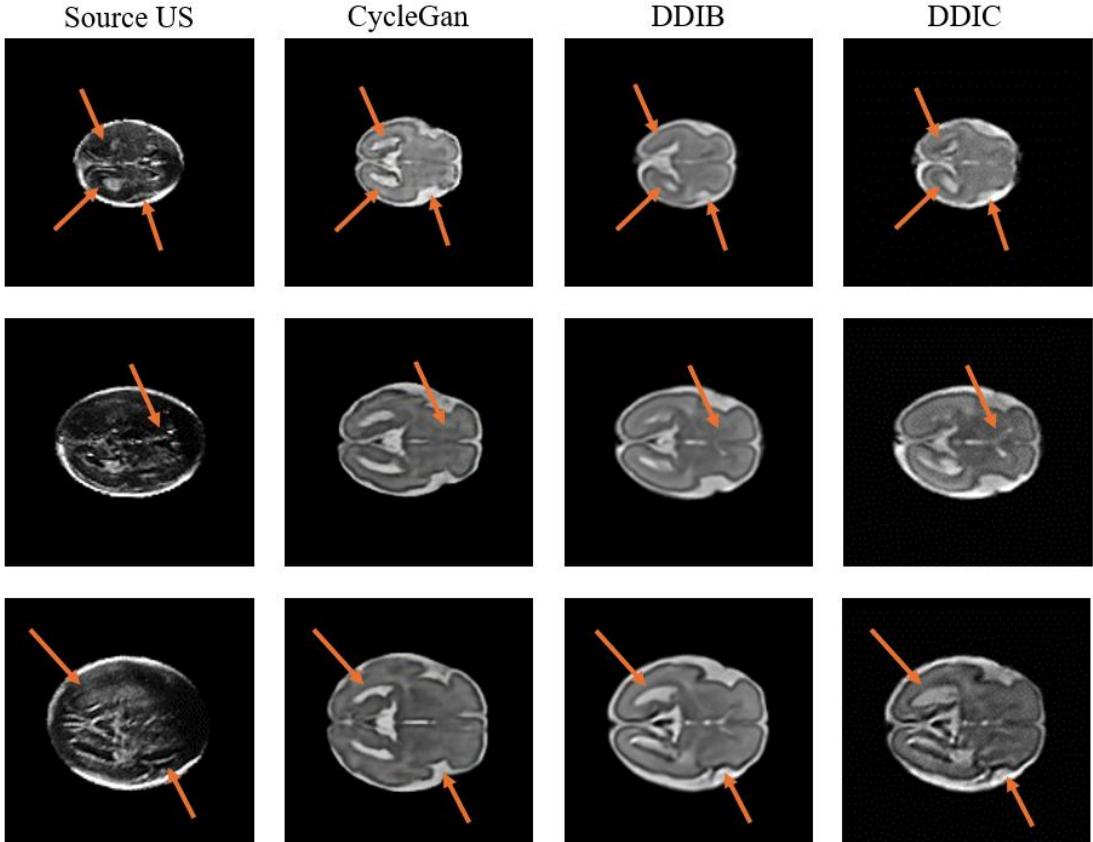


Figure 3: Qualitative evaluation of US to pseudo-MRI translation. Each row presents an exemplary sample from a different gestational age. (1st column) original US image with head segmentation. (2nd column) CycleGAN pseudo-MRI synthesis. (3rd column) DDIB pseudo-MRI synthesis. (4th column) DDIC pseudo-MRI synthesis. The arrows indicate zones of inconsistency between the pseudo-MRI reconstructions and the original US. Note the improved performance of the DDIC.

In comparison to CycleGAN and DDIB, the DDIC successfully preserves fine details between the original US image and its translated counterpart, as demonstrated in Figure 3. The arrows delineate regions wherein the DDIC algorithm exhibits enhancements in illustrating fine components of the fetal

brain structures. In the first row, the bottom right arrow highlights the insula region, wherein CycleGAN and DDIB methods exhibit distortion in the sulcation. In contrast, the DDIC method preserves the original US scan insula structure. The two other arrows indicate the lateral ventricles. CycleGAN failed to maintain the structure of the lateral ventricles, leading to their enlargement and elongation, wrongly indicative of a more advanced gestational stage compared to the original ultrasound image. Similarly, DDIB failed to preserve the distinctiveness of the lateral ventricles, amalgamating them with the flax midline. In contrast, the DDIC algorithm successfully separated the lateral ventricles from the flax midline and preserved their original anatomical structure. In the second row, the arrow denotes the Cavum Septum Pellucidum (CSP). As can be observed, the DDIC algorithm exhibits the highest level of sharpness and distinction. The top arrow in the third row indicates the left lateral ventricle. DDIC preserves the size and shape of the ventricle as depicted in the US image, as opposed to the other methods. The lower arrow signifies the maintenance of the insula sulcation, consistent with the depiction provided by the images in the first row.

The quantitative assessment is graphically depicted in Figure 4, wherein the evaluation metrics of MI, PSNR, and FID are presented. As can be observed, the DDIC yields the best results in all categories. These findings indicate that the performance of the proposed method for translating US to pseudo-MRI surpasses that of competing architectures.

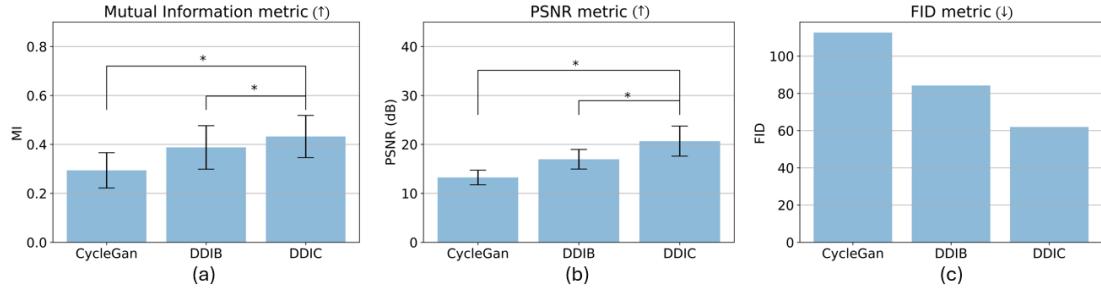


Figure 4: (a) Distributions of Mutual Information metric; (b) Distributions of PSNR metric; (c) Distributions of FID metric. *Indicates statistically significant two-sided P value < 0.001 with the unequal variances t-test.

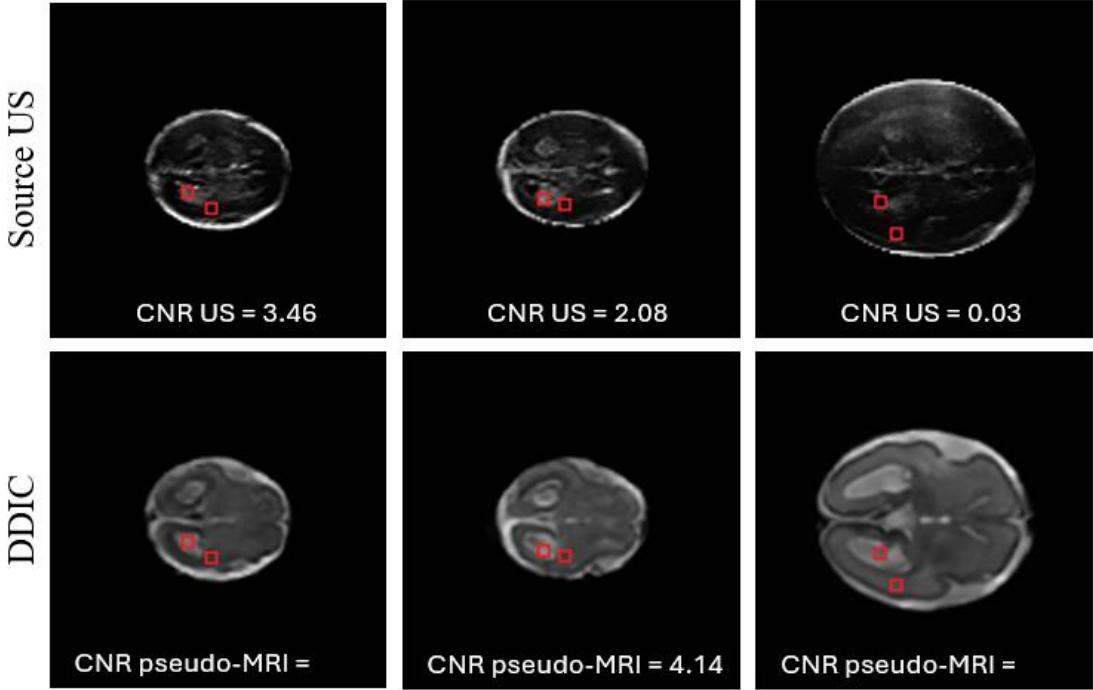


Figure 5: Three examples of the CNR zones selection. The US image is on the left, and the DDIC pseudo-MRI is on the right. The upper square on the lateral ventricle is the region of interest (ROI), and the lower square is the background for the CNR calculation.

It is noteworthy to mention our attempt to implement the methodology suggested by Jiao *et al.* (7), which translates fetal brain US to pseudo-MRI with a GAN-based model. However, we obtained meaningless pseudo-MRI images. We postulate that this outcome may be attributed to the comparatively limited size of the datasets available in our study in contrast to the extensive datasets employed by Jiao *et al.*

To demonstrate the enhancement in contrast of the pseudo-MRI images relative to the original ultrasound images, we computed the CNR as outlined in the method section. Representative contrast regions are illustrated in Figure 5. The results indicate that the contrast in the DDIC pseudo-MRI images is improved, from 1.37 ± 1.24 in the US images to 2.61 ± 1.75 in the pseudo-MRI images. The observed standard deviation can be attributed to variations in gestational age.

To illustrate the visual advantages of the DDIC-generated pseudo-MRI images compared to the original US images, we applied the "Segment Anything" algorithm on the test set images as described in the methods section. The pseudo-MRI images consistently yielded more reliable segmentation than both the full-resolution and low-resolution US images. Exemplary results are shown in Figure 7. In each of the presented examples, the segmentation of the pseudo-MRI images, compared to the US images, is noticeably smoother, free of irregularities, and successfully captures the entire target surface, including the contours of the lateral ventricles and the Sylvian fissure.

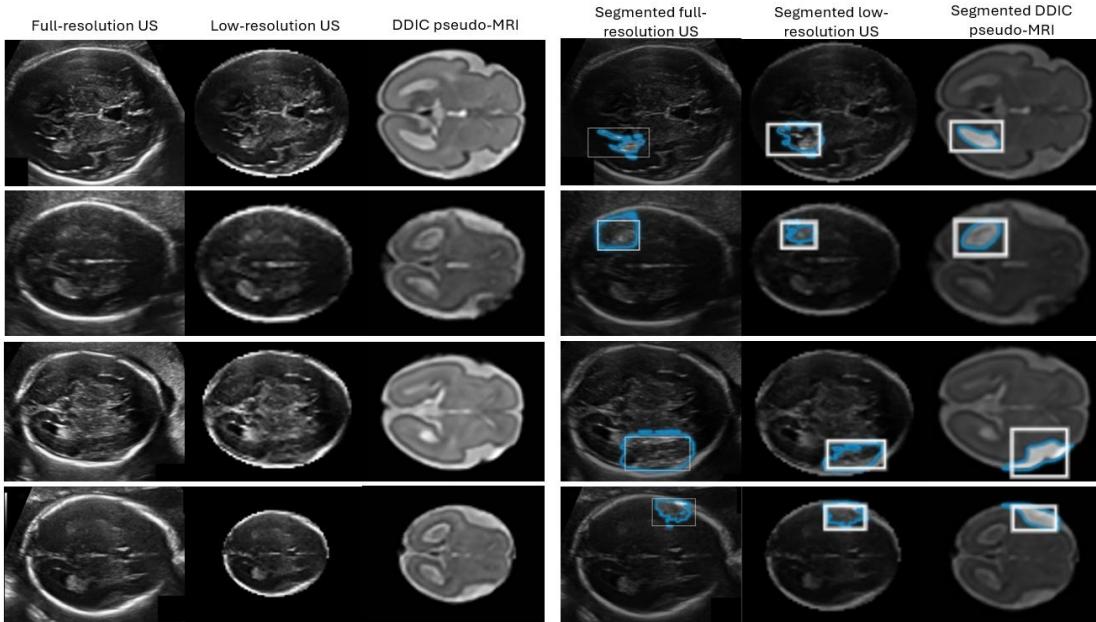


Figure 6: An example of Meta AI's "Segment Anything" algorithm demonstrates its application in segmenting anatomical features on three types of images: The original full-resolution US image as taken from the HC18 dataset, a low-resolution US image as preprocessed for data input to the DDIC algorithm, and the DDIC pseudo-MRI image. (1st row) Distal lateral ventricle segmentation. (2nd row) Proximal lateral ventricle segmentation. (3rd row) Distal Sylvian fissure segmentation, (4th row) Proximal Sylvian fissure segmentation. As can be observed, the segmentations in the pseudo-MRI images are noticeably smoother and more distinct compared to the segmentations of the original US images (both full and low resolution).

To validate the clinical value of the image translation, we conducted the MOT test as described in the methods section. The results are graphically presented in Figure 6. The findings reveal that, on average, the physicians reported an improvement in clinical information in approximately 81% of the evaluated images (see Figure 6a). The most significant improvement occurred in the proximal Sylvian fissure. Notably, there is a more significant enhancement in the proximal area, particularly concerning the lateral ventricle and the Sylvian fissure (see Figure 6b). These regions, typically not visible in ultrasound images (31), show marked improvement, potentially facilitating more accurate diagnoses.

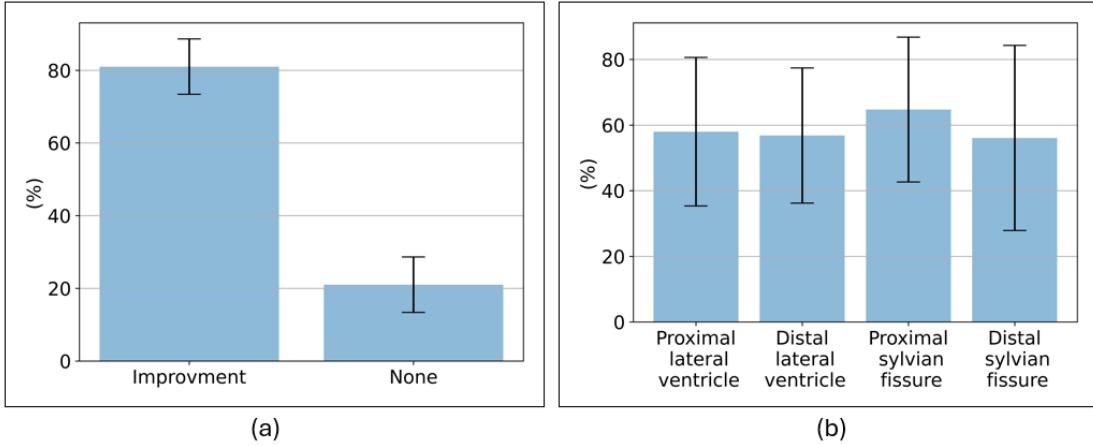


Figure 7: Graphical depiction of the MOT score. (a) Percentage of images in which the physicians have found improvement in at least one feature. (b) The percentage of images from the improved group in which each specific feature was improved. Notably, there is a significant enhancement observed in all features relative to the original US.

4 Discussions

Ultrasonic fetal images pose a challenge due to their inherent low image quality which is characterized by acoustic shadows and speckled texture. These obstructive factors are more prominent in the proximal zone and may be crucial when suspected anomalies may be present. In this study, we have demonstrated that the clinical merit of the acquired US images can be augmented by implementing image translation into pseudo-MRI display using the DDIC method. As observed, 81% of the tested translated images were improved according to the MOT test.

The primary advantage of the DDIC method lies in the improved visibility of the lateral ventricles and the Sylvian fissure (as shown in Figure 6b), potentially leading to a better fetal development diagnosis. This is achieved by the integrated optimization which imposes the correlation between the resultant pseudo-MRI image and the original ultrasound image. This correlation facilitates similarity in the fine details within the brain structure, leading to a more straightforward clinical interpretation.

Furthermore, as can be observed from the results, DDIC has improved image quality in all the quantitative metrics studied here (Figs. 3, 4, 5). Notably, the CNR measured for the distal ventricle has almost doubled in the generated images compared to the ultrasonic source.

Importantly, the Mutual Information (MI) metric was better relative to the reference translation methods applied in this study. The MI metric for translation quality evaluation allows for measuring the structural similarity between the US and MRI images without necessitating signal similarity, thereby facilitating direct comparison. This is very important since MRI and US have opposing depictions of certain anatomical features, for example, the ventricles appear black in the US but white in the MRI. In addition, the Fréchet inception distance (FID) which is a metric used to assess the quality of images created by a generative model, was also superior (smaller).

The DDIC algorithm offers a notable advancement in the application of segmentation algorithms for identifying brain anatomy, demonstrating superior performance on synthesized pseudo-MRI images compared to the source ultrasound images. In particular, identifying the lateral ventricles and the Sylvian fissure which are more distinctly visible in the pseudo-MRI images, exhibiting enhanced contrast and sharper boundaries, as confirmed by the MOT test. Figure 7 provides examples of the "Segment Anything" algorithm developed by Meta AI (30), which achieves significantly improved segmentations of the lateral ventricles and the Sylvian fissure in the pseudo-MRI images compared to the ultrasound images. These segmentations hold significant clinical value. As noted in the introduction, identifying the shape of the Sylvian fissure and evaluating the size of the lateral ventricles are crucial for diagnosing cortical dysplasia and fetal cerebral lateral ventriculomegaly respectively (17,31). We have demonstrated the application of the algorithm on both the original full-resolution ultrasound image and the low-resolution preprocessed ultrasound image to highlight the advantage of the DDIC even though it reduces the original resolution of the ultrasound images.

Another advantage of the DDIC method stems from the fact that it performs separate training for the ultrasound and MRI image databases. This separated training approach offers the potential to utilize databases taken from various sources. It enables training the US model in one hospital and the MRI model in another, eliminating the need for patient images to leave hospital premises and maintaining patient confidentiality. Moreover, in contrast to GAN-based models, employing a diffusion model for image translation enables the learning process to generate new images effectively even with a relatively small dataset. This capability represents a significant advantage, especially in the medical domain where access to extensive databases is often restricted.

Another point worth mentioning is the limited availability of databases containing fetal MRI images. This stems from the fact that unlike US, fetal MRI examinations are not routinely preformed. Hence presumably, US-MRI translation presents an opportunity to generate fetal pseudo-MRI datasets from widely accessible ultrasound databases. This process can facilitate the training of MRI-based models for various medical applications. Moreover, image translation can aid in the development of registration algorithms between US and MRI modalities (32). Such algorithms necessitate the identification of salient points present in both images. When the images are displayed in the same format, the process of identifying these points becomes more facile.

5 Conclusion

This study introduces a technique for translating fetal US images into a pseudo-MRI display, aimed at improving the visualization of fetal brain anatomical structures. Notably, the resulting pseudo-MRI Sylvian fissure, characterized by well-defined borders and significantly improved contrast clarity according to the MOT. This enhancement holds promise for expediting and enhancing the accuracy of fetal growth diagnosis, potentially leading to improved clinical outcomes.

Acknowledgments

This work was supported in part by the Ministry of Innovation, Science and Technology, Israel, under Grant #880011 and in part by the Technion Israel Institute of Technology scholarship.

Reference

1. Azhari H, Kennedy JA, Weiss N, Volokh L. From Signals to Image A Basic Course on Medical Imaging for Engineers.
2. Grover VPB, Tognarelli JM, Crossey MME, Cox JJ, Taylor-Robinson SD, McPhail MJW. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. Vol. 5, Journal of Clinical and Experimental Hepatology. Elsevier B.V.; 2015. p. 246–55.
3. Zhu JY, Park T, Isola P, Efros AA, Research BA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Monet Photos. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2223-2232.
4. Richardson E, Alaluf Y, Patashnik O, Nitzan Y, Azar Y, Shapiro S, et al. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 2287-2296.
5. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8789-8797.
6. Armanious K, Jiang C, Abdulatif S, Küstner T, Gatidis S, Yang B. Unsupervised medical image translation using Cycle-MeDGAN. In: European Signal Processing Conference. European Signal Processing Conference, EUSIPCO; 2019.
7. Jiao J, Namburete AIL, Papageorghiou AT, Noble JA. Self-Supervised Ultrasound to MRI Fetal Brain Image Synthesis. IEEE Trans Med Imaging. 2020 Dec 1;39(12):4413–24.

8. Vukovic D, Ruvinov I, Antico M, Steffens M, Fontanarosa D. Automatic GAN-based MRI volume synthesis from US volumes: a proof of concept investigation. *Sci Rep.* 2023 Dec 1;13(1).
9. Greenspan H, Madabhushi A, Mousavi P, Salcudean S, Duncan J, Syeda-Mahmood T, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 [Internet]. Cham: Springer Nature Switzerland; 2023. (Lecture Notes in Computer Science; vol. 14229). Available from: <https://link.springer.com/10.1007/978-3-031-43999-5>
10. Song Y, Ermon S. Generative Modeling by Estimating Gradients of the Data Distribution. In Advances in Neural Information Processing Systems, 2019, Vol. 32.
11. Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models. In Advances in Neural Information Processing Systems, 2020, Vol. 33. 6840–6851.
12. Zhang Q, Tao M, Chen Y. gDDIM: Generalized denoising diffusion implicit models. 2022 Jun 11; Available from: <http://arxiv.org/abs/2206.05564>
13. Dhariwal P, Nichol A. Diffusion Models Beat GANs on Image Synthesis. 2021 May 11; Available from: <http://arxiv.org/abs/2105.05233>
14. Melchiorre K, Bhide A, Gika AD, Pilu G, Papageorghiou AT. Counseling in isolated mild fetal ventriculomegaly. Vol. 34, Ultrasound in Obstetrics and Gynecology. John Wiley and Sons Ltd; 2009. p. 212–24.
15. Malinge G, Paladini D, Haratz KK, Monteagudo A, Pilu GL, Timor-Tritsch IE. ISUOG Practice Guidelines (updated): sonographic examination of the fetal central nervous system. Part 1: performance of screening examination and indications for targeted neurosonography. *Ultrasound in Obstetrics and Gynecology.* 2020 Sep 1;56(3):476–84.
16. Leitner Y, Goez H, Gull I, Mesterman R, Weiner E, Jaffa A, et al. Antenatal Diagnosis of Central Nervous System Anomalies: Can We Predict Prognosis? *Journal of Child Neurology.* 2004;19(6):435–438. doi:10.1177/088307380401900607
17. Guiraud L, Selleret L, Larroche JC, Buenerd A, Alias F, Gaucherand P, et al. Abnormal Sylvian fissure on prenatal cerebral imaging: Significance and correlation with neuropathological and postnatal data. *Ultrasound in Obstetrics and Gynecology.* 2008 Jul;32(1):50–60.
18. Song J, Meng C, Ermon S. Denoising Diffusion Implicit Models. 2020 Oct 6; Available from: <http://arxiv.org/abs/2010.02502>
19. Su X, Song J, Meng C, Ermon S. Dual Diffusion Implicit Bridges for Image-to-Image Translation. 2022 Mar 16; Available from: <http://arxiv.org/abs/2203.08382>
20. Parmar G, Singh KK, Zhang R, Li Y, Lu J, Zhu JY. Zero-shot Image-to-Image Translation. 2023 Feb 6; Available from: <http://arxiv.org/abs/2302.03027>
21. Milani HJF, de Sá Barreto EQ, Araujo Júnior E, Peixoto AB, Nardozza LMM, Moron AF. Ultrasonographic evaluation of the fetal central nervous system: Review of guidelines. Vol. 52, Radiologia Brasileira. Colegio Brasileiro de Radiologia; 2019. p. 176–81.
22. van den Heuvel TLA, de Bruijn D, de Korte CL, van Ginneken B. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS One.* 2018 Aug 1;13(8).
23. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Ismail LC, Lambert A, et al. International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *The Lancet.* 2014;384(9946):869–79.

24. Gholipour A, Limperopoulos C, Clancy S, Clouchoux C, Akhondi-Asl A, Estroff JA, et al. LNCS 8674 - Construction of a Deformable Spatiotemporal MRI Atlas of the Fetal Brain: Evaluation of Similarity Metrics and Deformation Models. 2014.
25. Payette K, de Dumast P, Kebiri H, Ezhov I, Paetzold JC, Shit S, et al. An automatic multi-tissue human fetal brain segmentation benchmark using the Fetal Tissue Annotation Dataset. *Sci Data*. 2021 Dec 1;8(1).
26. Jiao J, Namburete AIL, Papageorghiou AT, Noble JA. Self-Supervised Ultrasound to MRI Fetal Brain Image Synthesis. *IEEE Trans Med Imaging*. 2020 Dec 1;39(12):4413–24.
27. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*. 2004;69(6):16.
28. Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. In: Proceedings - International Conference on Pattern Recognition. 2010. p. 2366–9.
29. Yu Y, Zhang W, Deng Y. Frechet Inception Distance (FID) for Evaluating GANs Beijing, China:China Univ. Mining Technol. Beijing Graduate School, 2021.
30. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment Anything. 2023 Apr 5; Available from: <http://arxiv.org/abs/2304.02643>
31. Ganor Paz Y, Levinsky D, Rosen H, Barzilay E. Feasibility of Fetal Proximal Lateral Cerebral Ventricle Measurement. *Journal of Ultrasound in Medicine*. 2022 Dec 1;41(12):2933–8.
32. Fuerst B, Wein W, Müller M, Navab N. Automatic ultrasound-MRI registration for neurosurgery using the 2D and 3D LC2 Metric. *Med Image Anal*. 2014;18(8):1312–9.