

Evaluation of Early Gestational Age Fetal Brain Localisation Using Deep Learning

Chанго Бонг

Abstract—Fetal MRI provides valuable information for medical image analysis. However, analysis of MRI images of fetuses during gestation is challenging due to their constant motion, small size, and difficult positioning. In previous studies, neural network training has shown potential for solving various processing tasks for fetal MRI analysis. However, fetal imaging studies in early gestational age (GA) < 20 weeks are scarce and more challenging than conventional fetal brain MRI. Our study aims to provide an evaluation for localising the early gestational age fetal brain in MRI stacks affected by motion. Based on our analysis, we have observed that 3D DenseNET is the most suitable neural network for classification, achieving an impressive accuracy rate of 0.89. In terms of 3D brain segmentation, 3D AttentionUNet surpassed other models and demonstrated the highest level of similarity to the ground truth. Further research is required as the 2D Dynamic UNET, which was trained on a distinct dataset, had a low score during the evaluation and produced subpar 3D models. Our study indicates the feasibility of utilising 3D classification and 3D segmentation neural network techniques to analyse early GA fetal brain localisation, surpassing the performance of 2D neural networks.

I. INTRODUCTION

Magnetic resonance imaging, which is commonly referred to as MRI, is an advanced and highly effective technique that has recently been employed to capture images of fetal anomalies. However, Fetal imaging studies during the early GA were less common than regular fetal MRI scans conventionally performed after 28 weeks. Furthermore, the early GA fetal brain analysis was a daunting task due to their small size, constant movement, and varying position [1]. Over the past decade, various techniques have been used to process fetal MRI datasets for motion correction, classification and segmentation.

Single Shot Turbo Spin Echo (SSTSE) sequence has proven to be highly effective in reducing motion artefacts. However, this technique had limitations as The loss of structural continuity between slices and corruption of 3D volumetric information in 3D stacks is caused by fetal and maternal motion that occurs between slices.

To overcome this challenge, slice-to-volume registration (SVR) tools have been employed to reconstruct high-resolution 3D images of the fetal brain [2]. This technique has been instrumental in aiding the diagnosis of fetal anomalies, thereby allowing for more accurate and timely treatment. Moreover, A method called deformable SVR (DSVR) [3] has been

proposed as a solution for correcting non-rigid motion and reconstruction of the fetal trunk for diagnosis [4].

However, the SVR methodology based on classical registration cannot resolve large ($> 45 - 90^\circ$) rotations and translations of the fetal trunk [5]. Therefore, acceptable reconstruction quality can only be achieved for fetuses from the > 28 weeks gestational age range.

During the early gestational age, there is intra-uterine space within the uterus which can result in significant movements and rotations of the fetus. This can cause issues with the reconstruction of the thorax, as shown in Figure 1 where the fetus experienced large rotations and translations between the stacks.

Spatial transformer convolutional neural networks (CNN) have been used to overcome a limitation in imaging the fetal brain. This involves reorienting individual 2D slices to the standard radiological atlas space before reconstruction, as well as reconstructing already existing 3D volumes [6].

In this study, we evaluate various Deep Learning network for accurately localising the early GA (20 ~ 34 weeks) fetal brain MRI stacks affected by motion. Localising is an important initial step in most pipelines for analysing the fetal brain since it provides the basis for further processing.

Localising the early GA fetal brain is constructed as classification of the early GA of the fetal brain and segmentation of 3D volumes of the early GA of fetal brain. In classification, utilise different Deep Learning networks to classify fetal MRI stacks based on the GA into three categories: early, median, and last term. In Segmentation, different Deep Learning networks are used to segment fetal brain MRI stacks affected by motion, with the addition of various augmentations to create a 3D model.

II. RELATED WORKS

The previous studies that are similar to our study are presented in Table I. In 2023, an algorithm was proposed by Huang X. et al. [7]. for automatically segmenting fetal brain MRIs. The purpose of this algorithm is to diagnose congenital diseases or assess the impact of an intervention or treatment. In 2021, a group of researchers led by Dou H [8]. created a sophisticated neural network that utilises mixed kernel convolutions and attentive mechanisms to perform automatic segmentation of the cortical plate in fetal MRI scans. Axiomatic definition of label-set loss functions for training deep learning models with partially segmented

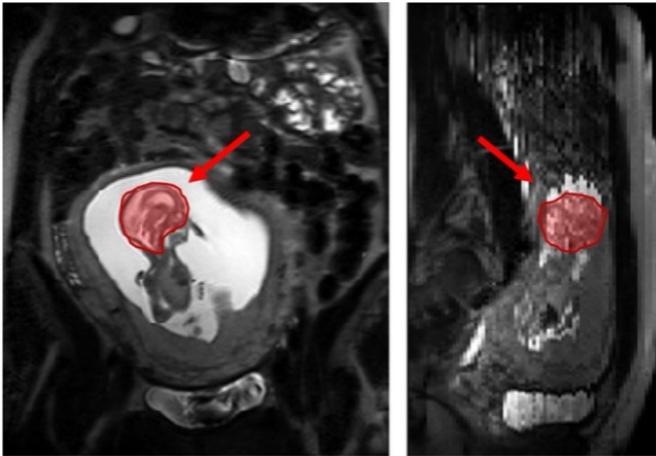


Fig. 1. An example of early GA (19 weeks) T2w MRI stack

images was presented by Fidon L. et al [9] in the same year. In 2023, Karimi D [10]. and colleagues discovered that the majority of teams developing segmentation techniques rely on UNET. They also observed that 3D UNET outperforms its 2D counterpart based on evaluation metrics. In 2018, a team led by Salehi S [11]. created a sophisticated neural network utilising mixed kernel convolutions to automatically segment the cortical plate in fetal MRI scans. In 2020, a deep learning framework was proposed by Li J et al. [12]. that is both effective and efficient in automatically extracting the fetal brain from fetal MRI. A group of researchers, led by Keraudren K. [13], presented a new technique in 2014 that utilises Random Forests and steerable features for the automatic identification of the heart, lungs, and liver in fetal MRI scans. In 2020, a framework was introduced by Ebner M. et al [14]. which is completely automated and accessible to the public. This framework is used for the purpose of localizing, segmenting, and reconstructing fetal brain MRI with super-resolution. Uus A. et al [3]. developed a new protocol for automated parcellation of 10 fetal body organ regions in 2022, which was used to train a neural network for multi-label segmentation. An evaluation conducted by Xie N H. et al. [15] in 2020 explored the possibility of utilizing deep learning algorithms based on CNNs to categorize fetal brain images as either normal or abnormal in the standard axial neurosonographic (SAN) planes. In 2018, Huang R. et al [16]. introduced VP-Nets, an efficient 3D detection approach for fetal neurosonography.

III. METHODS

A. Overview of methods

To provide an overview of our experiment process, we have included a comprehensive summary in Figure2. Our experiment begins with the preprocessing of datasets, which are then separated into training and validation datasets. For classification, we utilise 3D Deep Learning classification networks and evaluate their performance using recall, accuracy, precision, and F1-score. To ensure successful segmentation, we augment the datasets using 8 different types of augmentation techniques

in MONAI and train the data using 3D Deep Learning segmentation networks. The evaluation of segmentation performance is based on Dice score, IoU, Hausdorff Distance, and centroid Distance. Finally, we compare the 3D modelling results of each network to determine their effectiveness.

B. Data and Pre-processing

1) Fetal MRI images: This study employed fetal MRI data acquired at St.Thomas' Hospital, London under the "Individualised Risk prediction of adverse neonatal outcome in pregnancies that deliver preterm using advanced MRI techniques and machine learning" study (REC 21/SS/0082), (REC), "Placenta Imaging Project" (REC: 16/LO/1573) including 28 early GA datasets of subjects from 16 to 19 weeks GA range and 14 datasets from 24 to 32 weeks GA range. All datasets were acquired on 3T Philips Achieva MRI system using a 32-channel cardiac coil. The structural T2w (single-short turbo spin echo sequence) datasets include 5-6 stacks with TE=180ms, in-plane resolution 1.25 x 1.25mm, slice thickness 2.5mm and 0/-1.5mm gap.

2) Pre-processing: All stacks were resampled 128x128x128 grid using SVRTK package¹

3) Image Augmentation: When it comes to image classification and segmentation tasks, 3D deep learning algorithms are the go-to choice. However, these algorithms come with their own set of limitations, one of which is their need for a significant amount of data to achieve high performance and generalisability. This can be particularly problematic in the field of medical imaging, where large quantities of images are not always readily available. To address this issue, scientists have turned to the Data Augmentation technique, which is a general method for generating additional data for a training dataset. Studies have shown that this technique can significantly improve performance when evaluated on a dataset.

In this particular study, we applied Data Augmentation as transformations in the training batches, with a probability of 0.5 to ensure a level of randomness. They implemented a total of eight distinct transformations, as illustrated in Figure 3. By using this method, we were able to increase the size of their dataset and thereby improve its quality and performance.

- Random Adjust Contrast function utilises the default gamma of 4.5 to alter the image's intensity by randomly adjusting it within the range of 0.5 and our selected gamma.
- Random Affine function utilises an image geometrically without affecting parallelism. The affine automorphism is set between -0.1 and 0.1 for all three planes. This magnitude is randomly applied to the rotation, translation, and scale ranges.
- Random Gaussian Smooth function utilises Gaussian smoothing to an image. The sigma parameters for all three dimensions are chosen randomly between 0.25 and 1.5 for all axes.
- Random Gaussian Noise function utilises to add noise signal correlating to the normalised probability density.

¹SVRTK package: <https://github.com/SVRTK/SVRTK>

Reference	Method	Segmentation Type	Modality	GA	Subjects	Performance
Chen J. et al. 2021 [17]	3D U-Net	Automated	T2w	23-34	77	CSF, CNR
Khalili N. et al. 2019 [18]	multi-class U-Net	Automated	T2w	23-35	12	CB, BGT, vCSF, WM, BS, cGM, eCSF
Fidon L. et al. 2021b [19]	Multi-class nnU-Net-DRO	Automated	T2w	19-40	124	WM, LV, CB
Zhao L. et al. 2022 [20]	multi-class 3D U-Net	Automated	T2w	24.4-49.4	106	CSF, cGM, WM, dGM, CB, BS
Huang X. et al. 2023 [7]	CoT-CNN	Automated	T2w	20-35	80	eCSF, GM, WM, LV, CB, dGM, BS
Dou H. et al. 2021 [8]	Deep Attention CNN	Automated	T2w	16-39	57	CP
Fidon L. et al. 2021 [9]	3D U-Net	Automated	T2w	20-35	96	WM, LV, CB, eCSF, cGM, dGM, BS, CC
Karimi D. et al. 2023 [10]	multi-class nnUNet	Automated	T2w	19.6-38.9	294	Young (GA <32) Old (GA \geq 32)
Salehi S. et al. 2018 [11]	2D U-Net and Autocontext	Automated	T2w	22-39	315	Sections of the fetal brain
Li J et al. 2020 [12]	Shallow FCN and M-FCN	Automated	T2w	20-30	88	Fetal Brain Extraction
Keraudren K. et al. 2014 [10]	MSER, BoW with SIFT, CRF	Automated	T2w	22-39	66	Localisation and Segmentation the fetal brain
Ebner M. et al. 2020 [14]	Loc-Net, Seg-Net	Automated	T2w	21-31	85	Localisation, Segmentation and Super-Resolution Reconstruction of fetal brain
Uus A. et al. 2022 [3]	3D-CNN	Automated	T2w	21-36	178	Reconstruction of the fetal thorax region
Xie N H. et al. 2020 [15]	CNN	Automated	Ultrasound	22-29	12,780	Classification CNS Abnormalities
Huang R. et al. 2018 [16]	VP-Net	Automated	Ultrasound	20-29	285	CSP, Tha, CB, CM, LV

TABLE I
SUMMARY TABLE FOR PREVIOUS SEGMENTATION STUDIES IN FETAL BRAIN.

Healthy Subjects (HS), Challenging Subjects (ChS), Pathological Subjects (PS), spina bifida (sb), spina bifida after surgery (sbas), abnormal (ab), axial (ax), coronal (cor); sagittal (sag); Cerebellum (CB), White Matter (WM), Gray Matter (GM), cortical Gray Matter (cGM), deep Gray Matter (dGM), Lateral Ventricle (LV), Brain Stem (BS), CerebroSpinal Fluid (CSF), external CerebroSpinal Fluid (eCSF), ventricular CerebroSpinal Fluid (vCSF), Corpus Callosum (CC), Cortical Plate (CP), Basal Ganglia and Thalamus (BGT), Left and Right Hemisphere (LH, RH), Database (DB), Cadherin-related Neuronal Receptor (CNR), Cavum Septi Pellucidi (CSP), Thalamus (Tha), and Cisterna Magna (CM).

- Random Gaussian Sharpen function utilises a secondary transformation that is built on the Gaussian smoothed image. This process randomly sharpens the image, using sigma values that range between 0.5 and 1.0.
- Random Bias Field function utilises a series of linear collection of polynomial functions with a degree of 3 to adjust the MRI. The default coefficient range for this technique is set at (0.0, 0.1).
- Random Histogram Shift function utilises the intensity histograms of an image in a non-linear manner, introducing random variations.
- Random Shift Intensity function adjusts intensity with an offset value of 0.1.

C. Neural Network Architecture

1) **MONAI**: In this study, we utilised the MONAI framework which is open-source and commonly used in academic settings to create our models. It also has various qualities that are ideal for standardised healthcare imaging methods. [21] Our dataset consists of 3D medical images, and the PyTorch-based code

we used is adaptable for different network architectures, losses, and metrics. [22]

2) **Convolutional Neural Networks**: Convolutional Neural Networks (CNN) architecture has been the most commonly used to solve problems in computer vision and medical imaging field. [23] This Architecture utilises convolution, a mathematical operation, instead of general matrix multiplication in at least one of their layers. They are specifically designed to handle pixel data and are commonly applied in image recognition and processing. [24] Previous studies have shown that CNN architecture performs well for classifying or segmenting in the medical field. [23]

3) **UNET**: The UNET architecture is highly effective in various biomedical segmentation applications. It performs well even with limited training datasets and provides more accurate segmentations. Essentially, it enhances a standard contracting network by adding consecutive layers that utilise upsampling operators instead of pooling operators. These additional layers enhance the output's resolution. [25]

In our experiments, we utilise a 3D UNET architecture

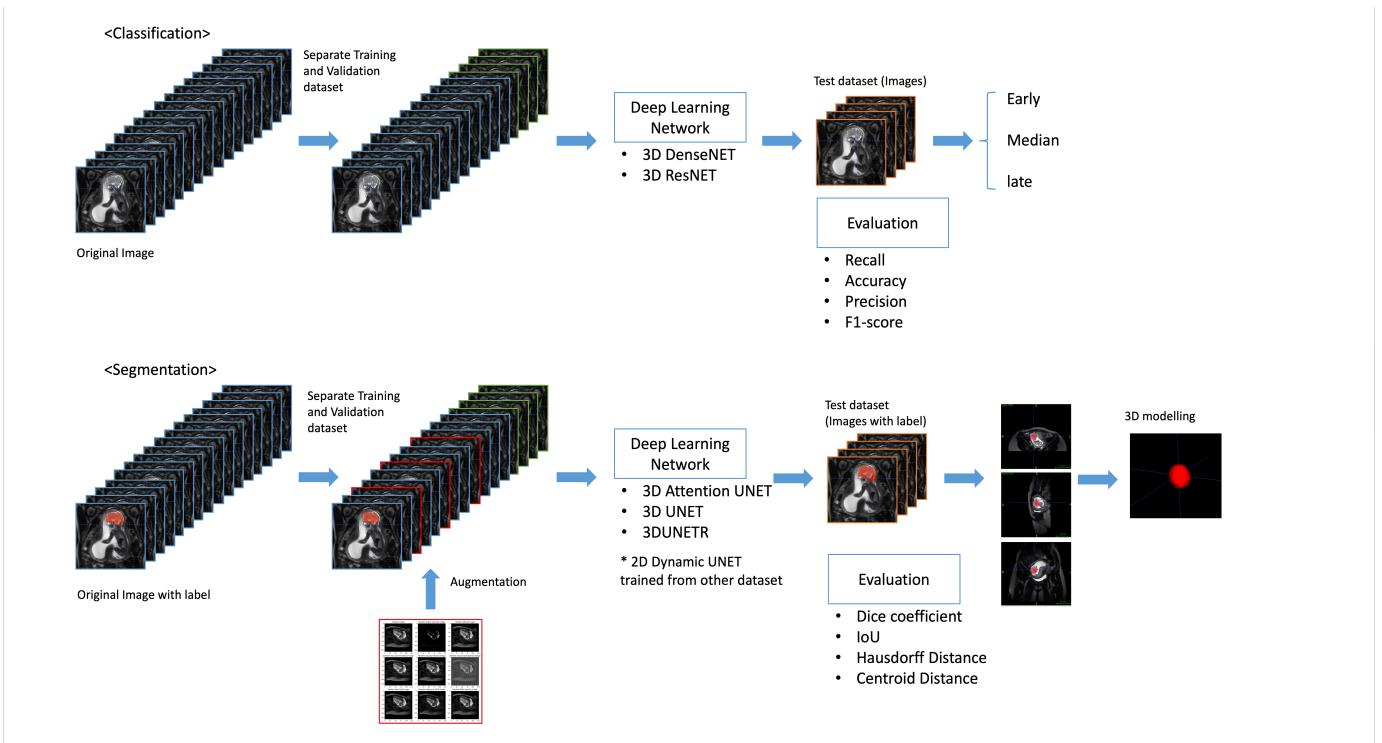


Fig. 2. Overveiw of classification and segmentation process

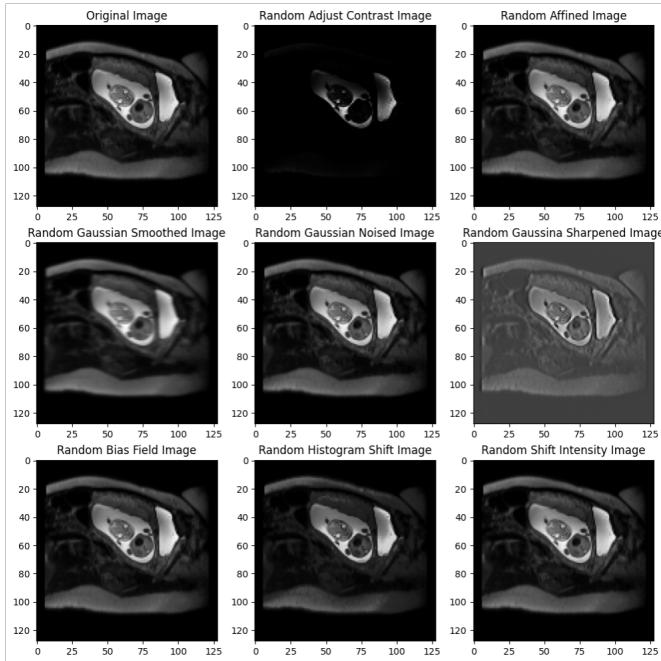


Fig. 3. Visualisations of transforms used in augmentation process.
Images are taken from the sagittal view of the MRI at slice 73 of 128

[26] with five encoding-decoding branches with 32, 64, 128, 256 and 512 channels, respectively. The residual Unit consists of blocks of $3 \times 3 \times 3$ convolutions with a stride of 2, instance Normalisation and PReLU activation, and repeat that blocks of $3 \times 3 \times 3$ convolutions without stride, instance Normalisation and PReLU activation. Upsampling consists

block of $3 \times 3 \times 3$ Convolution Transpose with a stride of 2, concatenate, $3 \times 3 \times 3$ Convolutions, Instance Normalisation and PReLU activation.

4) UNET Transformer: The UNET Transformer (UNETR) is able to effectively segment complex contexts by addressing the limitations of UNET's ability to model long-range contextual interactions and spatial dependencies. This is accomplished through the use of attention mechanisms at two levels. The self-attention module enables global interactions among encoder features, while cross-attention in the skip connections allows for accurate spatial recovery in the UNET decoder by filtering out non-semantic features. [27]

In our experiments, we utilise a 3D UNETR architecture with four encoding-decoding branches. Each encoder block consist of 2 repeated block of $3 \times 3 \times 3$ convolutions, Batchnorm and ReLU activation, while up sampling. Each decoder block consist of 3 repeated $2 \times 2 \times 2$ deconvolutions, $3 \times 3 \times 3$ convolution, Batchnorm and ReLU. other one consist of $2 \times 2 \times 2$ deconvolution. Lastly, the final output is fed into $1 \times 1 \times 1$ convolutional layer with softmax activation to generate voxel-wise semantic predictions.

5) Attention UNET: The combination of UNET and Attention Gate (AG) makes up the UNET Architecture. Through the use of skip connections, the AG filters the propagated features and utilises contextual information from coarser scales to achieve feature selectivity. This type of architecture has shown impressive results in medical imaging and can adapt to focus on target structures of varying shapes and sizes. The AGs have the ability to implicitly learn how to eliminate irrelevant regions in an input image, removing

the necessity for external tissue/organ localisation modules in cascaded convolutional neural networks (CNN). [28]

In our experiments, we utilise a 3D Attention UNET architecture with five encoding-decoding branches with 32, 64, 128, 256 and 512 channels, respectively. Each encoder block consist of 2 repeated block of $3 \times 3 \times 3$ convolutions and ReLU activation, while max pooling by 2. Each decoder block consist the same process with encoder block while upsampling by 2 and Attention Gate. The attention gate filter receives a combination of a gating signal from the previous block and a skip connection from the same level of the encoder and it through ReLU activation, Sigmoid, resampling. Finally multiplication with the original element from the skip connection.

6) Residual Network: Residual Network (ResNET) is a deep learning network that uses "Residual Units" or "Residual Blocks". These blocks contain skip connections that allow input data to pass directly to the next layer. This helps propagate gradients and makes learning easier as the network depth increases [29]. A residual block includes two convolutional layers and a skip connection between them. The skip connection is created by adding the input data. If the size of the output from the convolution layer is not the same as the input data, a 1×1 convolution is used to adjust the dimensions.

In our experiments, we utilise a 3D ResNET-152 architecture with five layers with 64, 128, 256, 512, 1024 and 2048 channels, respectively. The first layer block consists of $1 \times 1 \times 1$ convolutions, Batchnorm, ReLU activation, and Maxpooling. The second layer block consists of three repeated Residual block, which consists of two convolutions layer with activation and two convolutions layer without activation. The third layer consists of 4 Residual blocks, the fourth layer consists of 6 Residual blocks, and the fifth layer consists of 3 Residual blocks and adaptive average pooling.

7) Dense Convolutional Network: Dense Convolutional Network (DenseNET) is an improved architecture of RESNET, it connects every layer to each other in a feed-forward way. It uses the feature maps from all previous layers as inputs for each layer, and its own feature maps as inputs for all subsequent layers. This type of network has several benefits, including reducing the number of parameters, strengthening feature propagation, promoting feature reuse, and addressing the vanishing-gradient problem [30].

In our experiments, we use a 3D DenseNET-169 architecture with three Dense blocks and three Transition blocks. The dense block consists of $1 \times 1 \times 1$ convolutional layers and $3 \times 3 \times 3$ convolutional layers. The Transition block includes $1 \times 1 \times 1$ convolutional layers, average pooling by 2, and stride by 2. The first step is to apply a $7 \times 7 \times 7$ convolutional layer with a stride of 2 to the input image, followed by max pooling with a size by 3 and a stride by 2. Dense and Transition blocks are applied sequentially 3 times, followed by a classification layer which consists of $7 \times 7 \times 7$ global average pooling.

8) Dynamic UNET: Dynamic UNET is a modified version of the UNET structure. Its main feature is the ability to adjust the network architecture dynamically based on the size and spacing of input patches. This adjustment is crucial to ensure that the network's receptive field accurately reflects the size and characteristics of the input image. By changing the depth and downsampling step of the network to match the size and content of the input image, the segmentation process becomes more precise and versatile [31].

We utilised a trained 2D Dynamic UNET model from MONAIfb² in our experiment. The model was trained using 2D patches with a combined loss of Dice and Cross Entropy. Validation was performed in 3D, and the best-performing model snapshot was chosen based on its 3D Dice score. To post-process the results, we selected the largest connected component.

D. Loss Function

Machine learning techniques, especially deep neural networks, are primarily used for medical image segmentation tasks. However, due to the nature of medical data, data volume and class imbalance problems often appear, which are likely to represent low accuracy in segmentation. Using a loss function is a fundamental parameter in the neural network training process that can directly affect class imbalance and model convergence [32]. In the training process, after each iteration, the loss function is compared to the actual measurement to provide a calculated measure of 'accuracy' or prediction. Using these values, the model can optimise the parameters. A particular loss metric is used for semantic segmentation tasks because the criteria for good model performance differ from other classification problems. The optimisation surface is changed by adjusting the learning capacity of the network, and a better local optimisation can be found. These adjusted functions give more weight to selected pixels than others and provide stronger predictions.

In this paper, three different types of loss functions are used for our experiments: Dice Cross Entropy Loss, Dice Focal Loss, and BCE with Logits Loss. A loss function called Dice Cross Entropy Loss combines both the Dice and Cross-Entropy loss functions. The function calculates the weighted sum of these two losses, which smooths the Dice Loss and prevents vanishing gradient [33]. Dice Focal Loss also calculates a weighted sum between Dice and Focal Loss functions. Focal Loss scales down the weighting of easily learned pixels in order to penalise class imbalances and focus on misclassified labels. The combined loss function is effective in handling heavy class imbalances [32]. During the training process, BCE with Logits Loss utilizes a simple sigmoid function after calculating the Binary Cross Entropy between the output and target in a single class. This approach fosters numerical stability. [34].

²Dynamic UNET: <https://github.com/gift-surg/MONAIfb/tree/main>

E. Evaluation Metrics

Our study focused on estimating the performance of algorithms in localising data using metrics such as Dice Coefficient (DC), Intersection over Union (IoU) score, Hausdorff Distance, Recall, Accuracy, Precision, F-1 Score and Centred-point Distance (CD).

1) Dice Coefficient: Dice Coefficient(DC) is widely used for analysing medical images and performing image segmentation. It measures the overlap area between predicted and ground-truth maps, and is calculated by multiplying this area by two and then dividing it by the total number of pixels in both images. The ground truth and predicted segmentation maps are represented by A and B, respectively. The resulting value falls within the range of 0 to 1 [35].

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

2) Intersection over Union: Intersection over Union (IoU) is a prevalent metric used in semantic segmentation. Its purpose is to determine the level of overlap between the anticipated segmentation map and the actual ground truth by computing the ratio of the area of intersection to the area of union. [35].

$$IoU = J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

3) Hausdorff Distance: Hausdorff Distance(HD) is a metric used to measure segmentation errors [36]. The proximity between two images is determined by HD, which is computed by comparing the segmentation borders of the ground truth (Q) and predicted (P).

$$HD(P, Q) = \max \left(\max_{p \in P} \min_{q \in Q} \|p - q\|_2, \max_{q \in Q} \min_{p \in P} \|q - p\|_2 \right) \quad (3)$$

4) Recall: Recall is an important metric as it accounts for both under and over-segmentation when evaluating segmentation accuracy. If the recall score is low, it suggests that there may be under-segmentation. A True Negative(TN) is a pixel that is correctly identified as not being part of the ground truth, whereas a False Negative(FN) refers to an incorrectly predicted pixel that is part of the ground truth. A True Positive(TP) is a pixel that is correctly identified as belonging to the ground truth, while a False Positive(FP) is a pixel that is mistakenly identified as part of the ground truth. [36].

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

5) Accuracy: Accuracy pertains to the correctness of identified pixels in an image, also known as absolute pixel precision. It is a crucial performance metric, but it can be misleading in cases of class mismatch. When one segmented category outnumbers another, a category mismatch occurs. As a result, the accuracy of the prevailing class will outweigh the lower accuracy of other categories, leading to skewed outcomes. If there is no class mismatch, accuracy is a recommended measure for evaluating segmentation results in images [36].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

6) Precision: Precision is a crucial metric for evaluating segmentation outcomes. It refers to the fraction of cerebellum pixels in the automated segmentation that align with the ground truth cerebellum pixels. However, it's important to note that precision may be affected by over segmentation [36].

$$Precision = \frac{TP}{TP + FN} \quad (6)$$

7) F1-Score: To assess segmented images, Recall and Precision can be utilized together to gauge the proximity of the projected regions to the ground truth in regard to their location and level of detail. The F1-score, also referred to as Boundary F1 (BF), is a valuable metric when it comes to matching boundaries or contours between the predicted segmentation and ground truth, as it is the harmonic mean of recall and precision. [36].

$$F1-score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (7)$$

8) Centroid Distance: Centroid Distance is one of the significantly function in machine learning, it is a function that calculates the centre point of an object by finding the index of a non-zero pixel in both the binarised original centre point x_{ik} and comparison point x_{jk} and then calculates the Euclidean distance between the two points using Pythagorean theorem [37].

$$d_{ij} = \sqrt{\sum_{k=1}^3 (x_{ik} - x_{jk})^2} \quad (8)$$

IV. RESULTS

A. Evaluation of Classification

In this study, we evaluate the deep learning algorithms have the capability to accurately classify the gestational age of fetuses into early, median, and late stages before precisely localising the fetal brain.

The accompanying figure4 showcases the fetal brain images obtained during the data acquisition phase. Our study utilised a total of 45 high-quality images, with 31 images designated for training, 14 images for validation, and we tested 6 images in the new test dataset.

Table II provides a detailed evaluation of each network, including recall, accuracy, precision, and F1-score computed on the test dataset. Our analysis revealed that DenseNet outperformed RESNET in terms of recall (0.16),accuracy (0.11), precision (0.16) and F1-Score (0.16). Furthermore, it is noteworthy that the DenseNet model only required a few minutes for training, while the ResNet model took approximately two hours. Our GitHub repository offers the classification process.³

B. Augmentation Visualisation

Figure 3 shows visualisations of transformed images using augmentation in MONAI. The generated samples vary from the original image in brightness, darkness, size, and clarity, but samples maintained correct horizontal and vertical shifts.

³Classification network: <https://github.com/chanhobong/earlyGA>

	Recall	Accuracy	Precision	F1-Score
DenseNet-169	0.83	0.89	0.83	0.83
ResNet-152	0.67	0.78	0.67	0.67

TABLE II
EVALUATION PERFORMANCE COMPARISON OF DIFFERENT NETWORK

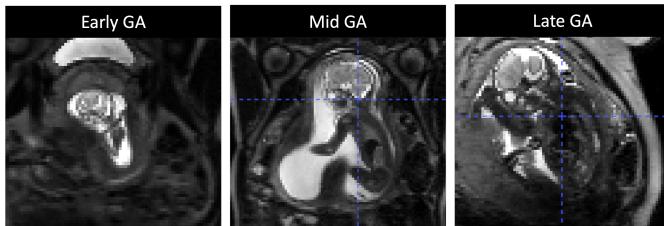


Fig. 4. Early, Median, and Late stage of fetal brain images of gestational age less than 22 weeks

C. Evaluation of Segmentation

According to the Dice Score evaluation, the Attention UNET method achieved the highest score of 0.927, while the 2D Dynamic UNET approach scored the lowest at 0.428. In the IoU evaluation, the Attention UNET method also obtained the highest score of 0.827, while the UNTER method scored the lowest at 0.18. Moreover, in the Hausdorff Distance evaluation, the Attention UNET method scored the lowest at 2.525, whereas the Dynamic UNET approach scored the highest at 11.82. Finally, in the Centroid Distance evaluation, the Attention UNET method scored the lowest at 1.779, while the UNTER approach scored the highest at 26.707.

We trained three different types of networks with and without augmentation in this experiment, we observed that the 3D Attention UNET with 2 augmentations represented that it scored overall well in every evaluation. Furthermore, it showed the most similar 3D modelling with ground truth compared with other networks. On the other hand, 3D UNTER without augmentation performed poorly in every evaluation, including 3D modelling except 2D Dynmaic UNET.

The result of evaluation demonstrated a box plot in Figure 6.

D. Localisation Visualisations of Fetal Brain

In Figure 7, We can observe segmentation labelled in Axial, Sagittal and Coronal planes and 3D modelling using ITK SNAP which is a software application used to segment structures in 3D and 4D biomedical images. The Attention UNET with 2 augmentations provides the best modelling images compared to other networks. On the other hand, 3D modelling of 2D Dynamic UNET demonstrated poorly compared with other 3D modelling of networks

V. LIMITATION

We utilised ResNet and Densenet to classify the gestational ages of fetuses, but the comparison method was lacking. After comparing DenseNet and ResNet, DenseNet had superior performance and trained in only a few minutes, whereas ResNet had poor performance and trained in an hour.

The segmentation performance between the network with augmentation and the network without augmentation was observed unclearly in the segmentation training process. There is a slight gap in the Dice score between augmentation and non-augmentation in UNET, but it cannot be distinguished in cases of Attention UNET. Moreover, The Attention UNET without augmentation performed segmentation and 3D modelling better than the Attention UNET with augmentation. Conversely, The UNET with augmentation performed segmentation and 3D modelling better than the UNET without augmentation.

VI. DISCUSSION

For the localisation of the fetal brain in this study, 39 images of the training set and 6 images of the test set were trained through three different Deep Learning networks for classification. For segmentation, 188 images and labels of the training set and 23 images and labels of the test set were used to train different types of Deep Learning networks under different augmentation conditions through MONAI using the v100 GPU with high RAM in Google Colab.

When it comes to training Deep Learning networks to classify data, the amount and quality of the training dataset are crucial. Insufficient data can lead to common issues in Deep Learning, such as overfitting or underfitting which are one of the reasons to score poorly in evaluation. Previous studies have found that increasing the number of images in the dataset can help overcome these problems and improve the overall performance of the classification model. In short, having a robust and diverse training dataset is essential for effective Deep Learning classification. Moreover, DenseNet outperformed ResNet in this study, but access to more datasets could improve its performance further.

Regarding our classification experiments, we used MRI scans of a whole fetus in uterus, as depicted in Figure 4. However, the classification of fetal brains at varying gestational ages, along with the background (womb) and other parts such as the body and organs of the fetus, can adversely impact the accuracy of the classification. We propose that entails a process of segmenting the fetal brain initially, followed by training the classification for each size of the fetal brain at various gestational ages. Removing the background womb, fetal body, and organs reduces the computing, allowing the network to focus on the brain . It represent a Figure5. Such an approach

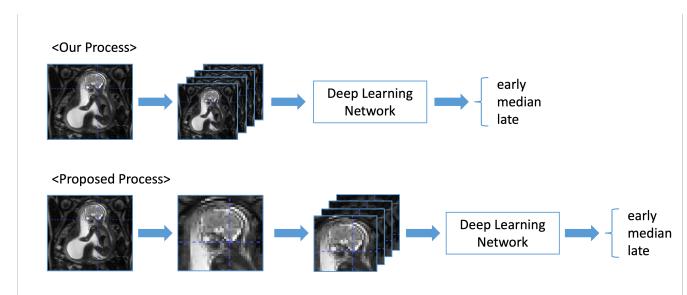


Fig. 5. Visualisation of fetal brain localisation to identify the optimal models for each neural network.

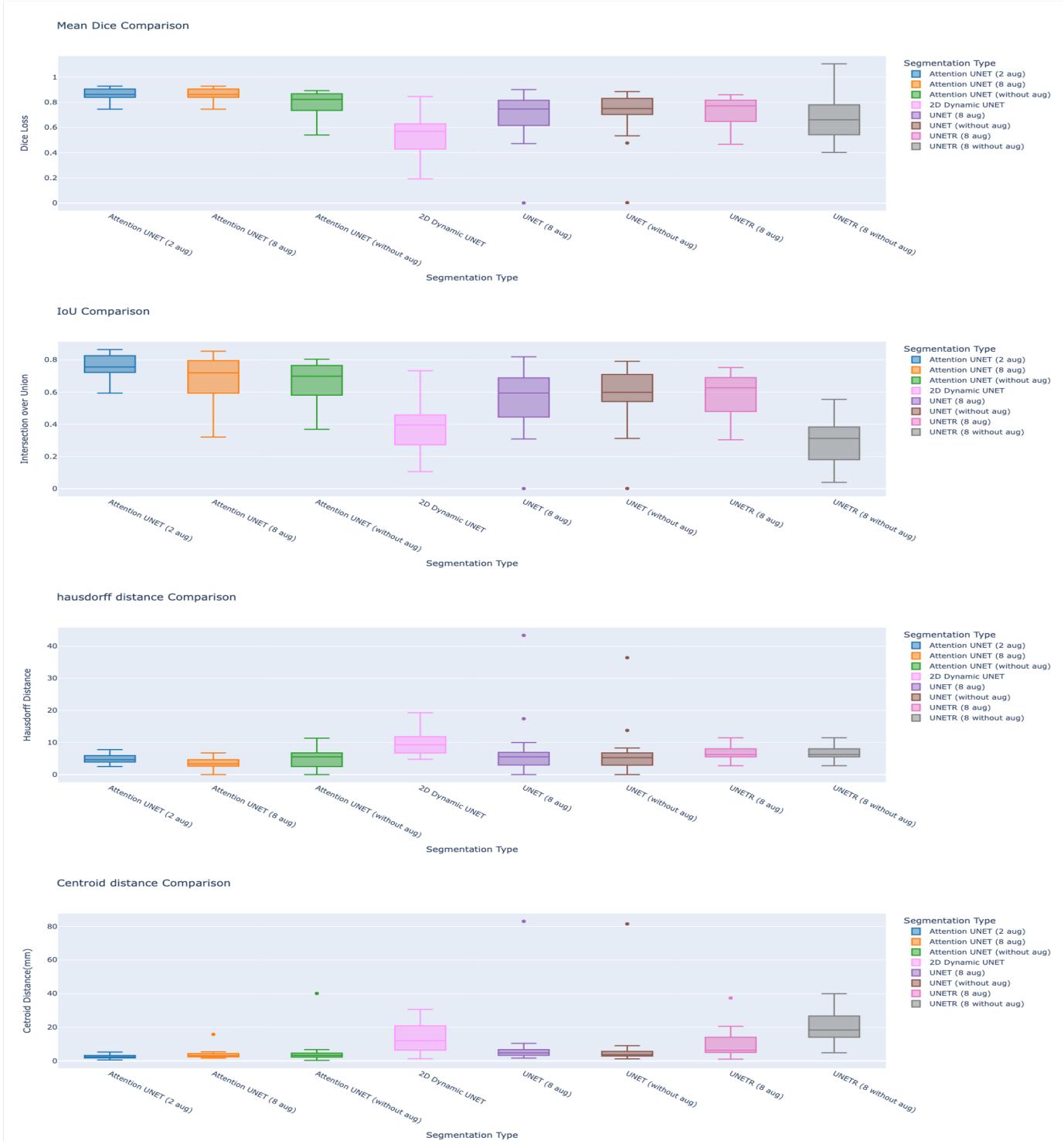


Fig. 6. Evaluation of localisation of fetal brain to identify the optimal models for each neural network.

can significantly enhance the efficiency and accuracy of the network, enabling it to deliver more reliable results.

Furthermore, using different deep learning classification networks aside from the one we used would make a good comparison group to improve our results.

In segmentation experiments, We added several augmentations to improve the segmentation result. However, finding a connection between the networks with and without augmen-

tation was hard.

Throughout our experiment, we utilised 8 different augmentation methods to improve the precision of our segmentation outcomes. The evaluation scores acquired from each network case did not exhibit significant differences when compared to one another. noticeably, we noted that certain networks that lacked any form of augmentation were able to perform 3D modelling with greater proficiency than those that did feature



Fig. 7. Visualisation of fetal brain localisation to identify the optimal models for each neural network.

augmentation. As a result, we suppose that exploring alternative augmentation methods or decreasing the augmentation employed may lead to improved modelling performance.

According to the evaluation scores in AttentionUNET, it was observed that three different types of networks got good scores but they had a difference in 3D modelling. Therefore, changing the hyperparameters of the network, such as reducing the batch size, changing the number of channels, or increasing

the training time, may result in better performance.

The spatial dimension setting is crucial for localisation. Most networks, except 2D Dynamic UNET, were used as 3D networks in this experiment. Despite Dynamic UNET being trained on 2D networks, it produced meaningful scores in our experiments. However, its localisation demonstrated poor results. These are interesting results, and it may be a challenging point for studying the reason and to improve performance.

Furthermore, we estimate that training 2D Dynamic UNET with our dataset may enhance localisation.

VII. CONCLUSION

We evaluated the effectiveness of various neural networks in accurately localising early GA (20 ~ 34 weeks) fetal brain MRI stacks that may be affected by motion. Our approach used state-of-the-art deep learning architectures, various augmentation methods, and other evaluation metrics to generate highly accurate classification and segmentation models and evaluate them.

In terms of 3D brain segmentation localisation, the AttentionUNet model outperformed other neural network training models, with notable differences in various evaluations. Additionally, it was found to be the most similar to the ground truth in the 3D localisation aspect, out of all the neural network training models. Additionally, 2D Dynamic UNET which trained on a different dataset, scored poorly in the evaluation and produced low-quality 3D models, indicating a need for further study. Our findings suggest that using 3D classification and 3D segmentation neural network methods is feasible for the analysis of early GA fetal brain localisation and they outperformed 2D neural networks.

VIII. ACKNOWLEDGMENT

We express gratitude to all those involved in acquiring and analysing the datasets at the Department of Perinatal Imaging and Health at Kings College London and St Thomas' Hospital. Our appreciation extends to all participants and their families. Lisa Story received support from NIHR Advanced Fellowship [NIHR30166], MRC Confidence in Concept [MC_PC_19041], and the NIH Human Placenta Project grant [1U01HD087202-01]. The NIHR Clinical Research Facility (CRF) at Guy's and St Thomas', as well as the National Institute for Health Research Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust, also provided support. The authors are responsible for the opinions expressed and not the NHS, the NIHR, or the Department of Health. I would also like to express my gratitude to my supervisor, Dr. Alena Uus, for her invaluable expert guidance and support throughout the project and for allowing me to work on it.

REFERENCES

- [1] A. U. Uus, A. E. Collado, T. A. Roberts, J. V. Hajnal, M. A. Rutherford, and M. Deprez, "Retrospective motion correction in foetal mri for clinical applications: existing methods, applications and integration into clinical practice," *The British Journal of Radiology*, 8 2022.
- [2] A. Gholipour, J. A. Estroff, and S. K. Warfield, "Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain mri," *IEEE transactions on medical imaging*, vol. 29, no. 10, pp. 1739–1758, 2010.
- [3] A. U. Uus, I. Grigorescu, M. P. van Poppel, J. K. Steinweg, T. A. Roberts, M. A. Rutherford, J. V. Hajnal, D. F. Lloyd, K. Pushparajah, and M. Deprez, "Automated 3d reconstruction of the fetal thorax in the standard atlas space from motion-corrupted mri stacks for 21–36 weeks ga range," *Medical image analysis*, vol. 80, p. 102484, 2022.
- [4] J. R. Davidson, A. Uus, J. Matthew, A. M. Egloff, M. Deprez, I. Yardley, P. De Coppi, A. David, J. Carmichael, and M. A. Rutherford, "Fetal body mri and its application to fetal and neonatal treatment: an illustrative review," *The Lancet Child & Adolescent Health*, vol. 5, no. 6, pp. 447–458, 2021.
- [5] M. Kuklisova-Murgasova, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, and J. A. Schnabel, "Reconstruction of fetal brain mri with intensity matching and complete outlier removal," *Medical image analysis*, vol. 16, no. 8, pp. 1550–1564, 2012.
- [6] B. Hou, B. Khanal, A. Alansary, S. McDonagh, A. Davidson, M. Rutherford, J. V. Hajnal, D. Rueckert, B. Glocker, and B. Kainz, "3-d reconstruction in canonical co-ordinate space from arbitrarily oriented 2-d images," *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1737–1750, 2018.
- [7] X. Huang, Y. Liu, Y. Li, K. Qi, A. Gao, B. Zheng, D. Liang, and X. Long, "Deep learning-based multiclass brain tissue segmentation in fetal mrIs," *Sensors*, vol. 23, no. 2, p. 655, 2023.
- [8] H. Dou, D. Karimi, C. K. Rollins, C. M. Ortinau, L. Vasung, C. Velasco-Annis, A. Oualam, X. Yang, D. Ni, and A. Gholipour, "A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal mri," *IEEE transactions on medical imaging*, vol. 40, no. 4, pp. 1123–1133, 2020.
- [9] L. Fidon, M. Aertsen, D. Emam, N. Mufti, F. Guffens, T. Deprest, P. Demaerel, A. L. David, A. Melbourne, S. Ourselin *et al.*, "Label-set loss functions for partial supervision: application to fetal brain 3d mri parcellation," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II* 24. Springer, 2021, pp. 647–657.
- [10] D. Karimi, C. K. Rollins, C. Velasco-Annis, A. Oualam, and A. Gholipour, "Learning to segment fetal brain tissue from noisy annotations," *Medical Image Analysis*, vol. 85, p. 102731, 2023.
- [11] S. S. M. Salehi, S. R. Hashemi, C. Velasco-Annis, A. Oualam, J. A. Estroff, D. Erdogan, S. K. Warfield, and A. Gholipour, "Real-time automatic fetal brain extraction in fetal mri by deep learning," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 720–724.
- [12] J. Li, Y. Luo, L. Shi, X. Zhang, M. Li, B. Zhang, and D. Wang, "Automatic fetal brain extraction from 2d in utero fetal mri slices using deep neural network," *Neurocomputing*, vol. 378, pp. 335–349, 2020.
- [13] K. Keraudren, M. Kuklisova-Murgasova, V. Kyriakopoulou, C. Malamatieni, M. A. Rutherford, B. Kainz, J. V. Hajnal, and D. Rueckert, "Automated fetal brain segmentation from 2d mri slices for motion correction," *NeuroImage*, vol. 101, pp. 633–643, 2014.
- [14] M. Ebner, G. Wang, W. Li, M. Aertsen, P. A. Patel, R. Aughwane, A. Melbourne, T. Doel, S. Dymarkowski, P. De Coppi *et al.*, "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri," *NeuroImage*, vol. 206, p. 116324, 2020.
- [15] H. Xie, N. Wang, M. He, L. Zhang, H. Cai, J. Xian, M. Lin, J. Zheng, and Y. Yang, "Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal," *Ultrasound in Obstetrics & Gynecology*, vol. 56, no. 4, pp. 579–587, 2020.
- [16] R. Huang, W. Xie, and J. A. Noble, "Vp-nets: Efficient automatic localization of key brain structures in 3d fetal neurosonography," *Medical image analysis*, vol. 47, pp. 127–139, 2018.
- [17] J. Chen, Z. Fang, G. Zhang, L. Ling, G. Li, H. Zhang, and L. Wang, "Automatic brain extraction from 3d fetal mr image with deep learning-based multi-step framework," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101848, 2021.
- [18] N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. de Heus, T. Kolk, M. A. Viergever, M. J. Benders, and I. Isgum, "Automatic brain tissue segmentation in fetal mri using convolutional neural networks," *Magnetic resonance imaging*, vol. 64, pp. 77–89, 2019.
- [19] L. Fidon, M. Aertsen, N. Mufti, T. Deprest, D. Emam, F. Guffens, E. Schwartz, M. Ebner, D. Prager, G. Kasprian *et al.*, "Distributionally robust segmentation of abnormal fetal brain 3d mri," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis: 3rd International Workshop, UNSURE 2021, and 6th International Workshop, PIPPI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 3*. Springer, 2021, pp. 263–273.
- [20] L. Zhao, J. Asis-Cruz, X. Feng, Y. Wu, K. Kapse, A. Largent, J. Quistorff, C. Lopez, D. Wu, K. Qing *et al.*, "Automated 3d fetal brain segmentation using an optimized deep learning approach," *American Journal of Neuroradiology*, vol. 43, no. 3, pp. 448–454, 2022.
- [21] M. Ranzini, L. Fidon, S. Ourselin, M. Modat, and T. Vercauteren, "MonaiFbs: Monai-based fetal brain mri deep learning segmentation," *arXiv preprint arXiv:2103.13314*, 2021.
- [22] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murray, A. Myronenko, C. Zhao, D. Yang *et al.*, "Monai: An open-source frame-

- work for deep learning in healthcare,” *arXiv preprint arXiv:2211.02701*, 2022.
- [23] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.
- [24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [25] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [26] E. Kerfoot, J. Clough, I. Oksuz, J. Lee, A. P. King, and J. A. Schnabel, “Left-ventricle quantification using residual u-net,” in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer, 2019, pp. 371–380.
- [27] O. Petit, N. Thome, C. Rambour, L. Themry, T. Collins, and L. Soler, “U-net transformer: Self and cross attention for medical image segmentation,” in *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer, 2021, pp. 267–276.
- [28] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [31] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert *et al.*, “nnu-net: Self-adapting framework for u-net-based medical image segmentation,” *arXiv preprint arXiv:1809.10486*, 2018.
- [32] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, “Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.
- [33] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, “Combo loss: Handling input and output imbalance in multi-organ segmentation,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, 2019.
- [34] Z. Kuang and X. Tie, “Flow-based video segmentation for human head and shoulders,” *arXiv preprint arXiv:2104.09752*, 2021.
- [35] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image segmentation using deep learning: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.
- [36] R. Sreelakshmy, A. Titus, N. Sasirekha, E. Logashanmugam, R. B. Begam, G. Ramkumar, and R. Raju, “An automated deep learning model for the cerebellum segmentation from fetal brain images,” *BioMed Research International*, vol. 2022, 2022.
- [37] R. Suwanda, Z. Syahputra, and E. M. Zamzami, “Analysis of euclidean distance and manhattan distance in the k-means algorithm for variations number of centroid k,” in *Journal of Physics: Conference Series*, vol. 1566, no. 1. IOP Publishing, 2020, p. 012058.