



# Cortical Plate Segmentation Using CNNs in 3D Fetal Ultrasound

Madeleine K. Wyburd<sup>1</sup>(✉), Mark Jenkinson<sup>2,3,4</sup>, and Ana I. L. Namburete<sup>1</sup>

<sup>1</sup> Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

`madeleine.wyburd@hertford.ox.ac.uk`

<sup>2</sup> Wellcome Centre for Integrative Neuroimaging, FMRIB, University of Oxford, Oxford, UK

<sup>3</sup> Australian Institute for Machine Learning (AIML), Department of Computer Science, University of Adelaide, Adelaide, Australia

<sup>4</sup> South Australian Health and Medical Research Institute (SAHMRI), North Terrace, Adelaide, Australia

**Abstract.** As the fetal brain develops, its surface undergoes rapid changes in shape and morphology. Variations in the emergence of the sulci on the brain surface have commonly been associated with diseased or at-risk pregnancies. Therefore, the process of surface folding is an important biomarker to characterise. Previous work has studied such changes by automatically delineating the cortical plate from MRI images. However, this has not been demonstrated from ultrasound, which is more commonly used for antenatal care. In this work we propose a novel method for segmenting the cortical plate from 3D ultrasound images using three varieties of convolutional neural networks (CNNs). Recent work has found improvements in medical image segmentations using multi-task learning with a distance transform regularizer. Here we implemented a similar method but found it was outperformed by the U-Net, which was able to segment the cortical plate with a Dice score of  $0.81 \pm 0.06$ .

**Keywords:** 3D ultrasound · Fetal brain · Gyrfication · Segmentation · U-net · Multi-task learning · CNN

## 1 Introduction

As a human fetus develops, the smooth (lissencephalic) cerebral cortex begins to fold, a process known as *gyrfication*, which continues through childhood and adolescence [23]. Gyrfication allows for the cortical surface area to scale almost linearly with volume, which is associated with increased cognitive abilities [23, 24]. In healthy fetal development the occurrence of major peaks and troughs in the folds, otherwise known as *gyri* and *sulci*, has been shown to be temporally consistent across a large population [2, 21]. Abnormal cortical folding has frequently been associated with severe intellectual disabilities, psychosis

and autism in adolescents and adults [3, 9]. Moreover, recent work has measured gyrification during pregnancies and found significant differences in timings and depths of sulci development between healthy and at-risk fetuses [13]. Having the ability to characterise and quantify gyrification during development is therefore of huge importance, as it could help monitor pregnancies to ensure that the cortical development is at the expected rate and identify any biomarkers for risk and disease.

Emerging around the twelfth gestational week, the Sylvian Fissure (SF) is the first major sulcus observed in fetal brain development and has been extensively studied [20]. Measuring the SF depth has a well established protocol in 2D ultrasound images. The standard transventricular axial plane of the fetal head is obtained which can be identified by the presence of the cavum septi pellucidi, atrium and posterior horns of the lateral ventricles [17]. SF depth is then defined as the straight line between the border of the inner cortex to the inner parietal bone, perpendicular to the midline. For ultrasound-based studies, manually-acquired linear measurements are still routinely performed. Using Magnetic Resonance Imaging (MRI), a similar approach was initially followed before methods became mainly automated [10]. Furthermore, advances in ultrasound techniques are sought to reduce the time consuming and tedious manual measurements performed by sonographers.

In recent work, the cortical plate has been automatically segmented from MRI volumes in order to analyse the rate of gyrification [5, 6, 13]. Delineating the cortical plate is commonly approached using atlas-aided tissue segmentation or MRI toolboxes. In [4] the authors non-rigidly registered a 4D labelled fetal atlas [11] to their dataset before manually correcting the labels, whilst [13] and [8] used an adapted version of the BrainVisa toolbox to segment the cortex [14]. These methods are limited by the availability of fetal atlases and compatible toolboxes which operate within a chosen image modality. Compared to fetal MRI, ultrasound has the advantages of being relatively cheap, available worldwide, and having shortened acquisition times (in the scale of seconds). However, to the best of our knowledge, the current methods of cortical plate delineation have been applied exclusively to MRI volumes and are incompatible with ultrasound data. In this work we address this by proposing a novel method for segmenting the developing cortex from ultrasound volumes.

The challenging task of structure segmentation and classification in medical images has been overcome in recent years by convolutional neural networks (CNNs). Dominating in semantic segmentation is the U-Net [22], which achieves high quality segmentation on medical datasets. Recent state-of-the-art networks have been designed by simply modifying its architecture [27]. One such modification adds auxiliary tasks to the network, known as multi-task learning (MTL). MTL networks exploit the hidden interdependencies across multiple related tasks by simultaneously optimising the performance on all tasks during training. In doing so, they learn to ignore task-specific noise, making the performance more generalisable [7].

MTL can be used to solve multiple problems simultaneously within one network, or to regularise the network with an auxiliary task. In the context of medical imaging the former has been commonly achieved [7, 16]. For example, the authors in [16] trained a single CNN for brain localization, structural segmentation and alignment using 3D ultrasound volumes and reported equivalent performance to that of a CNN trained on each task individually.

In this work we focus on whether using additional information as a regularizer aids the segmentation of the developing cortical plate. We propose an auxiliary regression task of predicting the distance from each voxel within the volume to the boundary of our segmentation, with the hypothesis that this could aid segmentation as it inherently provides a metric for depth, which is a characteristic that is regularly studied. A similar approach was taken in [7] who reported significantly improved segmentations of cardiac chambers by implementing a supplementary pixel-wise distance map regression task across three architecture types.

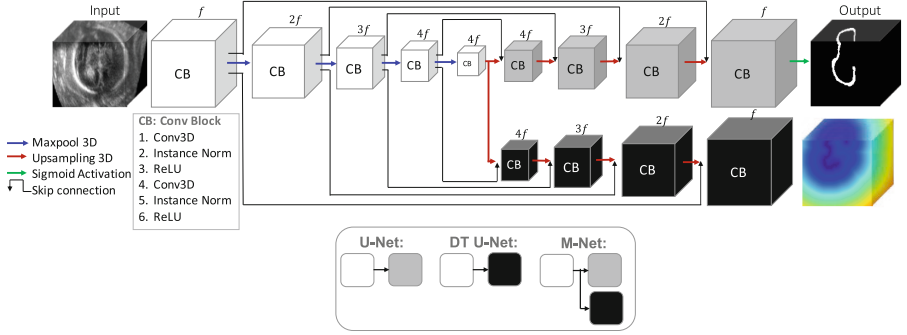
Here we investigate three architecture designs for cortical plate segmentation in fetal brain images: a task that to our knowledge has not been performed using neural networks or from ultrasound images. Specifically, we compare the results from the MTL network (M-Net), a modified U-Net, and a distance transform regression network (DT U-Net).

## 2 Network Design

Throughout this work we investigate whether MTL with a distance map regularizer (DMR) aids segmentation of the cortical plate. To act as a control we compare its performance to two encoder-decoder network designs. For the first control, we treat the segmentation as a simple classification task, which we refer to as the U-Net throughout. Secondly, we introduce the DT U-Net which uses regression to learn the distance transform of the cortical plate.

**U-Net:** The foundations for all three architectures follow the structure implemented by [22] with alterations in depth and number of feature maps, illustrated in Fig. 1. The size of the network was limited due to memory constraints.

The 3D encoder-decoder network was made up of  $l = 4$  convolutional blocks and down-sampling layers, followed by a bottleneck convolutional block, and finally  $l = 4$  upsampling and convolutional blocks. Each convolutional block consisted of a convolution with kernel size  $k = 3 \times 3 \times 3$  followed by instance normalisation [25] and a ReLu activation function, which was then repeated as shown in Fig. 1. Instance normalisation was used after each convolution as we were limited to a small batch size of  $b = 3$ . Skip connections were used to concatenate feature maps from the encoder blocks to the corresponding decoder blocks. For the first convolutional layer the initial number of feature maps was set to  $f = 8$  and doubled at each layer. Max-pooling and upsampling were each used with kernel size and stride of 2. To classify each voxel between 0 and 1 a sigmoid activation function was applied to the output of the final convolutional block.



**Fig. 1.** Schematic of the 3D CNNs used for cortical plate segmentation. The convolutional blocks consisted of two repeats of a 3D convolution, Instance normalisation and ReLU activation. Between the encoder and decoder convolutional blocks skip connections are used to concatenate the layers together. The U-Net cortical plate segmentation is generated using the light grey decoder, whereas the DT U-Net follows the dark grey decoder to produce the distance map of the cortical surface. For the multi task network (M-Net) the architecture is comprised of both the light grey and dark grey decoders. A simplified colour-coded encoder-decoder schematic of each network is shown at the bottom.

During training the network was optimised using Adam [12], a learning rate of  $L = 10^{-4}$  and a Dice loss function. The dice loss ( $\mathcal{L}_{DSC}$ ) between the cortical plate  $A$  and our prediction  $B$ , is calculated from the overlap over the total volume:

$$\mathcal{L}_{DSC} = 1 - \frac{(A \cap B)}{(A + B)}. \quad (1)$$

**DT U-Net:** A distance transform of the cortical plate labels was used as the ground-truth for the regression task. This was computed by calculating the Euclidean distance from each pixel to the nearest boundary on the binary segmentation within the volume.

In order to compare the performance between the classification and regression tasks, the encoder-decoder architecture was kept consistent with two minor changes in the decoder, as shown in dark grey in Fig. 1. The sigmoid activation function at the output layer was removed and the training loss was changed to the absolute deviations ( $\mathcal{L}_{L1}$ ) between the true distance transform and the predicted likelihood map:

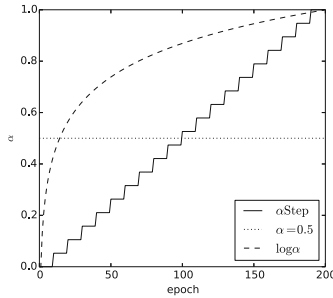
$$\mathcal{L}_{L1} = |A - B|. \quad (2)$$

**M-Net:** An MTL network with a DMR was the third network implemented. The architecture was a combination of the modified U-Net and DT U-Net. It consists of one encoder with two decoder streams branching off from the bottleneck: one for the binary segmentation and one for the distance transform regression task,

shown in Fig. 1. The network training loss was a weighted sum of the  $\mathcal{L}_{DSC}$  and  $\mathcal{L}_{L1}$ , for the binary segmentation and regression task, respectively. Throughout training  $\mathcal{L}_{L1}$  was normalised such that the highest value was equal to 1.

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{DSC} + (1 - \alpha)\mathcal{L}_{L1}. \quad (3)$$

The weighting ( $\alpha$ ) between the two loss functions was investigated. Initially, both loss functions had equal weighting ( $\alpha = 0.5$ ) during training before investigating the effect of adjusting  $\alpha$  over epochs. Preliminary results showed that having a larger weight on the regression task at the beginning of training (e.g.  $\alpha \approx 0$ ) and decreasing its influence with epochs performed better than decreasing  $\alpha$  with epochs. A logarithmic and a linear step  $\alpha$ -scale starting at  $\alpha = 0$  were therefore implemented, shown in Fig. 2.

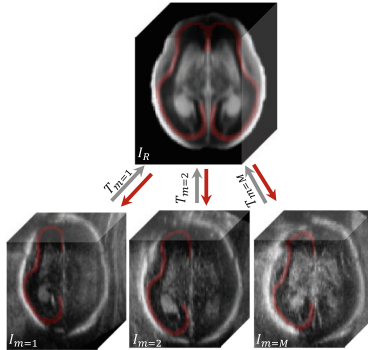


**Fig. 2.** The weighting between  $\mathcal{L}_{L1}$  and  $\mathcal{L}_{DSC}$  is described by  $\alpha$ . Step  $\alpha$  fractionally increases  $\alpha$  every ten epochs, reaching  $\alpha = 1$  in the final 10 epochs.  $\log\alpha$  logarithmically increases  $\alpha$  before plateauing to 1 at the final epoch.

### 3 Experimental Setup

#### 3.1 Dataset

A total of 307 fetal brain volumes between 126 to 160 days of gestation were used throughout the experiment. These volumes were obtained as part of INTERGROWTH-21<sup>st</sup>'s Fetal Growth Longitudinal Study (FGLS), using a Philips HD9 curvilinear probe [18]. The multi-site dataset is comprised of scans collected at 2–5 MHz wave frequency. Each participant was screened for optimal health and the infants were subject to cognitive tests at 2 years old, ensuring healthy development is being studied. In ultrasound volumes the hemisphere closest to the probe becomes distorted due to acoustic shadows of the beam, therefore, we are only able to study the distal hemisphere. The volumes were resampled to an isotropic voxel size of  $0.6 \times 0.6 \times 0.6$  mm before being aligned into a common reference space as described in [16].



**Fig. 3.** The propagation of  $M$  individual fetal brain volumes to a common space  $I_R$ , was performed by calculating a set of transformation vectors  $T_m$  in [15] and illustrated by the grey arrows. The cortical plate was manually annotated in the common space (in red) and propagated back to individuals using the inverse of the transformation fields, shown by the red arrows. (Color figure online)

### 3.2 Atlas-Based Label Propagation

Good quality ground-truth labels are required in deep learning for accurate model representation. Although methods such as semi-supervised learning and noisy labels are becoming more popular, usually some manual annotation is unavoidable, which is time consuming and can be expensive. Using atlas-based label propagation is one way to reduce this workload.

In [15] the authors describe how for each gestational week an atlas representing the population average was created using groupwise Demon registration from the same aligned set of  $M$  images used throughout this work. This involved calculating a set of transformations  $T_m$ , where  $m = (1, \dots, M)$ , needed to map each individual’s image  $I_m$  to the common reference space  $I_R$ , as illustrated in Fig. 3.

The inverse of the transformations  $T_m^{-1}$ , can be used to warp a volume from the reference space back into individual space, shown in red in Fig. 3. Here, we utilised this relationship to propagate manually annotated labels of the cortical plate from the atlases of the gestational ages of interest (18–22 weeks) to each individual fetal ultrasound. To ensure high correspondence between the generated label and cortical plate each volume was visually inspected and manually corrected using ITKsnap [26].

### 3.3 Network Implementation

The dataset was separated into 271 training and 36 testing volumes with an even distribution of age and observed cerebral hemisphere across the two groups. In total 5 models were explored, the U-Net, DT U-Net and M-Net, which in turn had three variations of task weighting,  $\alpha$ . To evaluate the performance of each design, five-fold cross-validation experiments were implemented. The training

volumes were subdivided into subsets with 217 used for training and 54 for validation, for each fold. The best-performing model across the five-folds was then re-trained using all 271 volumes and evaluated on the held-out testing set.

The training of each network was performed on a NVIDIA GeForce GTX 1080 GPU and implemented in PyTorch [19]. Each network was trained for 200 epochs, saving out the model with the highest Dice coefficient for the U-Net and M-Net, and the lowest  $\mathcal{L}_{L1}$  score for the DT U-Net. A binary segmentation was acquired from the predicted likelihood function by applying a threshold of  $t = 0.6$ , before evaluating performance using the Dice coefficient and binary cross-entropy.

## 4 Results

**Table 1.** The averaged Dice Score and binary cross-entropy (BCE) for each architecture and variation of  $\alpha$  across the 5-fold cross validation, between the predictions and the ground-truth labels. The results in bold illustrate the best performance for that measure.

	BCE	Dice
U-Net	<b>0.087 <math>\pm</math> 0.004</b>	<b>0.816 <math>\pm</math> 0.004</b>
DT U-Net	0.175 $\pm$ 0.062	0.307 $\pm$ 0.377
M-Net $\alpha = 0.5$	0.090 $\pm$ 0.001	0.811 $\pm$ 0.005
M-Net $\log(\alpha)$	0.088 $\pm$ 0.004	0.810 $\pm$ 0.005
M-Net Step $\alpha$	0.099 $\pm$ 0.015	0.799 $\pm$ 0.017

### 4.1 Cross Validation

The averaged performance of each network across the 5-fold cross validation is shown in Table 1. The U-Net and M-Net displayed consistent behaviour across the five folds, with similar performance across architecture types. Despite our hypothesis that MTL using a DMR had the ability to improve our results by providing additional supporting information, the performance of the U-Net was still superior (DSC:  $0.816 \pm 0.004$ ).

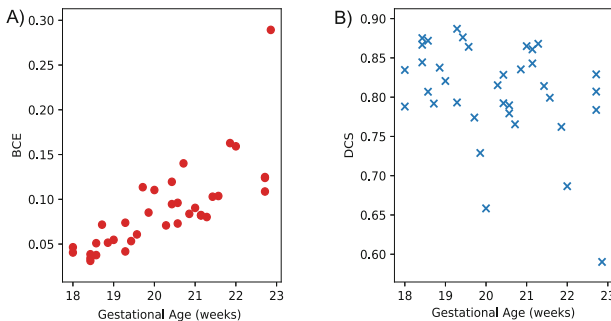
The volume of the cortical plate across one hemisphere occupied 0.5% of the total ultrasound volume. Therefore, the background voxels dominate, skewing the cross-entropy scores. Predicting a blank segmentation still generates a binary cross-entropy score of  $0.23 \pm 0.04$  on average, which should be considered whilst interpreting the results in Table 1. For the DT U-Net, three of the five folds’ scored approximately 0.23 in cross entropy and had a negligible Dice score and therefore likely to be producing blank predictions. The ground-truth distance maps on average had a range of values between 0 and 80. By visualising the outputs from these three folds, it is evident that this range was not being learnt throughout training and therefore, applying the threshold of  $t = 0.6$  to the

predicted distance maps did not have the desired effect of isolating the cortical plate. However, for  $k = 1$  and  $k = 4$  the binary cross-entropy and Dice score are comparable to the other networks. The inconsistency of this network suggests it is unstable which may be a result of the difficulty of using CNN-based feature maps to predict voxelwise distance values relative to some reference.

The performance across the three varieties of alpha schedule showed little variation between results or compared to the U-Net, which could imply that the network learnt to ignore the regression task and focus on the segmentation. This is particularly interesting for the  $\alpha = 0.5$  model, as the other cases purposefully suppress the regression task by increasing alpha. The segmented cortical plate results for  $\alpha = 0.5$  achieved an average Dice score of 0.811, however, the corresponding distance transform regression task only achieved a normalised  $\mathcal{L}_{L1}$  of 0.48, implying the task was too challenging or being ignored. In other words, throughout training the regression decoder could have been suppressed, effectively generating the same architecture as the U-Net and explaining the similarities between the results. This is supported by the unstable results demonstrated by DT U-Net, suggesting the regression task was too challenging to be modelled by the simple DT U-Net network and therefore, could be hindering the performance of the M-Net rather than helping.

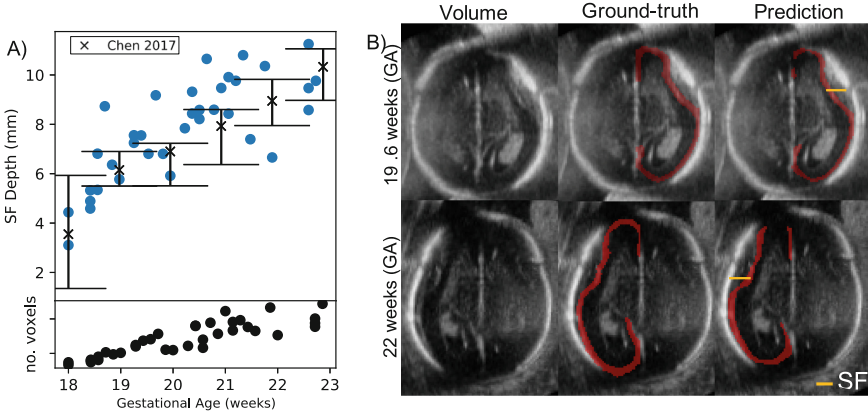
## 4.2 Cortical Plate Segmentation

As shown in Table 1, the U-Net outperformed the other networks and was therefore retrained across the whole dataset, applied to the test volumes and used for further experiments. On the 36 test volumes, this network achieved an average Dice score of  $0.81 \pm 0.06$  and a binary cross-entropy of  $0.09 \pm 0.05$ .



**Fig. 4.** The binary cross-entropy (BCE) and Dice score (DCS) compared to the gestational age of each individual. A low BCE and a high DCS is representative of a good segmentation.





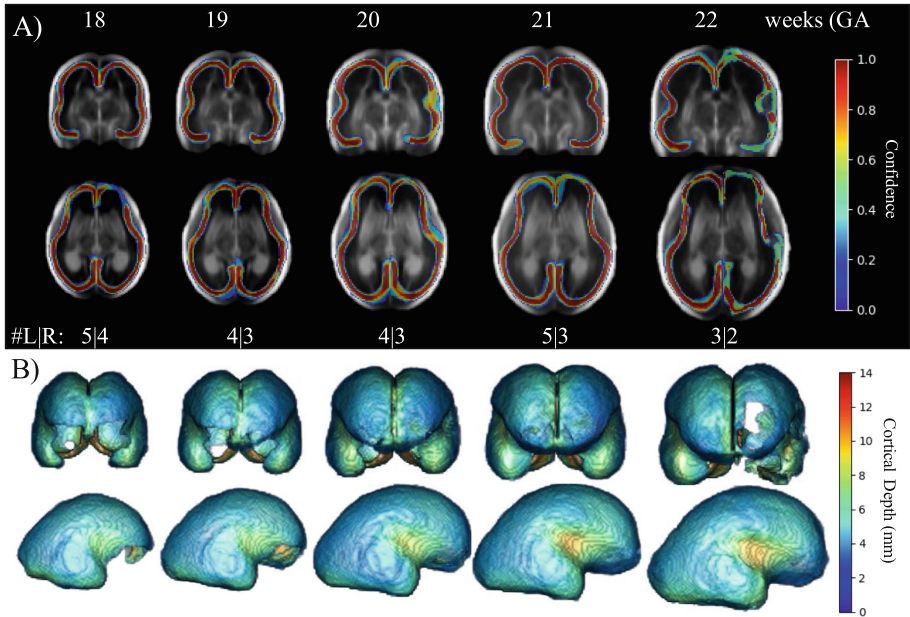
**Fig. 5.** **A)** The measured SF depth in the transventricular axial plane, shown in yellow in **B**. The range of SF depths measured by [1] for a similar age group is also plotted, with the bar representing shallowest and deepest depths measured for each gestational week. The number of voxel in the segmentation is shown in the bottom plot. **B)** The cortical plate labels and predicted masks for two gestational ages. The Sylvian Fissure measurement is shown in yellow. (Color figure online)

The error between each individual’s predicted and labelled cortical plate was compared to gestational age, shown in Fig. 4A and Fig. 4B. The performance of the cross-entropy linearly declines with gestation age ( $r = 0.765, p = 5.7^{-8}$ ) and in turn the volume of cortical plate, shown in Fig. 5A, a trend not observed in the Dice score. Cross-entropy is a count of correctly labelled voxels within the predicted volume, whereas the Dice coefficient measures the overlap between predicted and labelled cortical plate, and therefore they should both be independent of cortical plate size. However, as the cortical plate volume increases, the number of boundary voxels also increases, which notably are found to have the largest prediction uncertainty, as shown in Fig. 6A. A small number of incorrectly labelled boundary voxels will have a negligible effect on segmentation overlap (i.e. the Dice score) but will be significantly represented in cross-entropy loss.

### 4.3 Sylvian Fissure

To further evaluate our cortical plate segmentation, we measured the depth of the SF and compared the results to previous work [1]. For each of the 36 testing volumes, the transventricular axial plane was found by identifying the relevant anatomical landmarks and the SF depth was measured from the inner parietal bone to the cortical plate segmentation, using ITK-Snap [26] and shown in Fig. 5B.

The SF depth measurements compared to the range of depths found by [1] for similar gestational ages are shown in Fig. 5A. The depths measured from our segmentation follow the overall trend of increasing depth with gestational



**Fig. 6.** The result of propagating the cortical plate predictions back to the atlas. **A)** The averaged cortical plate volume produced for each gestational age (GA) and cerebral hemisphere in turn. Underneath the images show the number of volumes used to produce each atlas for the left and right hemispheres. **B)** The averaged cortical plate depth, produced by applying a threshold of  $t = 0.6$  to the volumes shown in **A**, before measuring the distance to the skull in 3D. The colour bar shows the depth in mm.

age and half of the volumes fell within the expected range, as exhibited by [1]. However, there is a cluster of measurement which express a higher depth to that previously measured. This could be a result of the difficulty in finding the transventricular axial plane in the ultrasound volumes. Due to shadowing and speckle throughout the volume the cavum septi pellucidum and posterior horns are not always visible, such shadowing is shown in Fig. 5B. Additionally, the parietal bone is not always well defined making the end point of measurement uncertain.

#### 4.4 Atlas Averages

For each gestational week studied, the predictions were propagated back to the common reference space,  $I_R$ , by applying the forward transformation vector,  $T_m$ , corresponding to each individual's volume,  $I_m$ . Once in the atlas space, it was possible to average over all predictions, for each gestational week and cerebral hemisphere, to illustrate the network's confidence across the volume, shown in Fig. 6A. Across the gestational weeks, the network confidently labelled the centre of the cortical plate while there was increased uncertainty at the boundaries.

This is not unexpected as the white matter boundary is not always apparent in the ultrasound volumes. One such region where the plate is not well defined is the anterior of the brain, which on Fig. 6A, corresponds to high uncertainty across the gestational weeks. The right hemisphere of gestational week 22 shows a high level of uncertainty in certain regions across both views. Out of all the age and hemisphere combinations, this had the smallest number of volumes during training (10/271) and only two test set volumes (2/36), therefore it is heavily influenced by variations within the small population.

A threshold of  $t = 0.6$  was then applied to the averaged cortical plate predictions and the depth between the parietal bone and the cortical surface was calculated, illustrated in Fig. 6B. As expected, with the increase of gestational age, the volume of the cortex also increases. Figure 6B illustrates the deepening of the SF with age which supports the findings in Fig. 5A. The 3D renderings also show that the SF is the deepest part of the cortical surface, which has commonly been reported in past literature for similar gestational ages [21].

## 5 Discussion and Conclusion

The work presented here looks at segmenting the developing fetal cortical plate from ultrasound volumes using neural networks, a task that has not been previously performed for this image modality or using CNNs. We investigated whether such a task was possible using the traditional U-Net, applied to a segmentation and regression task, before expanding to use a combination of the two in a MTL DMR network.

To test the capability of using neural networks for cortical plate segmentations, atlas-based label propagation, with manual corrections, was used to create the ground truth masks. Although this method produces high-quality segmentation, it was only possible due to previous work which used the same dataset to produce a population average atlas using a set of deformation fields [15]. In this work, we manually annotated the cortical plate in the atlas and propagated the labels back to the population of images used for the atlas construction. Therefore, although atlas-based label propagation can be used for high quality cortical plate segmentation, it requires a large set of labelled images along with manual corrections, whereas our proposed trained CNN method has the potential to be applied to new ultrasound volumes with no prior knowledge. However, the training of our network is intrinsically limited to the accuracy of label propagation.

Previous work had suggested MTL with DMR improved biomedical segmentation across a range of networks. However, for our biological structure we found the implementation of the DMR was outperformed by the standalone U-Net. We inferred that this could be due to the difficulty of the regression task. In [7] an MTL with a DMR was used for heart chamber segmentation, a solid volume generating a simpler distance transform map than ours, and perhaps making the DMR converge more easily. Moreover, the choice of MR, rather than ultrasound, may have also aided the authors regression task as it produces clearer image boundaries and well defined neighbouring structure which could help assign some initial distance measures.

We have been able to show that high quality segmentation of cortical plate can be achieved from ultrasound volumes using a simple encoder-decoder network, a method that is likely to work across imaging modalities. We demonstrated that the predicted results display similar sulcal depth characteristics, over a broad gestational age where notably morphological changes are present, to previous work, validating the accuracy of our method. In conclusion, our work has shown the initial steps towards characterising gyrification are possible using routinely collected ultrasound scans, laying the foundation for a clinically applicable diagnostic tool.

**Acknowledgements.** MW is supported by the Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC) [grant number EP/L016052/1]. MJ is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), and this research was funded by the Wellcome Trust [215573/Z/19/Z]. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust [203139/Z/16/Z]. AN is grateful for support from the UK Royal Academy of Engineering under the Engineering for Development Research Fellowships scheme.

## References

1. Chen, X., et al.: Ultrasonographic characteristics of cortical sulcus development in the human fetus between 18 and 41 weeks of gestation. *Chin. Med. J.* **130**(8), 920 (2017)
2. Chi, J.G., Dooling, E.C., Gilles, F.H.: Gyral development of the human brain. *Ann. Neurol.* **1**(1), 86–93 (1977). <https://doi.org/10.1002/ana.410010109>
3. Chung, Y.S., Hyatt, C.J., Stevens, M.C.: Adolescent maturation of the relationship between cortical gyrification and cognitive ability. *NeuroImage* **158**, 319–331 (2017)
4. Clouchoux, C., Guizard, N., Evans, A.C., Du Plessis, A.J., Limperopoulos, C.: Normative fetal brain growth by quantitative in vivo magnetic resonance imaging. *Am. J. Obstet. Gynecol.* (2012). <https://doi.org/10.1016/j.ajog.2011.10.002>
5. Clouchoux, C., et al.: Quantitative in vivo MRI measurement of cortical development in the fetus. *Brain Struct. Funct.* **217**(1), 127–139 (2012)
6. Corbett-Detig, J., et al.: 3D global and regional patterns of human fetal subplate growth determined in utero. *Brain Struct. Funct.* **215**(3–4), 255–263 (2011)
7. Dangi, S., Linte, C.A., Yaniv, Z.: A distance map regularized cnn for cardiac cine mr image segmentation. *Med. Phys.* **46**(12), 5637–5651 (2019)
8. Dubois, J., et al.: Mapping the early cortical folding process in the preterm newborn brain. *Cerebral Cortex* **18**(6), 1444–1454 (2008). <https://doi.org/10.1093/cercor/bhm180>
9. Fernández, V., Llinares-Benadero, C., Borrell, V.: Cerebral cortex expansion and folding: what have we learned? *The EMBO J.* **35**(10), 1021–1044 (2016). <https://doi.org/10.15252/embj.201593701>
10. Garel, C., et al.: Fetal cerebral cortex: normal gestational landmarks identified using prenatal mr imaging. *Am. J. Neuroradiol.* **22**(1), 184–189 (2001)
11. Guizard, N., Lepage, C., Fonov, V., Hakyemez, H., Evans, A., Limperopoulos, C.: Development of fetus brain atlas from multi-axial MR acquisitions. In: *Proceedings of the Sixteenth Annual Meeting of the International Society for Magnetic Resonance in Medicine*, vol. 672, p. 132 (2008)

12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
13. Lefèvre, J., et al.: Are developmental trajectories of cortical folding comparable between cross-sectional datasets of fetuses and preterm newborns? *Cereb. Cortex* **26**(7), 3023–3035 (2016). <https://doi.org/10.1093/cercor/bhv123>. <https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhv123>
14. Mangin, J.F., Lopez-Krahe, J.: From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. Technical report (1995)
15. Namburete, A.I.L., van Kampen, R., Papageorghiou, A.T., Papież, B.W.: Multi-channel groupwise registration to construct an ultrasound-specific fetal brain atlas. In: Melbourne, A., et al. (eds.) PIPPI/DATRA-2018. LNCS, vol. 11076, pp. 76–86. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00807-9\\_8](https://doi.org/10.1007/978-3-030-00807-9_8)
16. Namburete, A.I., Xie, W., Yaqub, M., Zisserman, A., Noble, J.A.: Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning. *Med. Image Anal.* **46**, 1–14 (2018)
17. Paladini, D., Malinger, G., Monteagudo, A., Pilu, G., Timor-Tritsch, I., Toi, A.: Sonographic examination of the fetal central nervous system: guidelines for performing the ‘basic examination’ and the ‘fetal neurosonogram’. *Ultrasound Obstet. Gynecol.* **29**(1), 109–116 (2007)
18. Papageorghiou, A.T., et al.: International standards for fetal growth based on serial ultrasound measurements: the fetal growth longitudinal study of the intergrowth-21st project. *Lancet* **384**(9946), 869–879 (2014)
19. Paszke, A., et al.: Automatic differentiation in PyTorch (2017)
20. Poon, L.C., et al.: Transvaginal three-dimensional ultrasound assessment of sylvian fissures at 18–30 weeks’ gestation. *Ultrasound Obstet. Gynecol.* **54**(2), 190–198 (2019)
21. Rajagopalan, V., et al.: Local tissue growth patterns underlying normal fetal human brain gyrification quantified in utero. *J. Neurosci.* **31**(8), 2878–2887 (2011)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation, May 2015. <http://arxiv.org/abs/1505.04597>
23. Striedter, G.F., Srinivasan, S., Monuki, E.S.: Cortical folding: when, where, how, and why? *Annu. Rev. Neurosci.* **38**(1), 291–307 (2015)
24. Sun, T., Hevner, R.F.: Growth and folding of the mammalian cerebral cortex: from molecules to malformations, April 2014. <https://doi.org/10.1038/nrn3707>
25. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv preprint [arXiv:1607.08022](https://arxiv.org/abs/1607.08022) (2016)
26. Yushkevich, P.A., et al.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**(3), 1116–1128 (2006)
27. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS-2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1)