

BEAN: Brain Extraction and Alignment Network for 3D Fetal Neurosonography

Felipe Moser^{a,*}, Ruobing Huang^b, the INTERGROWTH-21st Consortium^c,
Bartłomiej W. Papież^{b,d}, Ana I.L. Namburete^{a,e}

^a Oxford Machine Learning in Neuroimaging laboratory, OMNI, Department of Computer Science, University of Oxford, Oxford, UK

^b Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

^c Nuffield Department of Women's and Reproductive Health, John Radcliffe Hospital, University of Oxford, Oxford, UK

^d Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK

^e Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom

ARTICLE INFO

Keywords:

Fetal brain development
Convolutional neural network
Deep learning
3D ultrasound
Brain extraction
Brain alignment

ABSTRACT

Brain extraction (masking of extra-cerebral tissues) and alignment are fundamental first steps of most neuroimaging analysis pipelines. The lack of automated solutions for 3D ultrasound (US) has therefore limited its potential as a neuroimaging modality for studying fetal brain development using routinely acquired scans. In this work, we propose a convolutional neural network (CNN) that accurately and consistently aligns and extracts the fetal brain from minimally pre-processed 3D US scans. Our multi-task CNN, Brain Extraction and Alignment Network (BEAN), consists of two independent branches: 1) a fully-convolutional encoder-decoder branch for brain extraction of unaligned scans, and 2) a two-step regression-based branch for similarity alignment of the brain to a common coordinate space. BEAN was tested on 356 fetal head 3D scans spanning the gestational range of 14 to 30 weeks, significantly outperforming all current alternatives for fetal brain extraction and alignment. BEAN achieved state-of-the-art performance for both tasks, with a mean Dice Similarity Coefficient (DSC) of 0.94 for the brain extraction masks, and a mean DSC of 0.93 for the alignment of the target brain masks. The presented experimental results show that brain structures such as the thalamus, choroid plexus, cavum septum pellucidum, and Sylvian fissure, are consistently aligned throughout the dataset and remain clearly visible when the scans are averaged together. The BEAN implementation and related code can be found under www.github.com/felipemoser/kelluwen.

1. Introduction

Accurate brain extraction and alignment are crucial for reliable neuroimaging pipelines, with fetal neuroimaging being no exception. Improperly aligned images can make inter- and intra-subject comparisons a challenge, and potentially misrepresent structural differences. Similarly, an accurate brain extraction (i.e., binary masking of the brain) ensures that extra-cerebral structures such as the skull, amniotic fluid, and maternal tissues do not affect any subsequent analysis, with the added benefit of improving performance by reducing the amount of processed information. However, manual alignment and extraction of brain images are tedious, time-consuming tasks that require a high level of expertise, especially for three-dimensional (3D) images. The need for automated solutions to these tasks has resulted in several algorithms being proposed for more traditional 3D neuroimaging modalities such as Mag-

netic Resonance Imaging (MRI) (Klein et al., 2009, 2010; Makropoulos et al., 2018). However, although ultrasound (US) is the preferred imaging modality for prenatal care and fetal growth assessment, there is a lack of development of such solutions for 3D US, limiting its potential as a neuroimaging modality. Therefore, in this work we propose an automated solution specifically designed for the tasks of automated fetal brain alignment and extraction from 3D US scans. While MRI is arguably the most widely used neuroimaging modality and therefore has had the largest amount of development for automated fetal brain alignment and extraction (Makropoulos et al., 2018; Studholme, 2011; Torrents-Barrena et al., 2019), this imaging modality is not used as part of the routine clinical practice during pregnancy. Although MRI is generally considered safe for fetal imaging (Ray et al., 2016), patients are typically referred for a secondary MRI scan for diagnostic confirmation, only if abnormalities are suspected during the routine second-trimester anomaly US scan (Prayer et al., 2017). In contrast, the use of US to monitor and assess the development of the fetus during pregnancy is standard clinical practice worldwide (Haratz and Lerman-Sagie, 2018; Kim et al., 2008; Namburete et al., 2018), and it is used during the routine clinical

* Corresponding author.

E-mail address: felipe.moser@univ.ox.ac.uk (F. Moser).

<https://doi.org/10.1016/j.neuroimage.2022.119341>.

Received 27 September 2021; Received in revised form 8 April 2022; Accepted 28 May 2022

Available online 30 May 2022.

1053-8119/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

examination to assess the anatomical integrity of the brain throughout gestation.

While the clinical adoption of 3D US over 2D US has been relatively slow due to higher costs and limited perceived benefits (Prager et al., 2010), current research has highlighted its potential, and the advantages of this modality have resulted in its recommendation by the updated International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) Practice Guidelines of 2021 (Paladini et al., 2021).

The ubiquity of US and availability of large 3D US datasets, such as the INTER-GROWTH-21st study (Papageorgiou et al., 2014) offer the possibility to study and assess the development of the brain from its early stages up until birth. However, the difficulty of analyzing 3D US brain data has limited its clinical use to 2D assessment, in which biometric measurements are manually collected for growth monitoring, or in which structures are qualitatively assessed (Paladini et al., 2007). This, in turn, severely limits the amount of anatomical information on which the assessment is performed. Additionally, 3D fetal neurosonograms can be used to perform substantially more complex tasks, such as automated extraction and evaluation of oblique planes (Carneiro et al., 2008; Huang et al., 2015; Li et al., 2017; 2018; Yaqub et al., 2015), anatomy-based gestational age prediction (Namburete et al., 2015, 2017; Wyburd et al., 2021), and 3D segmentation of fetal brain structures for volumetric assessment (Becker et al., 2010; Cerrolaza et al., 2017; Gutiérrez-Becker et al., 2013; Hesse et al., 2022; Huang et al., 2018; Velásquez-Rodríguez et al., 2015; Yaqub et al., 2013). A reliable, automated solution to fetal brain extraction and alignment would facilitate further development of such algorithms as well as their potential implementation for clinical assessment.

While clinical guidelines for acquisition of 3D neurosonograms have been proposed in order to minimize positional variability of the brain between scans and facilitate analysis (Paladini et al., 2007; Pilu et al., 2007), the results rely heavily on the expertise of the sonographer and cannot always account for varying fetal pose during image acquisition. This results in an improved but still highly inconsistent brain alignment between fetal scans. A reliable and accurate solution for the tasks of brain alignment and extraction would therefore be paramount for 3D neurosonograms to become standard clinical practice.

To date, little work has been published for the automated extraction and alignment of the fetal brain from 3D US scans (Cerrolaza et al., 2017; Chen et al., 2012; Cuingnet et al., 2013; Kuklisova-Murgasova et al., 2013; Moser et al., 2020; Namburete et al., 2018). This can be attributed, in part, to the intrinsic challenges associated with this imaging modality, as well as the effect that temporal changes of the fetal head have on image quality. Specifically, during gestation, the brain grows in size and the internal structures undergo morphological changes (Prayer et al., 2006), causing a large intrinsic variability in the fetal cohort. The increasing ossification of the skull during gestation affects the interaction of the tissues with the US beam, resulting in a variation of shadows, occlusions, and reverberation artifacts throughout pregnancy which vary depending on the relative orientation and position of the probe relative to the skull. Finally, due to the acoustic cavity created by the increasingly calcified skull, the hemisphere of the brain that is located closest to the US probe has most of its structural information obscured, resulting in an intrinsically asymmetric representation of the brain. All of these challenges make the development of a reliable and accurate automated solution for brain alignment and extraction difficult.

Deep Learning (DL)-based approaches proposed for medical imaging have increased dramatically (Suzuki, 2017), with the effectiveness of networks such as the U-Net (Ronneberger et al., 2015) and Spatial Transformer Networks (Jaderberg et al., 2015) having a profound impact in the development of segmentation and alignment solutions for medical data, respectively. Specifically, DL solutions have shown great potential for processing US image data, obtaining accurate predictions in tasks such as standard plane detection (Baumgartner et al., 2017; Chen et al., 2015; Yaqub et al., 2017) and fetal structure detection, localization, and segmentation (Gutiérrez-Becker et al., 2013a), in spite

of the aforementioned image quality challenges. This makes DL an ideal candidate for developing a generalized solution to the automated brain alignment and extraction of 3D neurosonograms.

In this work, we build upon our preliminary results for automated fetal brain extraction from 3D neurosonograms (Moser et al., 2020) and propose an end-to-end, multi-task Convolutional Neural Network (CNN) that extracts and aligns the fetal brain from 3D US scans. Our multi-task CNN, Brain Extraction and Alignment Network (BEAN), consists of two independent networks: 1) a fully-convolutional, encoder-decoder network for brain extraction of unaligned scans, and 2) a cascade regression-based network for rigid alignment of the brain to a common coordinate space. These concurrent but independent networks result in a fast, accurate, and consistent alignment and extraction of the fetal brain from minimally pre-processed 3D US scans.

In summary, contributions proposed in this work are:

- We adapt the fully-convolutional fetal brain extraction network published in Moser et al. (2020) to predict masks for 160x160x160 with a spline upsampling step, which results in better performance without significant computational cost.
- We propose a fully-automated alignment regression network with a cascade architecture that accurately aligns the fetal brain from minimally pre-processed 3D US scans to a common coordinate space.
- We combine both networks into a single, end-to-end, Brain Extraction and Alignment Network (BEAN), that returns the extracted and aligned brain from a minimally-preprocessed 3D US scan, aiming to help facilitate the development of neuroimage analysis pipelines for intrauterine brain development.

2. Related works

The importance of reliable and accurate brain extraction and alignment for neuroimage analysis has led to an extensive list of methods being proposed for MR imaging (Eskildsen et al., 2012; Guha Roy et al., 2018; Han et al., 2018; Huo et al., 2019; Isensee et al., 2019; Kleesiek et al., 2016; Klein et al., 2009; 2010; Mohseni Salehi et al., 2017; Nestares and Heeger, 2000), with libraries and toolboxes such as FSL (Jenkinson et al., 2012) and BrainVISA (Cointepas et al., 2001) being commonplace in most pipelines. In particular, there is an extensive list of methods proposed for fetal brain extraction and alignment from 3D MRI images. In contrast, considerably fewer solutions have been proposed for fetal brain 3D US scans, in spite of the increased use of 3D US during routine clinical assessment and the recent endorsement of this modality by ISUOG (Paladini et al., 2021). Therefore, in order to provide a richer technical context and insight to our work, in this section we summarize relevant works proposed for both imaging modalities. However, it is important to note that MRI and US images are acquired through entirely different physical interactions, which results in differences in the structural information rendered, as well as other intensity properties, such as contrast and sharpness. Therefore, while a method that works well for MRI could potentially be adapted to work for 3D US, such an implementation is non-trivial and the lack of one-to-one intensity mapping between US and MRI would potentiate the need for major modifications to most intensity-based approaches.

2.1. Brain extraction: 3D US

Several publications have proposed methods for automated segmentation of cerebral structures from 3D US (Becker et al., 2010b; Gutiérrez-Becker et al., 2013b; Rodríguez et al., 2015; Velásquez-Rodríguez et al., 2015b; Yaqub et al., 2013; 2012; Hesse et al., 2022). However, there is currently very limited work published on the topic of automated fetal brain extraction from 3D US scans. Cerrolaza et al. (2017) proposed using a structured geodesic random forest approach that implements structured labels and semantic features for automated segmentation of the fetal skull. While this solution showed good performance, it resulted in a significantly lower sensitivity when compared to a classic

non-structured random forest method, as well as a U-Net based CNN method. In a subsequent study (Cerrolaza et al., 2018), their improved method for 3D skull segmentation used a two-step CNN approach. The first CNN performed an initial segmentation of the skull, which was then used to derive an Incidence Angle Map (IAM) and a Shadow Casting Map (SCM), which provided relevant complementary information of the underlying physics of US image acquisition. The initial segmentation was then passed along with the two derived maps as input for the second CNN, which returned the final segmentation.

Namburete et al. (2018) proposed a multi-task fully convolutional neural network that addressed brain localization, segmentation, and alignment to a canonical coordinate system. The alignment task of this work is discussed in Section 2.2. The brain segmentation task consisted of predicting the fetal brain segmentation of 2D US image slices with a CNN network. These predictions were then stacked together, creating a 3D volume to which an ellipsoid was fitted as an approximation of the fetal brain. However, although this network showed good performance in their testing, it has two important limitations that affect its potential. Firstly, by predicting the brain segmentations from 2D images, the 3D context of the original volume is lost. This context is important information that could be used to improve the segmentation. Secondly, the ellipsoid approximation of the shape of the brain does not represent its actual morphology and will therefore always include background voxels in its brain segmentation.

To the best of our knowledge, our preliminary work (Moser et al., 2020) is currently the only approach that directly extracts the brain from fetal 3D US scans, without relying on shape approximations. The work presented here is an extension of that network.

2.2. Brain alignment: US

Similarly to the task of fetal brain extraction from 3D US, there is very limited work published on the topic of automated fetal brain alignment from 3D US. Chen et al. (2012) proposed a method based on a coarse-to-fine strategy for fetal head registration and segmentation. The eye of the fetus was first detected through Gabor Features (Kamarainen, 2012) to identify head-pose, which was then followed by registering a reference model of the fetal head to the imaged head. However, the reliance on a reference model, as well as the presence of strong edges in the image strongly limit its implementation. It is important to note that this pipeline required the user to manually select a Volume Of Interest (VOI) in the original scan before the registration takes place.

Circumventing the need for user intervention, Cuingnet et al. (2013) proposed a multi-step approach for automated alignment of the fetal brain from 3D US scans. This relied on an anisotropic plate diffusion filter, a spheroidal shell approximation of the general shape of the skull, weighted Hough transform to detect the mid-sagittal plane, and a minimum intensity gradient across the skull surface to detect the neck location. This method was tested on an age range of 19–24 gestational weeks. The performance of this approach was evaluated by computing the maximal distance inside the skull between three ground-truth planes (transthalamic, transventricular, and transcerebellar planes) and the ones obtained from their alignment. However, this approach is strongly limited by the reliance on a clear view of the neck and eye, limits its implementation in cases where these structures are occluded or affected by acoustic shadows.

As mentioned in Section 3.1.3, Namburete et al. (2018) proposed the use of a multi-task fully Convolutional Neural Network (CNN) that addressed brain alignment to a referential coordinate system. The alignment was achieved by defining a parametric coordinate system based on skull boundaries, location of the eye sockets, and head pose. The network performed 5 tasks for each 2D axial slices of a 3D US scan: 3 classification tasks (slice located near crown or neck, ante-posterior orientation of the head, and presence or absence of eye in the slice) and 2 segmentation tasks: eye segmentation, and brain segmentation. The

transformation matrix for rigid alignment of the brain was then calculated by multiplying four separate matrices based on the combined prediction of the stacked 2D predictions of the 5 tasks. This was tested on a gestational age range of 21 to 30 weeks. However, one strong limitation of this approach is the loss of 3D context for the predictions, since they are done for each 2D slice independently. This also means that there is a need for a post-processing step to fuse the network prediction results to estimate the transformation matrix for alignment. The transformation matrix for the alignment is calculated based on these 2D segmentations, causing any discrepancies between slices to impact the alignment result. Similarly to Cuingnet et al. (2013), the reliance on a clear view of the eye also limits its implementation.

More recently, Wright et al. (2019) proposed using the LSTM spatial co-transformer solution proposed in Wright et al. (2018) to co-align images of the fetal head to a canonical pose, showing remarkable results. The network performs an initial alignment, which is then iteratively refined by groupwise registration with a saliency-weighted compounded volume. The predictions are performed independently for each image, allowing for the alignment of single scans. However, analogous to Wright et al. (2018), the network relies on multiple views of the same subject for training, which is not always available.

Also relevant to US brain alignment are the works of Kuklisova-Murgasova et al. (2013) and Wright et al. (2018) which are described in Section 2.4.

2.3. Brain extraction: MRI

Several solutions have been proposed for the task of fetal brain extraction from MRI scans. Anquez et al. (2009) proposed using a template-matching approach to locate the fetal eye, subsequently segmenting the surrounding skull bone using a graph cut approach. However, this method relies on the eye structure to be clearly visible inside the scan, as well as a clear contrast between the skull and the brain. Automated solutions that rely on brain templates have also been proposed (Gholipour et al., 2012; Habas et al., 2010; Taimouri et al., 2015; Taleb et al., 2013; Tourbier et al., 2017; Wright et al., 2014). While these works showed good results for fetal brain segmentation of MRI scans, such solutions tend to struggle on scans that show a large deviation from the average, such as structurally abnormal and pathological cases.

Machine Learning (ML) solutions have also been explored for fetal brain extraction from MRI scans. Kainz et al. (2014) proposed a Random Forest classifier trained on 3D Gabor descriptors, with a subsequent segmentation refinement using a 2D level-set. Keraudren et al. (2014) used Scale-Invariant Feature Transform (SIFT) features for localization clustered with k-means, and subsequently classified with a Support Vector Machine (SVM) classifier and refined using a RF classifier on patches of the 2D slices.

More recently, the high performance of CNNs for this task has been thoroughly demonstrated by the works of Rajchl et al. (2016b), Rajchl et al. (2016a), Mohseni Salehi et al. (2017), Khalili et al. (2017), Khalili et al. (2019), Ebner et al. (2020), and Ranzini et al. (2021). These solutions rely on an initial 2D brain segmentations from each slice using CNNs, which can then combined into a 3D volume. Although such an approach results in a more compact network, it greatly limits the contextual information available to the network. However, in the case of fetal brain MRI segmentation, the larger inter-slice spacing and potential motion corruption between neighbouring slices make this a compelling trade-off (Ebner et al., 2020).

2.4. Brain alignment: MRI

In the case of fetal MRI, research is heavily focused on the reconstruction of the 2D slices into a 3D volume rather than alignment of the 3D brain itself due to the nature of how the MRI scans are generated (Ebner et al., 2020; Hou et al., 2018; Miao et al., 2016). While fast imaging methods such as Single-Shot Fast Spin Echo (SSFSE) can

minimize the effects of fetal motion in-plane, motion-corruption is still a common issue between slices of the stack. In contrast, while motion artifacts can still occur when acquiring 3D US scans, the shorter acquisition time of this imaging modality strongly minimizes their occurrence. The acquisition can also simply be repeated if such artefacts are indeed noticed by the sonographer. A reconstruction of 2D slices into a 3D volume is therefore not needed for 3D US. However, there are two particular methods that particularly relevant for this work. Kuklisova-Murgasova et al. (2013) proposed a method for registration of 3D US to a 3D MRI image by creating a simulated US image from the MRI data, on which they used block-matching and normalized cross-correlation to align it to the US volume. The registration was divided into two age-specific groups: one for 18 to 22 gestational weeks, and one for 29 gestational weeks. To assess the performance for the younger group, 4 brain structures were manually segmented in both US and MRI scans and compared to registration using label consistency as a similarity measure and gradient descent as the optimization method. Since the same structures were not visible in the older group, 5 other structures were manually segmented and the mean Target Registration Error (mTRE) of these labels after registration was calculated. Wright et al. (2018) proposed a LSTM spatial co-transformer for group-wise co-registration of MRI and 3D US scans to a canonical pose. This method refines the alignment by iteratively predicting the residual transformations and subsequently warping the images to update them.

2.5. Proposed method

For this work we aimed to develop a network that could accurately extract and align the brain from 3D US with a minimum amount of pre- or post-processing, while yielding state-of-the-art performance. This led us to design a network that performs each of these tasks independently, allowing for the optimal network architecture to be used for each. The brain extraction step (the first half of our proposed BEAN network), was inspired by the U-Net (Ronneberger et al., 2015), which has demonstrated to be a reliable, flexible, and accurate approach for image segmentation in a multitude of image modalities. As for the brain alignment step (the second half of BEAN), our initial source of inspiration was the Spatial Transformer Networks work of (Jaderberg et al., 2015), since we wanted a simple yet robust solution to the challenging task of automated alignment of the fetal brain. Their work was an essential starting point for BEAN, where we've modified their architecture to allow for serial transformations to be performed on the original input, minimizing the loss of information caused by the intermediate resampling steps.

3. Methods

3.1. Network architecture

Our proposed BEAN network, as shown in Fig. 1, is sub-divided into two networks: one for the task of brain extraction and the other for brain alignment. While these networks receive the same input and operate concurrently, they operate entirely independently from each other and the results of one have no impact on the results of the other.

3.1.1. Notation

To facilitate the understanding of our equations and results, we introduce the notation used for the volumes and parameters involved. We denote the original 3D US scan as $S \in \mathbb{R}^3$ and the ground-truth brain mask as $M \in (\mathbb{R})^3$, with the value of their individual voxels described as $S(x, y, z)$ and $M(x, y, z)$, respectively. Similarly, the ground-truth alignment parameters are represented by p , with individual parameters shown as p_i . Their predicted counterparts are denoted by \hat{M} and \hat{p} , respectively. Finally, if a scan or brain mask has been aligned, the parameters used will be noted as a superscript. For example, the ground-truth mask aligned with the predicted alignment parameters is

represented by $M^{\hat{p}}$, while the predicted mask aligned with the ground-truth parameters is \hat{M}^p .

3.1.2. Network pipeline

The BEAN network takes in the original 3D US scan S as input and it subsequently passes it as input to the *Extraction* and *Alignment* components, as shown in Fig. 1. The predicted brain mask \hat{M} as well as the predicted brain alignment parameters \hat{p} are then passed as input, along with the original scan S into the BEAN transformation layer \mathcal{T} . In this layer, the predicted parameters \hat{p} are used to calculate the corresponding similarity transform, which is then used to align the predicted brain mask ($\hat{M}^{\hat{p}}$) as well as the original scan ($S^{\hat{p}}$) to the common coordinate space. Finally, the aligned scan $S^{\hat{p}}$ is multiplied with the aligned mask $\hat{M}^{\hat{p}}$, resulting in the masked and aligned scan $S^{\hat{p}}_{\hat{M}}$. We note that the input volumes S are scaled down from $160 \times 160 \times 160$ to $80 \times 80 \times 80$ inside the networks before being passed to their first convolutional layers, as shown in Fig. 1. This drastically reduces the memory requirements by 87% for both the *Extraction* and *Alignment* networks, allowing for training on a single consumer-grade GPU (Nvidia GTX 1080 Ti). However, the prediction of both networks are for the original size of $160 \times 160 \times 160$. The *Extraction* network performs a final spline up-sampling on the predicted mask, which is then binarised, while parameters predicted by the *Alignment* network are independent of size, since the translation is normalised, as described in Section 3.1.4. All results and figures contained in this work have been generated at the $160 \times 160 \times 160$ scale.

3.1.3. Brain extraction network

The brain extraction network is based on an encoder-decoder architecture that inputs the original scan S and outputs a predicted binary volume \hat{M} containing the extracted brain. The encoder is comprised of an AvgPool Layer that downsamples the input scan from $160 \times 160 \times 160$ to $80 \times 80 \times 80$, followed by 5 Convolutional Blocks (see Fig. 1) with a $2 \times 2 \times 2$ MaxPooling Layer between them. The decoder consists of 4 Convolutional Blocks, with a $2 \times 2 \times 2$ Upsampling Layer and concatenating skip connections between them, as shown in Fig. 1. A Sigmoid Layer of kernel size $1 \times 1 \times 1$, predicts the brain probability maps. These are then upsampled back to the original size of $160 \times 160 \times 160$ using a spline interpolation, and subsequently a threshold of 0.5 returns the predicted binary mask \hat{M} of the brain inside the original 3D US scan. All 3D Convolutional Layers of the extraction network have a kernel size of $3 \times 3 \times 3$ with 8, 16, 32, 64, 64, 64, 64, and 64 output channels, respectively. The only differences between this network and the one published in our preliminary results in Moser et al. (2020) are the downsampling and upsampling steps.

3.1.4. Brain alignment network

The brain alignment network consists of two cascading regression subnetworks that input the original scan S and output the predicted transformation parameters needed to align the input to a common coordinate space. These parameters p consist of three Euler angles p_α , p_β and p_γ , three Cartesian shift values p_x , p_y and p_z , normalized to $[-0.5, 0.5]$, and a scaling factor p_s . We have chosen a single scaling factor for all 3 dimensions of the volume in order to keep the voxel spacing isotropic and preserve the aspect ratio. Each subnetwork has the same architecture, with an AvgPool Layer that downsamples the input scan from $160 \times 160 \times 160$ to $80 \times 80 \times 80$, followed by 5 Convolutional Blocks (see Fig. 1) and two Fully Connected Layers (FCL). The 3D Convolutional Layers of both subnetworks have a kernel of size $3 \times 3 \times 3$ and 32, 64, 128, 256, and 512 output channels respectively. The output of the final Convolutional Block is passed to two FCL with 256 and 7 output channels respectively. The output of the first subnetwork (Coarse) are the global similarity transform parameters \hat{p}_c , which are then passed to the Coarse transformation layer \mathcal{T}_c along with the original scan S . Inside \mathcal{T}_c , a volume-centered similarity transform matrix is calculated using \hat{p}_c , which is subsequently applied to S , resulting in the coarsely

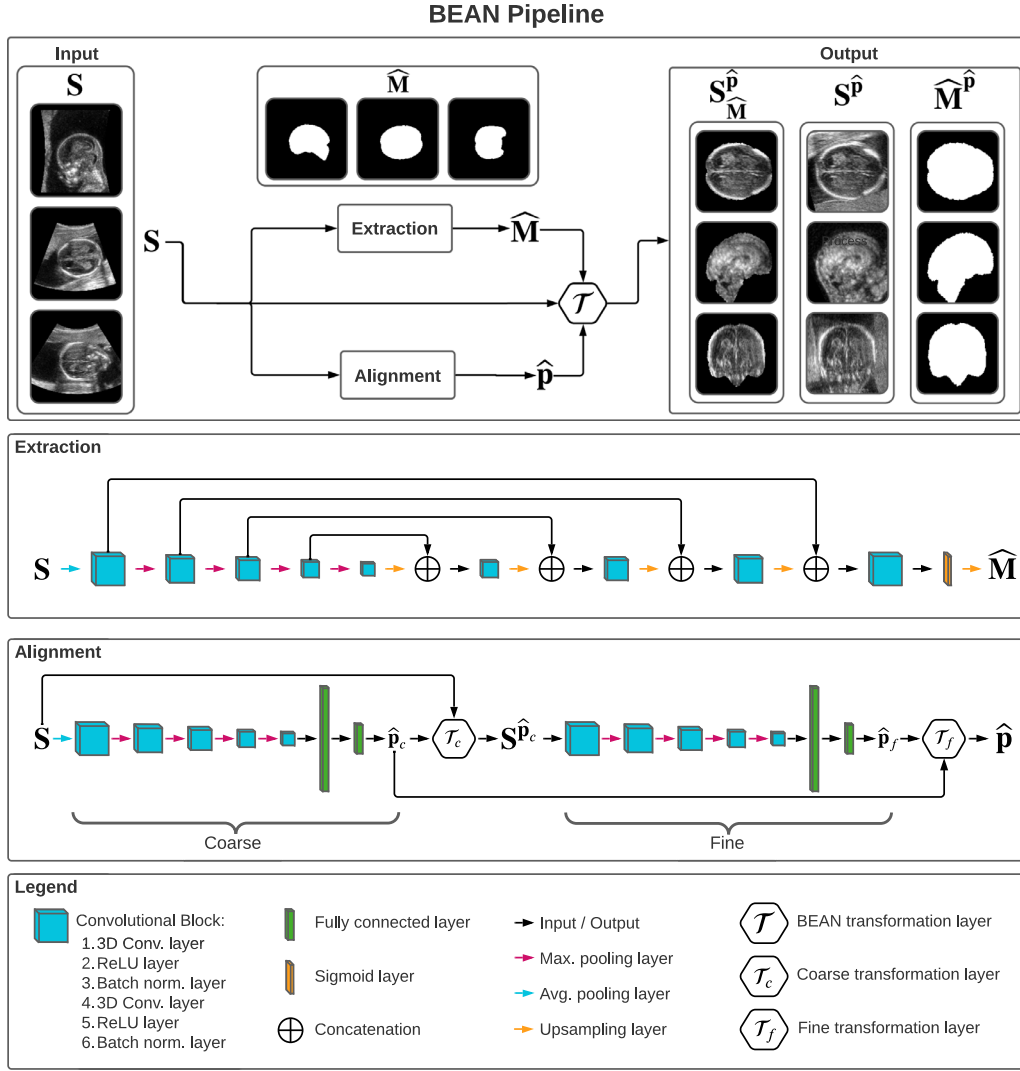


Fig. 1. Schematics of the BEAN network and its pipeline. The input S is a $160 \times 160 \times 160$ 3D US scan of the fetal brain, which is passed as input to the Extraction and Alignment networks, resulting in the predicted brain mask \hat{M} and brain alignment parameters \hat{p} . The BEAN transformation layer \mathcal{T} combines S , \hat{M} , and \hat{p} , returning the aligned brain mask $\hat{M}^{\hat{p}}$, the aligned scan $S^{\hat{p}}$, and the masked and aligned scan $\hat{S}_M^{\hat{p}}$. The specific architecture of the Extraction and Alignment networks is shown below the BEAN pipeline. For each volume, the orthogonal midplanes of an example scan are shown. Note that the perceived change in the order of midplanes between input and output is due to the differences between the original position of the fetal brain as acquired by the sonographer, and the canonical coordinate space.

aligned scan $\hat{S}^{\hat{p}_c}$. This $\hat{S}^{\hat{p}_c}$ is then passed as input to the second sub-network (Fine), which predicts the alignment parameters \hat{p}_f required to refine its alignment to the common coordinate space. Finally, \hat{p}_f is passed as input to the Fine transformation layer \mathcal{T}_f , along with \hat{p}_c . Here, the volume-centered similarity transforms calculated from \hat{p}_c and \hat{p}_f are multiplied, resulting in the final similarity transform, which is ultimately decomposed into the predicted alignment parameters \hat{p} .

3.2. 3D US Data

A total of $n = 1185$ 3D US scans of the fetal brain were used for training (829) and testing (356) BEAN. These were obtained from the INTERGROWTH-21st (Papageorgiou et al., 2014) study dataset and represent the gestational range of 14.4 to 30.9 gestational weeks (GW) distributed as shown in Fig. 2. The 3D scans were obtained in eight sites in eight countries using a Philips HD-9 ultrasound machine with a V7-3 curvilinear abdominal transducer. The volumes were re-sampled to an isotropic voxel size of 0.6 mm in each dimension, and subse-

quently centre-cropped to a size of $160 \times 160 \times 160$ voxels. For reference, the median dimensions of the original scans were $444 \times 254 \times 111$ (IQR= $59 \times 88 \times 36$) while their median voxel size was [0.32, 0.51, 0.85] mm (IQR=[0.05, 0.15, 0.28] mm).

3.3. Data annotation

In order to train the proposed BEAN network, a separate set of ground-truth labels were generated for the brain extraction and alignment components of the framework.

To generate the ground-truth alignment parameters, each brain was manually aligned to a canonical coordinate space using a custom-built tool, as shown in Fig. 3a. The brain was iteratively translated and rotated until the mid-sagittal, mid-coronal, and mid-axial planes corresponded to the central orthogonal planes of the volume. Finally, each brain was isometrically scaled to match the average brain size at 30 GW. The translation, rotation, and scaling transforms applied in each step were multiplied together, and the resulting transform was deconstructed into the

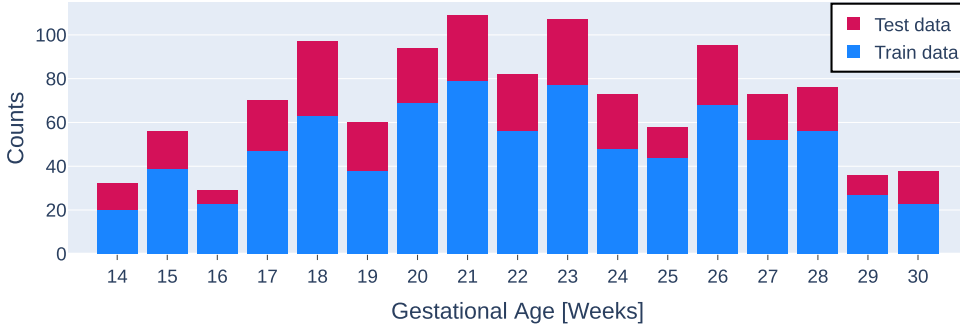


Fig. 2. Histogram of the distribution of the gestational age of the dataset (Papageorgiou et al., 2014), separated into training, and testing data.

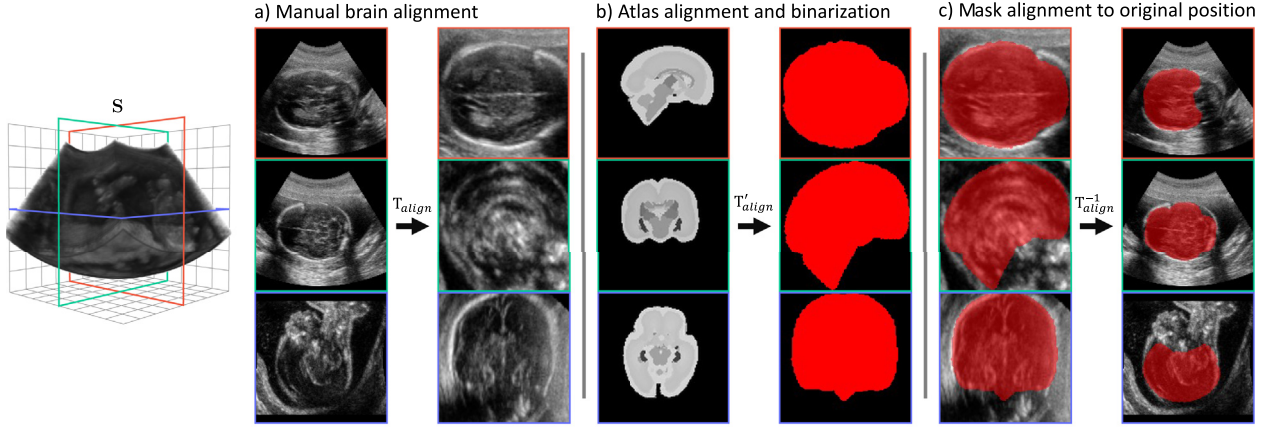


Fig. 3. Demonstration of the process of data labeling. First, the individual 3D US scans are manually aligned with to a common canonical space, as shown in a), in order to obtain the ground-truth alignment parameters. This alignment is performed with the transform T_{align} . Then, as shown in b), the spatiotemporal atlases (tissue) from Gholipour et al. (2017) are binarised and aligned to the same canonical space as the scans (T'_{align}). Finally, c) shows that the corresponding binary brain masks are aligned to the original position of the fetal brain for each GW using the inverse transformation T_{align}^{-1} from the ground-truth parameters.

translation parameters p_x , p_y , and p_z , the rotation parameters p_a , p_θ , and p_γ , and the scaling parameter p_s . These were set as ground-truth parameter labels \mathbf{p} for the brain alignment, which are used to generate the corresponding translation, rotation, and scaling transforms T_T , T_R , and T_S , respectively, that make up the volume-centered alignment transform T_{align} as shown in Eq. (1).

$$T_{align} = T_T \cdot T_R \cdot T_S \quad (1)$$

Due to the complexity of visually assessing 3D US images, manual labeling of the brain extraction masks by a clinician would be extremely time-consuming and require a high degree of expertise. In order to circumvent this problem, we used the fetal brain atlas labels generated by Gholipour et al. (2017) as templates. These atlases were generated using MRI images and represent the average brain shape for each GW in the range of 21 to 38 weeks. An example of these atlases is shown in Fig. 3(b). We binarised the atlas to convert it into a brain mask and we manually aligned each mask to the same coordinate space as the 3D US scans, as shown in Fig. 3(b). Finally, we aligned each brain mask to the 3D US scans of the same gestational age by transforming it using the inverse of the corresponding alignment similarity transform T_{align}^{-1} as shown in Fig. 3(c). These binary masks of the brain in their original locations are the ground-truth brain extraction labels \mathbf{M} .

3.4. Implementation details

BEAN was implemented in Python using the Tensorflow (v2.8) (Abadi et al., 2015) and Keras (v2.8) (Chollet et al., 2015) libraries. The networks were developed on an Intel Xeon E-2146G CPU (3.50GHz, 6 cores) and an Nvidia GTX 1080 Ti. Additionally, the Numpy (v1.21.2) (Harris et al., 2020), Pandas

(v1.3.5) (McKinney et al., 2010), Scipy (v1.7.3) (Virtanen et al., 2020), Scikit-learn (v1.0.1) (Pedregosa et al., 2011), Matplotlib (v3.5.1) (Hunter, 2007), and Plotly (v5.1.0) (Inc., 2015) libraries were used for analysis and figure generation. The *Extraction* network trained in 7 hours, and had an inference time of 0.17s per volume. The complete (Alignment) network trained in 12 hours, and had an inference time of 0.027s per volume.

3.5. Training

In order to train and test BEAN with representative data of the entire dataset, we randomly selected 30% (356 volumes) for a holdout testing dataset. We performed a three-fold cross-validation with the remaining data, splitting them into 70% (580 volumes) for training and 30% (249 volumes) for validation. Note that the random selection of scans was performed proportionally to the gestational ages in order to ensure that the training, validating, and testing datasets represent the entire gestational range of the original dataset, with a Kolmogorov-Smirnov p-value > 0.6 for all data splits. The best model was then retrained with the entire dataset and its performance was evaluated with the holdout testing dataset. The same split was used to train the *Extraction* and *Alignment* components of BEAN.

The *Extraction* network was trained end-to-end for 200 epochs saving the best version, with a batch size of 4, the Adam optimizer, and an empirically selected learning rate of 10^{-3} . The loss function \mathcal{L}_{ext} was a based on the Dice Similarity Coefficient (DSC) described in Section 4, and is shown in Eq. (2).

$$\mathcal{L}_{ext} = \mathcal{L}_{DSC}(\mathbf{M}, \hat{\mathbf{M}}) = 1 - \text{DSC}(\mathbf{M}, \hat{\mathbf{M}}) \quad (2)$$

The *Alignment* network was trained in two steps. First, the *Coarse Alignment* was trained for 200 epochs, with a batch size of 4, the Adam optimizer, and an empirically selected learning rate of 10^{-4} . The loss function \mathcal{L}_{align} was the sum of the Mean Absolute Error (MAE) between the predicted alignment parameters $\hat{\mathbf{p}}$ and the ground-truth parameters \mathbf{p} , and the \mathcal{L}_{DSC} of the ground-truth brain extraction mask \mathbf{M} aligned with \mathbf{p} and $\hat{\mathbf{p}}$, $\mathbf{M}^{\mathbf{p}}$ and $\mathbf{M}^{\hat{\mathbf{p}}}$, respectively. This is described in Eq. (3) with $w = 7$ denoting the number of alignment parameters.

$$\begin{aligned}\mathcal{L}_{align} &= \mathcal{L}_{MAE}(\mathbf{p}, \hat{\mathbf{p}}) + \mathcal{L}_{DSC}(\mathbf{M}^{\mathbf{p}}, \mathbf{M}^{\hat{\mathbf{p}}}) \\ &= \frac{1}{w} \sum_{i=1}^w |p_i - \hat{p}_i| + \mathcal{L}_{DSC}(\mathbf{M}^{\mathbf{p}}, \mathbf{M}^{\hat{\mathbf{p}}})\end{aligned}\quad (3)$$

After training the *Coarse Alignment*, the *Fine Alignment* was trained in the same manner but with two important differences. Firstly, the trained weights of the *Coarse Alignment* were transferred to the *Fine Alignment* as a warm start to help it converge faster. Secondly, the input data is now the original scan aligned with the predictions from the *Coarse Alignment* $\hat{\mathbf{p}}_c$, and the predicted parameters $\hat{\mathbf{p}}_f$ aim to finish aligning this input to the original scan aligned with \mathbf{p} .

3.6. Ethics statement

The 3D US scans used in this work were obtained from the INTERGROWTH-21st study, which was approved by the Oxfordshire Research Ethics Committee 'C' (ref: 08/H0606/ 139), as well as the research ethics committee and corresponding regional health authorities of the participating institutions. Written consent to be involved in the project was provided by all participants Villar et al. (2013).

4. Evaluation

In order to fully analyze the accuracy and reliability of BEAN, we define seven measures that will describe the performance of the predictions from the network against the ground-truth.

Statistical Significance

To assess the statistical significance when comparing these measures, we first determined the normality with a D'Agostino and Pearson's test d'Agostino (1971), followed by a paired Student's *t*-test Student (1908) or a Wilcoxon signed-rank test Wilcoxon (1992), for normal and non-normal samples, respectively. We used a significance threshold of $p < 0.05$ for both tests.

Centroid Distance

The Centroid Distance (CD) measures the Euclidean distance between the centroids of two binary volumes, with an ideal value of 0. For the *Extraction* network, it is calculated between the predicted mask $\hat{\mathbf{M}}$ and ground-truth \mathbf{M} , and it helps evaluate how accurately the predicted mask locates the brain in the scan S . A more detailed description can be found in Appendix A.

Dice Similarity Coefficient

The Dice Similarity Coefficient (DSC) gives a measure of the overlap between the predicted brain mask and the ground-truth, with an ideal value of $DSC = 1$ if the overlap is perfect, and $DSC = 0$ if there is no overlap at all. A more detailed description can be found in Appendix A. For the *Extraction* network, analogous to Eq. (2), the DSC is calculated between the predicted brain mask $\hat{\mathbf{M}}$ and the ground-truth mask \mathbf{M} . For the *Alignment* network, the DSC is calculated between $\mathbf{M}^{\mathbf{p}}$ and $\mathbf{M}^{\hat{\mathbf{p}}}$ instead.

Hausdorff Distance

One of the limitations of DSC is that since it is a global measure, local discrepancies are often overlooked. To determine whether the predicted mask is consistent throughout the brain or if there are specific areas of discrepancy with the ground-truth, we use the Hausdorff Distance (HD) as a measure, as described in Huttenlocher et al. (1993). For the *Extraction*, this is calculated between $\hat{\mathbf{M}}$ and \mathbf{M} , while for the *Alignment* it is calculated between $\mathbf{M}^{\hat{\mathbf{p}}}$ and $\mathbf{M}^{\mathbf{p}}$.

Mean Squared Error

The Mean Squared Error (MSE) measures the average squared difference between the predicted alignment parameters $\hat{\mathbf{p}}$ and the ground-truth \mathbf{p} , as described in Appendix A

Symmetry Coefficient

As mentioned in Section 1, there is an asymmetry of structural information between the two brain hemispheres due to the interaction between the skull and the US beam. To assess if the *Extraction* network is predicting both sides of the brain symmetrically, we align the predicted brain mask $\hat{\mathbf{M}}$ with the ground-truth alignment parameters \mathbf{p} , obtaining $\hat{\mathbf{M}}^{\mathbf{p}}$. We then mirror the right hemisphere through the sagittal plane and calculate the DSC between both hemispheres, as shown in Eq. (4). We name this measure Symmetry Coefficient (SC), and a visual representation is shown in Fig. 4.

$$SC(\hat{\mathbf{M}}) = DSC(\hat{\mathbf{M}}_{left}^{\mathbf{p}}, \hat{\mathbf{M}}_{right(mirrored)}^{\mathbf{p}}) \quad (4)$$

Spatial Spread of Landmarks

To assess the impact of the *Alignment* network of BEAN on the spatial distribution of the brain structures of the original scan, a single point was placed on each face as well as on the geometric center of each aligned 3D US scan, as shown in Fig. 8a. Each color represents a different landmark (left, right, top, bottom, front, back, center). These landmarks were then transformed back to their respective unaligned position, resulting in the original Spatial Spread of Landmarks (SSL) of the unaligned dataset shown in Fig. 8b. Finally, the parameters predicted by using a single alignment subnetwork (Coarse) and the two alignment subnetworks (Coarse + Fine) were then used to align the landmarks (Fig. 8c and 8d, respectively).

We define two measures to analyze the SSL of each case: the Mean Shift of SSL and the Mean Spread of SSL. To obtain them, we first calculate the mean position of each landmark. If we define the position of every individual landmark by \mathbf{k}_j^i and their goal position by \mathbf{u}_j , where j is the landmark and i is the scan number, we can calculate the mean position of each landmark over all scans with Eq. (5), where m denotes the total number of scans.

$$\mathbf{k}_j = \frac{1}{m} \sum_{i=1}^m \mathbf{k}_j^i \quad (5)$$

The Mean Shift of SSL assesses the accuracy of the alignment. It is the mean Euclidean distance between the mean positions \mathbf{k} and the goal positions \mathbf{u} , as shown in Eq. (6) with $g = 7$ representing the total number of landmarks.

$$\text{Mean Shift of SSL}(\mathbf{k}, \mathbf{u}) = \frac{1}{g} \sum_{j=1}^g \|\mathbf{k}_j - \mathbf{u}_j\| \quad (6)$$

The Mean Spread of SSL assesses the precision of the alignment. It describes the spread of each landmark point cloud and is determined by the mean Euclidean distance between all points of a landmark and their mean position as shown in Eq. (7).

$$\text{Mean Spread of SSL}(\mathbf{k}) = \frac{1}{gm} \sum_{j=1}^g \sum_{i=1}^m \|\mathbf{k}_j^i - \mathbf{k}_j\| \quad (7)$$

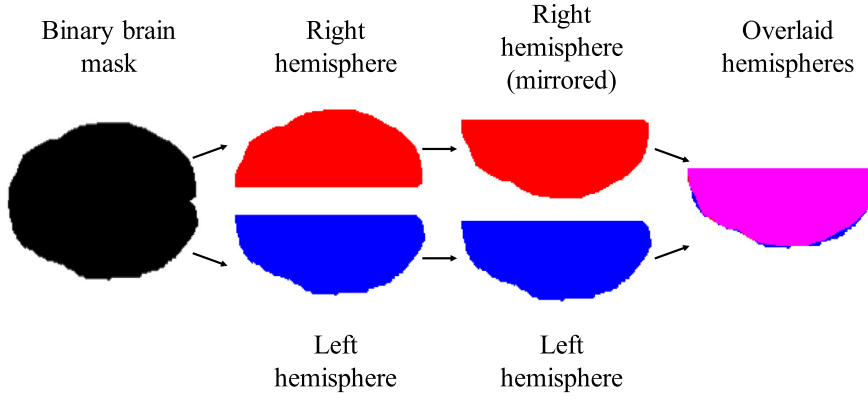


Fig. 4. Visual representation of how the Symmetry Coefficient is (SC) calculated. The SC is the Dice Similarity Coefficient (DSC) between the left hemisphere and right (mirrored) hemisphere, with the intersection between them shown in magenta.

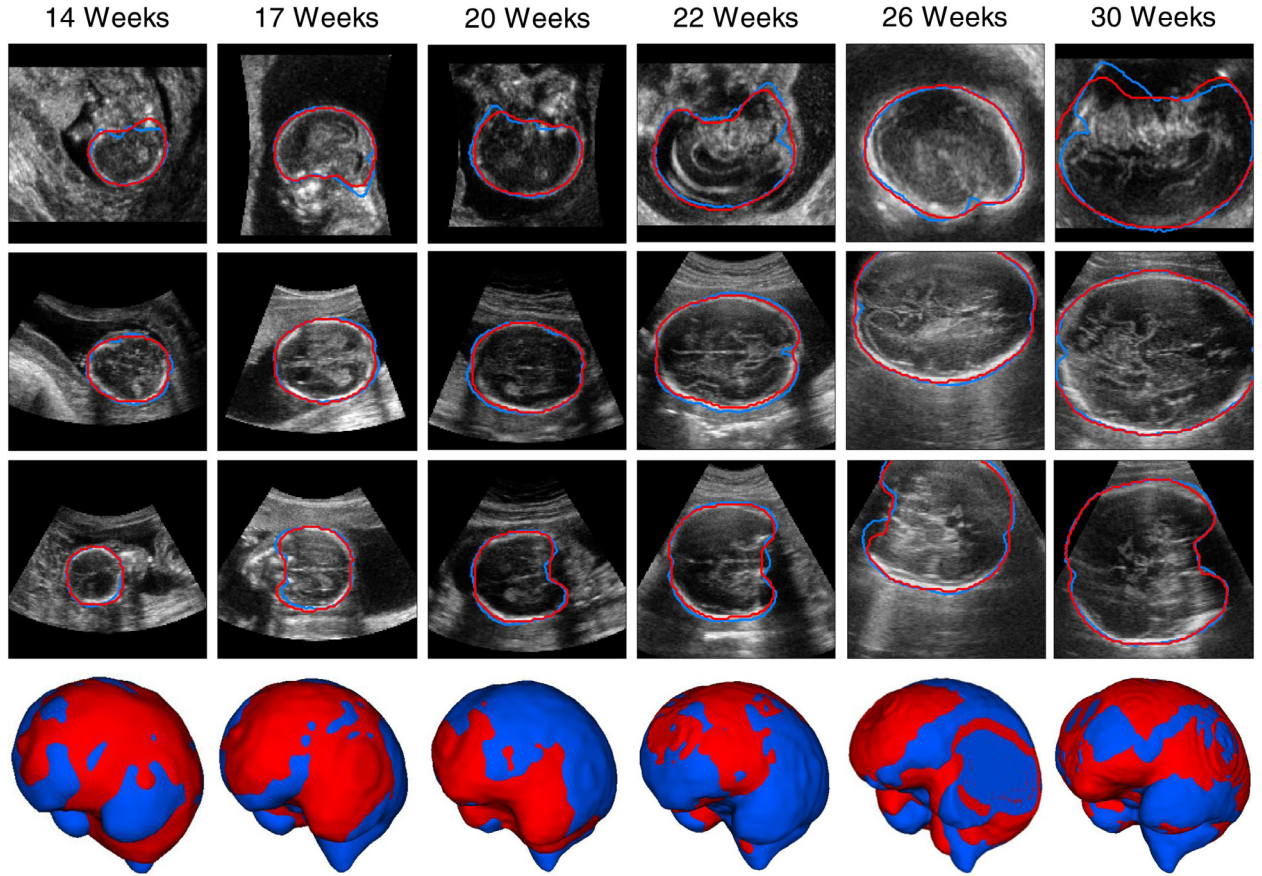


Fig. 5. Comparison of the *Extraction* network prediction (red) and the ground-truth (blue) superimposed onto the mid-planes of the 3D US volumes for different gestational ages. These volumes were selected to demonstrate the amount of variation between each scan. Top: XY-plane. Middle: XZ-plane. Bottom: YZ-plane. The GW of each volume is displayed above, and the bottom row shows an aligned and scaled 3D rendering of the predicted mask (red) overlaid on the ground-truth mask (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Network Consistency

Since the *Extraction* and *Alignment* networks of BEAN operate independently of each other, it is important to determine if the results of both are consistent. For this, we use the volume v of the predicted mask \mathbf{M} and the predicted scaling parameter p_s . As mentioned in Section 3.3, all scans have been scaled to match the average brain size at 30 GW, which has a volume denoted by v_{30} . Therefore, the scaling parameter p_s is proportional to the ratio between v_{30} and the volume v of extraction brain mask \mathbf{M} as shown in Eq. (8).

$$\frac{v_{30}}{v} = p_s^3 \quad (8)$$

We define the Network Consistency (NC) as shown in Eq. (9), with an ideal value of NC = 1.

$$\text{NC}(v, p_s) = \frac{v_{30}}{v \cdot p_s^3} \quad (9)$$

5. Results

5.1. Brain extraction

As stated in Section 4, we assess the performance of the *Extraction* network of BEAN by calculating the Center Distance (CD), Dice

Similarity Coefficient (DSC), and Hausdorff Distance (HD) between the predicted brain mask \hat{M} and the ground-truth M , as well as the Symmetry Coefficient (SC) of \hat{M} . As a baseline comparison, we use the results already presented in Moser et al. (2020) which comprise the approximation of the brain mask generated by the method proposed in Namburete et al. (2018), as well as the best-performing configuration of the preliminary results of our network. As shown in Table 1, BEAN outperforms the solution presented in Namburete et al. (2018) in every measure. The *Extraction* network of BEAN obtained a mean CD of 0.98 mm, a DSC of 0.94, and a HD of 7.87 mm. These results represent a 276%, 9.6%, and 92% improvement over the results obtained with the method used in Namburete et al. (2018). Additionally, the standard deviations of our results also show a much higher precision for BEAN. Finally, BEAN shows a 22% improvement of the SC with a result of 0.95

when compared to 0.74 obtained with Namburete et al. (2018). The addition of an upsampling step before the thresholding the brain mask results in a small improvement when compared to our previous work. The CD and HD results are 27.9% and 13% smaller, both statistically significant (see Section 4). In contrast, DSC and SC showed no statistically significant changes.

Examples of the outline of the predicted and ground-truth extractions can be seen for different views of a range of gestational ages in Fig. 5. The network performs accurately for the entire gestational range of our dataset, regardless of scale, orientation, and rotation. As the examples for gestational ages 26 and 30 show, the prediction works reliably even if parts of the brain are not contained in the 3D US beam or volume. When analyzing the performance difference between scans that were imaged from the left or right cerebral hemispheres, i.e., where the probe was

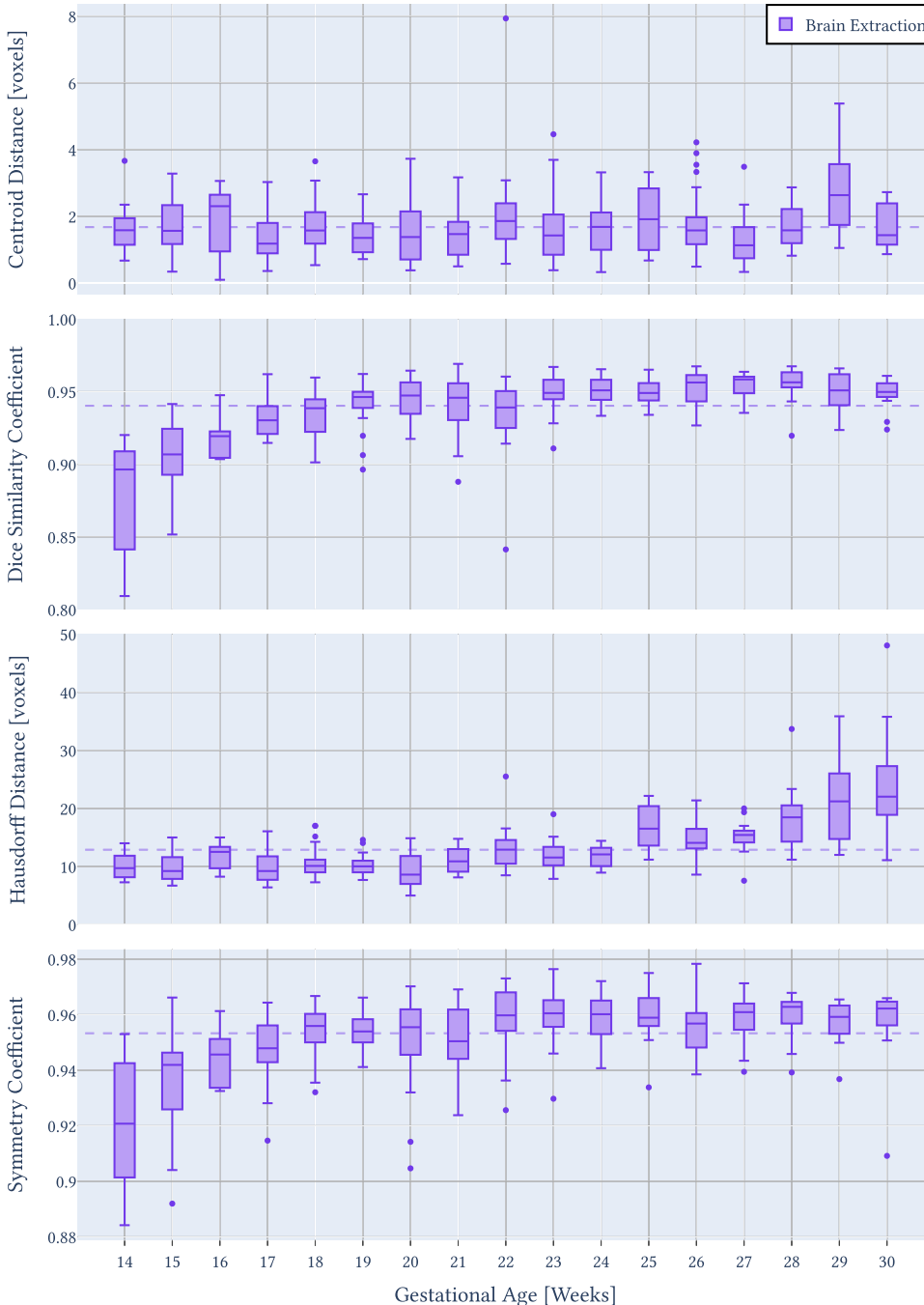


Fig. 6. Evaluation performance of the *Extraction* network of BEAN for each GW. The average result of each measure is shown as a dashed line.

Table 1

Testing results of the fully-trained *Extraction* network compared against method from Namburete et al. (2018) and our preliminary work from Moser et al. (2020). The best result for each measure has been highlighted in boldface font, with an arrow indicating whether a higher (up) or lower (down) value is desirable. CD: Centroid Distance of binary masks. HD: Hausdorff Distance. DSC: Dice Similarity Coefficient. SC: Symmetry Coefficient.

Method	CD [mm] ↓	DSC ↑	HD [mm] ↓	SC ↑
Namburete et al. (2018)	3.68 ± 4.44	0.85 ± 0.08	15.12 ± 5.24	0.74 ± 0.06
Moser et al. (2020)	1.36 ± 0.72	0.94 ± 0.02	9.05 ± 3.56	0.95 ± 0.02
BEAN (ours)	0.98 ± 0.53	0.94 ± 0.02	7.87 ± 3.12	0.95 ± 0.02

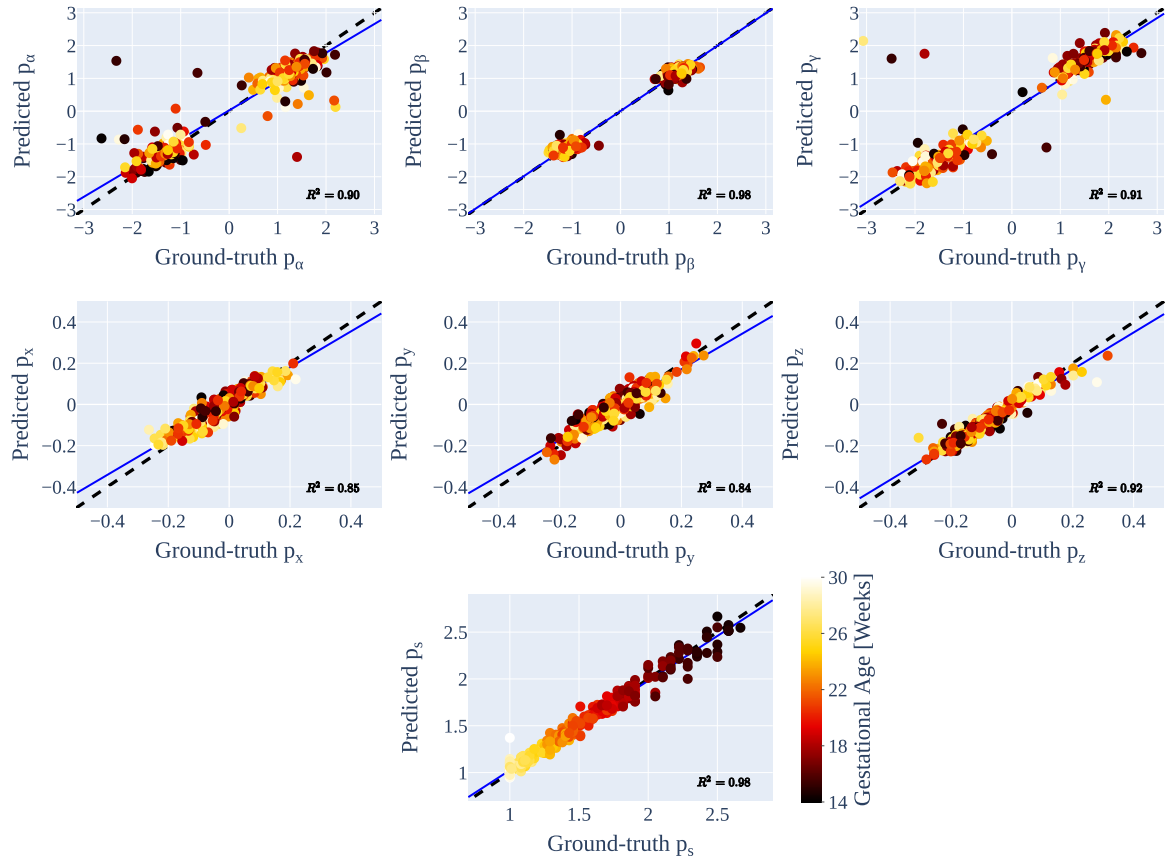


Fig. 7. The value of every parameter predicted by the *Alignment* network has been plotted against their target value, with the color of each value representing the GW of the 3D US scan. The linear regression for each parameter is shown in blue, with its R^2 score displayed in the bottom-right corner. The line of equality is shown as a dashed line on every plot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

located on the left or right side of the fetal brain, only the DSC result was statistically significant but the relative difference between them was less than 1%.

5.1.1. Brain extraction by gestational week

Fig. 6 shows the performance of each performance measure for the *Extraction* network for each GW. The performance of the CD shows a consistent trend throughout the gestational period of 14 to 30 weeks. The DSC, however, shows a drop in performance between weeks 14 and 19. This is likely due to the smaller brain volume of the earlier GWs, which increases the DSC sensitivity to mislabeled voxels since the overall true-positive voxels are fewer. We observe a similar trend for the SC, which is also likely to be due to the same reason. Regardless of this effect, we note that the lowest average DSC is still around 0.90. The performance of the HD, however, shows a decrease from 25 GW onward. As mentioned in Section 3.3, the ground-truth extraction masks are based on the atlases from Gholipour et al. (2017), which are created from MRI data. Unlike 3D US, this imaging modality has the ability to separate the cerebrospinal fluid from the brain tissue. As a result, the ground-truth masks are structurally different from what can be observed

with US around the edges of the brain. This is particularly evident for the gap between the occipital cortex and the cerebellum, which can be seen in Fig. 3 and was discussed in depth in Moser et al. (2020). Since these gaps widen as the fetal brain develops, so would the HD between the ground-truth and the predicted masks increase as well.

5.2. Brain alignment

The *Alignment* network of BEAN is evaluated by calculating the MSE between the predicted parameters $\hat{\mathbf{p}}$ and the ground-truth parameters \mathbf{p} , as well as the DSC and HD between the ground-truth mask \mathbf{M} aligned with the predicted alignment parameters $\hat{\mathbf{M}}^{\hat{\mathbf{p}}}$ and with the ground-truth parameters $\mathbf{M}^{\mathbf{p}}$. As a baseline comparison, we assess the performance of the *Alignment* network using a simple Mean Absolute Error (MAE) loss function \mathcal{L}_{MAE} during training, as well as using only one subnetwork (Coarse) instead of two (Coarse+Fine). The results in Table 2 show that the combined loss function $\mathcal{L}_{comb} = \mathcal{L}_{MAE} + \mathcal{L}_{DSC}$ outperforms a basic \mathcal{L}_{MAE} in every measure, with an improvement of 14.3% for MSE, 1.3% for DSC, and 10.6% for HD. However, this improvement is only statistically significant for the latter two measures, highlighting

Table 2

Testing results of the *Alignment* network. The results are shown for two types of loss functions used during training: Mean Absolute Error (\mathcal{L}_{MAE}) of the predicted parameters compared to the ground-truth, and the sum of MAE with the Dice Similarity Coefficient ($\mathcal{L}_{MAE} + \mathcal{L}_{DSC}$) loss of 3D ground-truth brain-masks transformed with the predicted and ground-truth parameters. The results of using a single alignment subnetwork (Coarse) are shown to display the improvement from our cascading two-subnetwork approach (Coarse + Fine). The performance tests are Mean Squared Error (MSE) of the predicted parameters, as well as Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) of the transformed masks. The values for HD are shown in voxels of the aligned volume. The best result for each measure has been highlighted in boldface font, with an arrow indicating whether a higher (up) or lower (down) value is desirable.

Loss Fn.	Align Step	MSE ↓	DSC ↑	HD ↓
\mathcal{L}_{MAE}	Coarse	0.063±0.340	0.858±0.065	18.68±7.38
	Coarse+Fine	0.014±0.039	0.913±0.062	13.34±7.25
$\mathcal{L}_{MAE} + \mathcal{L}_{DSC}$	Coarse	0.065±0.348	0.893±0.049	15.77±6.58
	Coarse+Fine	0.012±0.042	0.925±0.043	11.92±6.23

the fact that simply assessing the predicted parameters is not enough for a reliable alignment performance evaluation. We note here that using \mathcal{L}_{DSC} as a loss function by itself was not enough supervision for training and therefore never managed to converge. [Table 2](#) also shows that alignment with two cascaded subnetworks (Coarse + Fine) outperforms a single subnetwork (Coarse) in every measure, with an improvement of 81.5% for MSE, 3.6% for DSC, and 24.4% for HD, all of which are statistically significant. We also notice an improvement of 87.9% for the standard deviation of MSE, 12.2% for the standard deviation of DSC, and 5.3% for the standard deviation of HD, showing that the

Table 3

Mean squared error results of the *Alignment* network separated by parameter type: rotation parameters p_α , p_β , and p_γ ($MSE_{\alpha\beta\gamma}$), translation parameters p_x , p_y , and p_z (MSE_{xyz}), and scaling parameter p_s (MSE_s). The best result has been highlighted in boldface font.

Loss Fn.	Align Step	$MSE_{\alpha\beta\gamma}$ [10^{-3}]	MSE_{xyz} [10^{-3}]	MSE_s [10^{-3}]
\mathcal{L}_{MAE}	Coarse	139.7±792.3	1.8±2.1	1.7±3.1
	Coarse+Fine	30.3±9.1	1.1±1.4	1.7±4.9
$\mathcal{L}_{MAE} + \mathcal{L}_{DSC}$	Coarse	148.8±812.5	1.2±1.4	4.8±11.8
	Coarse+Fine	28.8±8.9	0.9±1.2	1.6±4.5

Coarse + Fine alignment is not only more accurate but also more consistent. We also remark that a third subnetwork was considered but early results showed an inability to generalise and was therefore not further developed.

[Table 3](#) shows the MSE results shown in [Table 2](#) separated for rotation parameters ($MSE_{\alpha\beta\gamma}$), translation parameters (MSE_{xyz}), and scaling parameters (MSE_s). These results show that while the inclusion of a Fine alignment step improves the performance for all parameters, the largest effect is on the rotation parameters, for which $MSE_{\alpha\beta\gamma}$ shows a decrease of 78% and 80% for \mathcal{L}_{MAE} and $\mathcal{L}_{MAE} + \mathcal{L}_{DSC}$, respectively. In contrast, the added supervision of \mathcal{L}_{DSC} has a smaller but more equitable impact on the parameters. For Coarse, the performance improvements shown in [Table 2](#) are due to a 33% lower MSE_{xyz} which has a bigger impact on the overall performance at this stage than the 6.5% and 182% increase of $MSE_{\alpha\beta\gamma}$ and MSE_s , respectively. However, \mathcal{L}_{DSC} improves performance for all parameters when using a Coarse + Fine alignment.

The performance for each individual parameter can be seen in [Fig. 7](#), where they have been plotted against their target value, along with their linear regression and the corresponding R^2 score. The Euler angles p_α ,

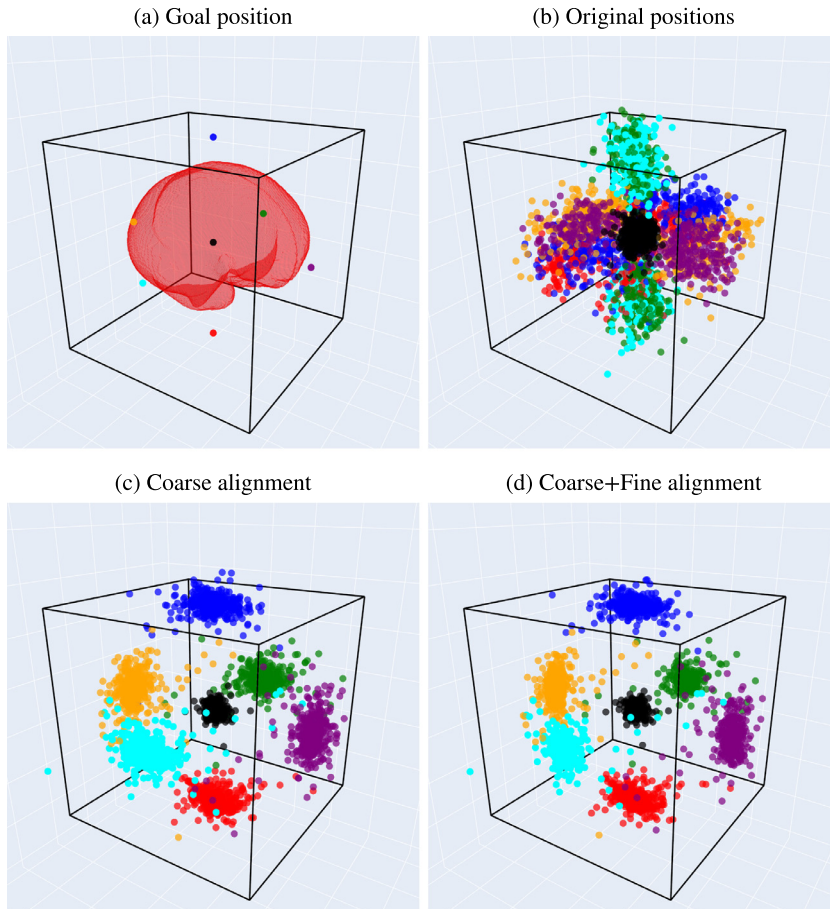


Fig. 8. Visualization of Spatial Spread of Landmarks (SSL) as described in [Section 4](#). A single point has been located at the middle of each face of each 3D US scan: Top (blue), bottom (red), front (purple), back (orange), left (green), right (cyan), and center (black). These points are the goal positions for the aligned volumes, shown in (a) along with an outline of an ideally aligned brain of 30 GW for reference. Using the ground-truth alignment parameters, they have been transformed to their original positions in their unaligned scans (b). The possible orientations of the head when acquiring the scan causes the clusters to be divided among opposing faces of the volume. The results of a single alignment subnetwork (Coarse) and of our cascading two-subnetwork (Coarse + Fine) architecture are shown in (c) and (d) respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

p_β , and p_γ lie consistently on the line of equality with only a handful of outliers and no visible correlation with gestational age. We observe that the angles are distributed in two distinct clusters as a result of the clinical acquisition protocol, which does not take into account the top-bottom and left-right orientation of the brain, as shown in the XY-planes of Fig. 5. The predictions of p_β also show better performance than p_α and p_γ , which seem to be more sensitive to the quality of the scan, since the outliers had large acoustic shadows or a higher amount of motion blur. The predicted Cartesian shifts p_x , p_y , and p_z are highly consistent, lying on top of the line of equality. These 6 parameters show only a weak negative correlation with the gestational age of the fetus, with their Pearson's correlation coefficients lying between -0.04 and -0.26. Finally, the predicted scaling p_s shows a high degree of consistency, with a single outlier being a 30.4 GW scan with particularly strong acoustic shadows and some sections of the brain located outside of the 3D US beam. The expected correlation with gestational age is also clearly visible (Pearson's correlation coefficient of -0.94), with a slight drop in performance for earlier GWs where the required scaling is largest.

As described in Section 4, Fig. 8 shows the SSL, with Fig. 8(a) showing the goal position of our 7 landmarks (top, bottom, left, right, front, back, and center) as well as an outline of an ideally aligned brain of

30 GW for reference. Fig. 8(b) shows the positions of these landmarks in their original, unaligned scans. Here, the landmark positions are dispersed widely. The spread of the central landmark (black) shows a large variation in the location of the brain with respect to the scan. For the remaining landmarks, there is a large radial spread from the central landmark which represents the scaling variation between volumes, as well as a large axial spread which represents the rotational variation. Fig. 8(b) also shows that each landmark cluster is split along two sides of the volume. In comparison, Fig. 8(c) and 8(d) show the landmark positions after Coarse and Coarse + Fine alignments, respectively. In both cases, the clusters have been dramatically reduced in size. The cluster of the central landmark is now considerably smaller, which indicates that the center of the brains are closer to the center of the volume. The predicted scaling has greatly reduced the radial spread, while the rotation parameters have reduced the axial spread. We observe that the cluster split seen in Fig. 8(a) is no longer present. In comparison with Coarse alignment, Fine alignment manages to reduce the spread of the clusters even further, but this improvement is not as remarkable as that from the original positions to the aligned positions.

For a quantitative assessment of these results, the Mean Shift and Mean Spread of SSL as described in Section 4 of the original positions,

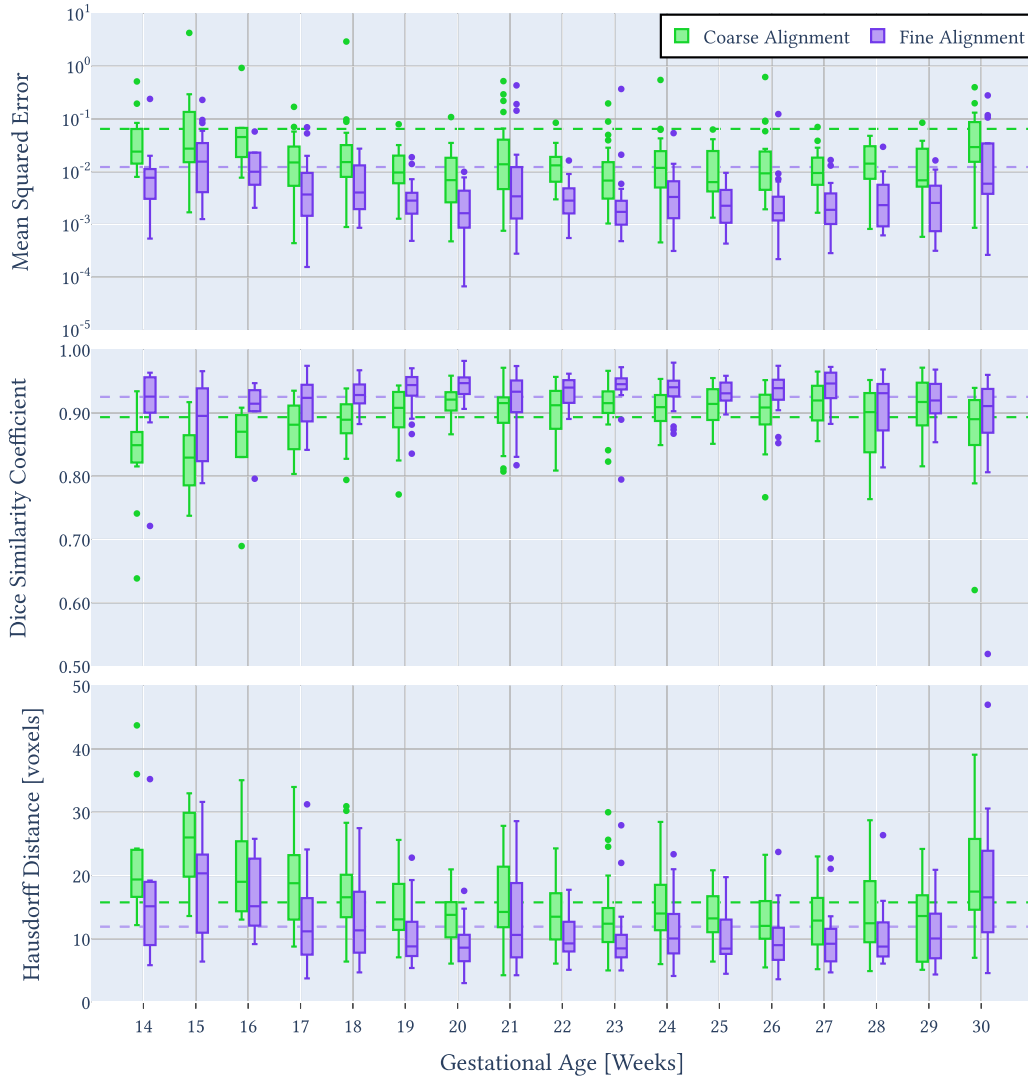


Fig. 9. Performance of the Alignment network of BEAN for each GW. The testing results of Mean Squared Error (MSE) of the predicted parameters, as well as the Dice Similarity Coefficient and Hausdorff Distance of the transformed binary brain masks are plotted for both the Coarse and Fine alignments. The average of each alignment step is shown as a dashed line of the corresponding color. Note that the y-axis scale for MSE Parameters is logarithmic, in order to show all outliers while still showing the improvements between coarse and fine alignments.

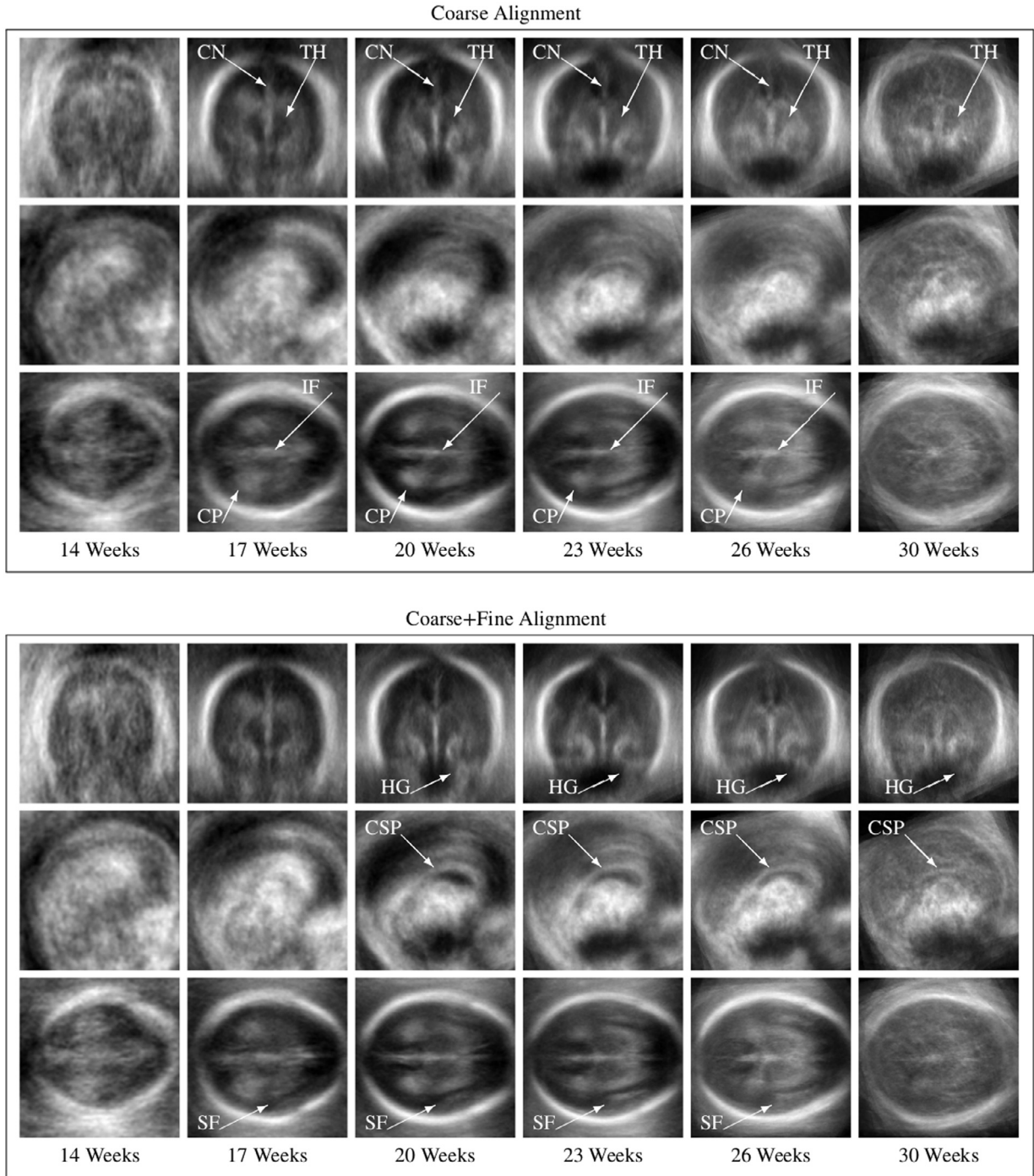


Fig. 10. Average volumes after alignment using BEAN, separated by GW. Top-, middle-, and bottom-row represent the XY, XZ, and YZ central orthogonal planes respectively. Top: One alignment subnetwork (Coarse). Bottom: Two alignment subnetworks (Coarse + Fine). Relevant brain structures have been marked to highlight the alignment performance. CN: Caudate Nucleus, TH: Thalamus, IF: Interhemispheric Fissure, CP: Choroid Plexus, HG: Hippocampal Gyrus, CSP: Cavum Septum Pellucidum, SF: Sylvian Fissure.

as well as the positions after Coarse and Fine alignment are shown in Table 4. We observe a reduction of the Mean Shift of 96.9% after Coarse alignment, and 97.19% after Fine alignment. The Mean Spread shows a 69.1% reduction after Coarse alignment, and a 73.9% reduction after Fine alignment.

5.2.1. Brain alignment by gestational week

To provide a deeper understanding of the performance of the *Alignment* network, we assess the predictions for each GW of our dataset,

Table 4

Evaluation results of the Spatial Spread of Landmarks (SSL) shown in Fig. 8>. The best result for each measure has been highlighted in boldface font. The equations describing the Mean Shift of SSL and the Mean Spread of SSL are found in Section 4.

Data	Mean Shift of SSL [voxel]	Mean Spread of SSL [voxel]
Original	65.15	50.42
Coarse	2.00	15.60
Fine	1.88	13.17

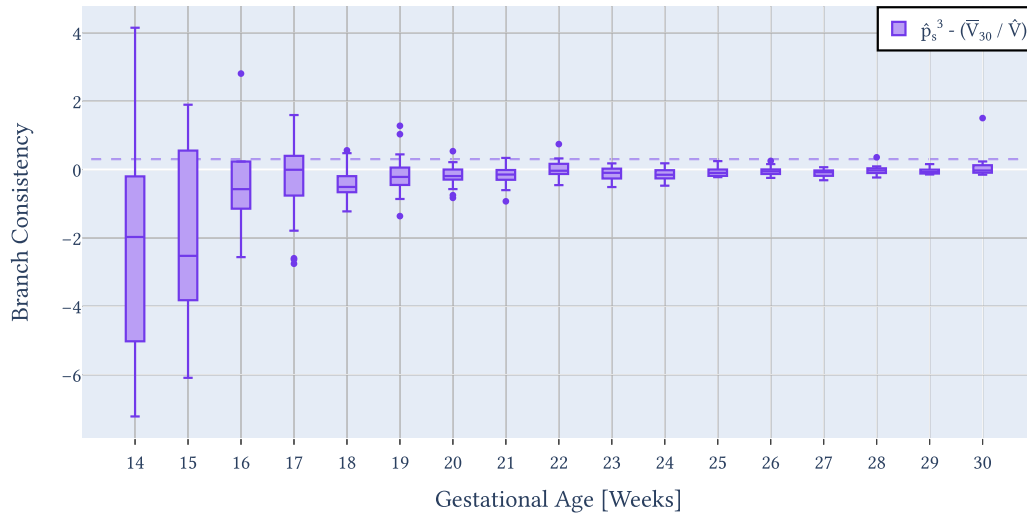


Fig. 11. Evaluation of the Network Consistency (NC) measure as described in Section 4 for each GW. It represents the difference between the predicted scaling parameter p_s of the *Alignment* network, and the volume of the brain mask predicted by the *Extraction* network \hat{M} in relation to the average brain volume of the brain at 30 GW \tilde{V}_{30} . A result of zero represents ideal consistency between both networks. The mean results are shown as a dashed line.

which ranges from 14.4 to 30.9 weeks. Fig. 9 shows a breakdown of each performance measure for the Coarse and Fine alignments.

The performance improvement as well as the consistency throughout the range of gestational ages when comparing both alignment steps is evident. The Fine alignment increases performance for every gestational age, especially around the extrema of the gestational age range. All three measures show a decline in performance for 14–16 GW as well as 30 GW, which is expected due to their smaller sampling sizes as shown in Fig. 2. However, there are certain characteristics of the data that are most likely related to this performance drop as well. In the case of the early GWs (14–16 GW), this is most likely a result of the larger transformation required, since all scans are aligned and scaled to the 30 GW scans. While scaling to the largest brains minimizes the information loss caused by downsampling, it results in errors in the predicted angles and shifts of smaller brains being amplified by the relatively larger scaling factor. We note that although the aforementioned performance drops are noticeable, the median performance is still high for every week, with GW 15, for example, obtaining a median MSE of 0.027 and a median DSC of 0.895.

As a final evaluation of the alignment performance, Fig. 10 shows the average volumes for several GWs. These have been split into Coarse and Fine alignment to qualitatively illustrate the differences between them.

5.3. Full network results

In Section 5.1 and 5.2, we have shown the results of our full assessment of the performance of the brain alignment and brain extraction of BEAN. Since the *Extraction* and *Alignment* networks are fully independent, this represents the complete performance assessment of BEAN. However, in order to test if both networks are consistent with one another, we must analyze the Network Consistency (NC) score introduced in Section 4. On average, BEAN obtained a NC score of 0.31 ± 1.10 , indicating a very high degree of consistency between the *Extraction* and *Alignment* networks. This performance drops for earlier GWs, which is most likely due to the relative impact that each voxel has on the total predicted brain mask, as well as the larger scaling parameters. As shown in Fig. 11, the performance of 14 GW and 15 GW shows the lowest NC, which is consistent with the results seen in Section 5.1 and 5.2.

For a final demonstration of the performance of the network, we have chosen an example gestational age of 22 weeks, and compared the mean results of the full network with the mean results obtained using

the ground-truth mask and alignment parameters, which is shown in Fig. 12. The results of BEAN are similar to the expected ground-truth, both in terms of alignment and extraction, highlighting its exceptional performance. The only major difference is the discrepancy in the space between the occipital cortex and the cerebellum that can be clearly seen in the sagittal plane view, as mentioned in Section 5.1.

6. Discussion

In this paper we present BEAN, a multi-stage convolutional neural network, that automatically extracts the fetal brain from minimally pre-processed 3D US scans and aligns it to a common space. BEAN extracts the fetal brain without the need for shape approximations, allowing for the actual shape of the brain to be analysed, which is fundamental to the analysis of brain maturation. It also predicts the brain alignment parameters using the entire scan for context. As a result, it does not require specific structures such as the eyes to be clearly visible, something that is not always possible due to the unpredictable positional variability of the fetal head in the womb during gestation. The modularity of BEAN also allows for the predictions to be used as needed. The independent nature of the *Extraction* and *Alignment* subnetworks allows for brain extraction to be performed without the need for alignment and vice-versa. Similarly, the parametric prediction of the *Alignment* network makes it possible to perform partial alignments, e.g. the brains can be aligned with and without scaling. As a result, BEAN is a fast, modular, age-independent, and consistent solution for the extraction and alignment of the fetal brain from 3D US with state-of-the-art performance.

The methodology presented in this work aims to help automate the tasks of brain alignment and extraction from 3D neurosonograms in order to facilitate the development of neuroimage analysis pipelines for intrauterine brain development (Namburete et al., 2015; Yaqub et al., 2015). The lack of robust automated solutions for these tasks, in part due to the relatively lower contrast and higher imaging artifacts when compared to other modalities such as MRI, has limited the usage of 3D US both for research of brain maturation. It has also limited the potential of 3D US for clinical use, since clinical evaluation 3D volumes without automated assistance tools is considerably harder than that of 2D planes. The recent recommendation by ISUOG to use 3D scans when possible (Paladini et al., 2021) has made the need for such automated brain extraction and alignment tools even more urgent.

To the best of our knowledge, BEAN is the first network that extracts the 3D fetal brain directly from a 3D scan without the need to analyse

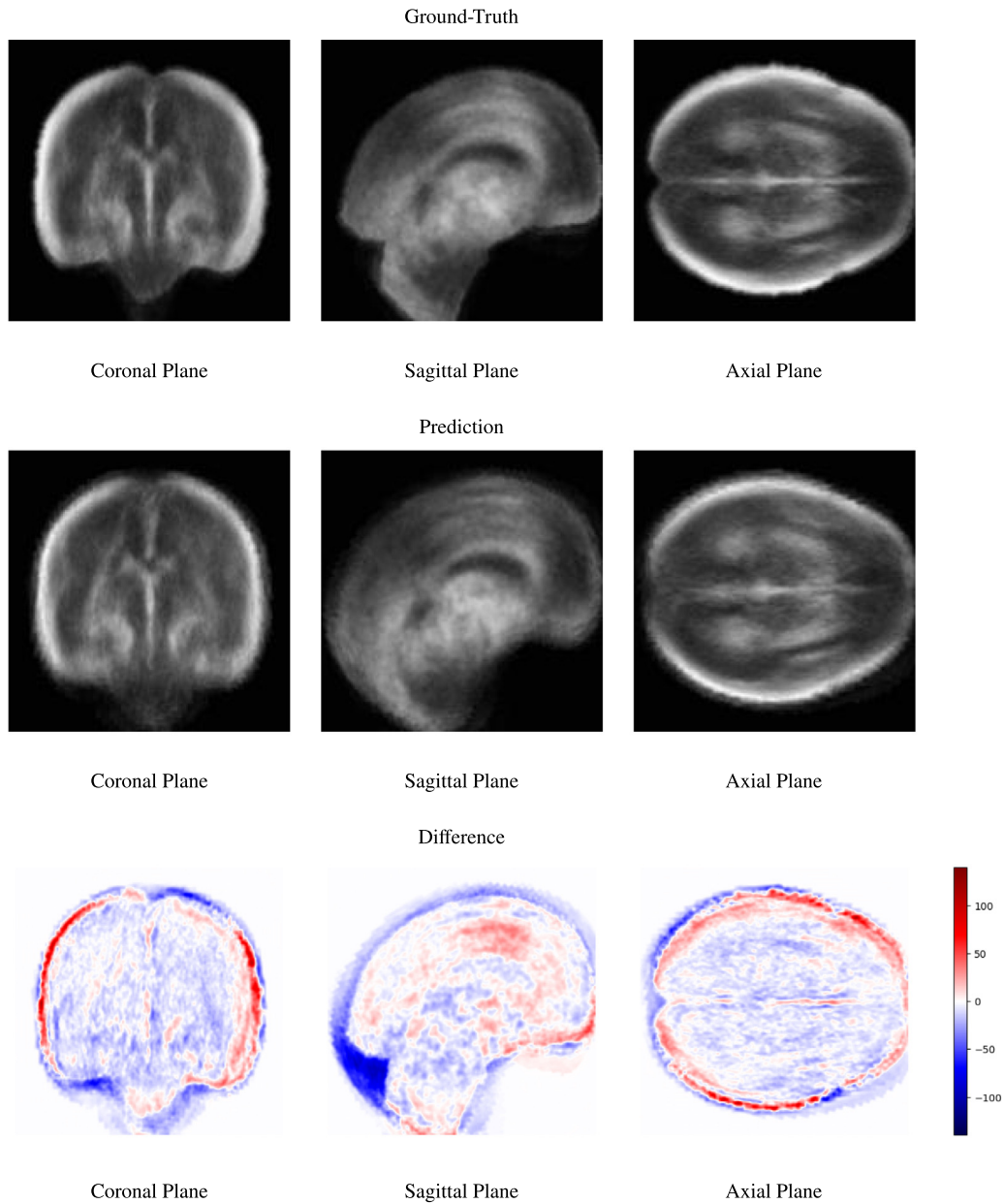


Fig. 12. Comparison of mean predicted output of BEAN, i.e., *Extraction* and *Alignment*, and the ground-truth for gestational week 22. Top: Ground truth. Middle: Prediction. Bottom: Difference between ground truth and prediction.

each 2D frame independently. As a result, the BEAN has access to much richer contextual information, which results in better performance and consistency, as shown in Section 5. The extracted brain is a direct segmentation from the original 3D US scan and does not require any approximations of the brain shape. Additionally, while the performance improvements gained in comparison to Moser et al. (2020) by adding the upsampling step are small, it allows for state-of-the-art performance without requiring any changes to the network architecture or size, and with a negligible additional computational cost. The extraction performance achieved by BEAN rivals to that of state-of-the-art solutions for fetal MRI mentioned in Section 2 such as Ebner et al. (2020), which achieved a similar DSC score of 0.94.

BEAN was trained on 829 and tested on 356 3D US scans spanning the gestational age range of 14 to 30 weeks, representing healthy fetuses from several ethnic and geographical groups. During this gestational period, the fetal head undergoes rapid structural changes, including ossification of the skull and development of brain structures, which result in a large intrinsic variability of the structural information of the

3D US scans. Nevertheless, without the need of any additional inputs or pre-processing (besides isotropic sampling and volume centre-cropping) our results show that BEAN performs consistently across this gestational range and structural variability with only some performance loss around the edges of the gestational range, which is expected due to the broad age distribution of our dataset (Fig. 2).

Although BEAN was trained on data from a multi-site study, every 3D US scan collected using the same model of US machine and transducer. As such, a possible limitation is that the performance of BEAN could vary for scans obtained with different equipment. While multi-site datasets of other modalities are being harmonised to remove scanner biases (Dinsdale et al., 2021), in 3D US the operator settings (e.g. time-gain compensation) have a much larger impact on the characteristics of the image. Therefore, the variability introduced by the settings used in different sites should counterbalance this limitation. Additionally, the data used to train BEAN was from the INTERGROWTH-21st (Papageorgiou et al., 2014) study, which was specifically designed to image healthy women that experienced a normal pregnancy. As a

result, the data used to train and test BEAN excludes scans exhibiting particularly abnormal brain development. This could potentially hamper BEAN's applicability in such cases and needs to be explored. The independent nature of the *Extraction* and *Alignment* subnetworks might also prove limiting in the performance of BEAN. Although this architecture has many benefits as already discussed, training each task independently might be limiting the overall performance. Multi-task learning has shown to improve the performance and flexibility of Deep Learning methods (Namburete et al., 2018; Ruder, 2017), therefore a performance comparison of such an approach against the current architecture of BEAN is still needed.

As mentioned in Section 5, BEAN shows a decreased performance for scans with particularly poor quality, i.e., large amount of acoustic shadows or motion blur.

In summary, BEAN presents a fast, consistent, and robust DL solution for the automated extraction and alignment of the fetal brain from 3D US scans to a common coordinate space. As a result, BEAN represents a helpful tool to aid the development of neuroimage analysis pipelines for the development of the fetal brain during gestation. The performance of the network should only improve with larger, more diverse datasets, and it will be interesting to see how it performs with abnormally developed brains.

Data and code availability statement

The 3D US scans used for this work are available from the INTERGROWTH-21st Consortium upon reasonable request. The spatiotemporal atlas used to generate the 3D brain masks was developed by Gholipour et al. (2017) and is available at http://crl.med.harvard.edu/research/fetal_brain_atlas/. All network training code as well as the weights of the trained networks can be requested by emailing the corresponding author.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Felipe Moser: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ruobing Huang:** Conceptualization, Software, Data curation. **Bartłomiej W. Papież:** Supervision, Conceptualization, Writing – review & editing. **Ana I.L. Namburete:** Supervision, Conceptualization, Data curation, Writing – review & editing, Project administration, Funding acquisition.

Acknowledgement

Felipe Moser acknowledges the support and funding from the Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC) (EP/L016052/1), as well as the support from University College Oxford and its Oxford-Radcliffe benefaction. Ana Namburete is grateful for the support from the UK Royal Academy of Engineering under the Engineering for Development Research Fellowships scheme and Pembroke College. Bartłomiej W. Papież acknowledges Rutherford Fund at Health Data Research UK (HDR UK) (MR/S004092/1) and the Nuffield Department of Population Health (NDPH) Senior Research Fellowship.

Appendix A. Measures

Consider \mathbf{M} and $\hat{\mathbf{M}}$ to be the binary ground-truth and predicted masks, respectively. We define m as the total number of voxels in the

volume, and \mathbf{M}_i and $\hat{\mathbf{M}}_i$ the binary values of the voxels $\mathbf{M}(x_i, y_i, z_i)$ and $\hat{\mathbf{M}}(x_i, y_i, z_i)$, respectively. The ground-truth alignment parameters are defined as $\mathbf{p} = [p_\alpha, p_\beta, p_\gamma, p_x, p_y, p_z, p_s]$, analogous for the predicted parameters $\hat{\mathbf{p}}$.

Centroid distance

The centroid $(\bar{x}_M, \bar{y}_M, \bar{z}_M)$ of \mathbf{M} (analogous for $\hat{\mathbf{M}}$) can be calculated as shown in Eq. (A.1).

$$(\bar{x}_M, \bar{y}_M, \bar{z}_M) = \frac{1}{\sum_{i=1}^m \mathbf{M}_i} \left(\sum_{i=1}^m (\mathbf{M}_i \cdot x_i), \sum_{i=1}^m (\mathbf{M}_i \cdot y_i), \sum_{i=1}^m (\mathbf{M}_i \cdot z_i) \right) \quad (\text{A.1})$$

This follows analogous for all our binary brain masks. Therefore, the Centroid Distance between \mathbf{M} and $\hat{\mathbf{M}}$ is calculated as shown in Eq. (A.2)

$$\text{CD}(\mathbf{M}, \hat{\mathbf{M}}) = \sqrt{(\bar{x}_M - \bar{x}_{\hat{M}})^2 + (\bar{y}_M - \bar{y}_{\hat{M}})^2 + (\bar{z}_M - \bar{z}_{\hat{M}})^2} \quad (\text{A.2})$$

Dice Similarity Coefficient

The Dice Similarity Coefficient between \mathbf{M} and $\hat{\mathbf{M}}$ is calculated as shown in Eq. (A.3).

$$\text{DSC}(\mathbf{M}, \hat{\mathbf{M}}) = \frac{2 \cdot \sum_{i=1}^m (\mathbf{M}_i \cdot \hat{\mathbf{M}}_i) + s}{\sum_{i=1}^m \mathbf{M}_i + \sum_{i=1}^m \hat{\mathbf{M}}_i + s} \quad (\text{A.3})$$

For this work, we used a smoothing factor $s = 1$.

Mean Squared Error

The Mean Squared Error between the ground-truth and predicted alignment parameters, \mathbf{p} and $\hat{\mathbf{p}}$, respectively, is calculated as shown in Eq. (A.4), with $w = 7$.

$$\text{MSE}(\mathbf{p}, \hat{\mathbf{p}}) = \frac{1}{w} \sum_i^w (p_i - \hat{p}_i)^2 \quad (\text{A.4})$$

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/>.
- Anquez, J., Angelini, E.D., Bloch, I., 2009. Automatic segmentation of head structures on fetal MRI. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, pp. 109–112.
- Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D., 2017. Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans Med Imaging 36 (11), 2204–2215. doi:10.1109/TMI.2017.2712367. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6051487/>.
- Becker, B.G., Cosio, F.A., Huerta, M.E.G., Benavides-Serralde, J.A., 2010. Automatic segmentation of the cerebellum of fetuses on 3D ultrasound images, using a 3d point distribution model. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, pp. 4731–4734. <https://ieeexplore.ieee.org/abstract/document/5626624>.
- Becker, B.G., Cosio, F.A., Huerta, M.E.G., Benavides-Serralde, J.A., 2010. Automatic segmentation of the cerebellum of fetuses on 3d ultrasound images, using a 3d point distribution model. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, pp. 4731–4734.
- Carneiro, G., Georgescu, B., Good, S., Comaniciu, D., 2008. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans Med Imaging 27 (9), 1342–1355.
- Cerrolaza, J.J., Oktay, O., Gomez, A., Matthew, J., Knight, C., Kainz, B., Rueckert, D., 2017. Fetal skull segmentation in 3D ultrasound via structured geodesic random forest. In: Fetal, Infant and Ophthalmic Medical Image Analysis. Springer, pp. 25–32. https://link.springer.com/chapter/10.1007/978-3-319-67561-9_3.

- Cerrolaza, J.J., Sinclair, M., Li, Y., Gomez, A., Ferrante, E., Matthew, J., Gupta, C., Knight, C.L., Rueckert, D., 2018. Deep learning with ultrasound physics for fetal skull segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 564–567.
- Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P.A., 2015. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Health Inform* 19 (5), 1627–1636. doi:10.1109/JBHI.2015.2425041. <http://www.ncbi.nlm.nih.gov/pubmed/25910262> <http://ieeexplore.ieee.org/document/7090943/>.
- Chen, H.-C., Tsai, P.-Y., Huang, H.-H., Shih, H.-H., Wang, Y.-Y., Chang, C.-H., Sun, Y.-N., 2012. Registration-based segmentation of three-dimensional ultrasound images for quantitative measurement of fetal craniofacial structure. *Ultrasound in Medicine & Biology* 38 (5), 811–823. doi:10.1016/j.ultrasmedbio.2012.01.025. <http://www.sciencedirect.com/science/article/pii/S0301562912000579>.
- Chollet, F., et al., 2015. Keras. <https://keras.io>.
- Cointepas, Y., Mangin, J.-F., Garner, L., Poline, J.-B., Benali, H., 2001. Brainvisa: software platform for visualization and analysis of multi-modality brain data. *Neuroimage* 13 (6, Supplement), 98. doi:10.1016/S1053-8119(01)91441-7. Originally published as Volume 13, Number 6, Part 2.
- Cuingnet, R., Somphone, O., Mory, B., Prevost, R., Yaqub, M., Napolitano, R., Papageorgiou, A., Roundhill, D., Noble, J., Ardon, R., 2013. Where is my baby? a fast fetal head auto-alignment in 3d-ultrasound. *Proc IEEE Int Symp Biomed Imaging* 768–771. doi:10.1109/ISBI.2013.6556588.
- d'Agostino, R.B., 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58 (2), 341–348.
- Dinsdale, N.K., Jenkinson, M., Namburete, A.I.L., 2021. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *NeuroImage* 228, 117689.
- Ebner, M., Wang, G., Li, W., Aertsen, M., Patel, P.A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., David, A.L., Deprest, J., Ourselin, S., Vercateren, T., 2020. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri. *Neuroimage* 206, 116324. doi:10.1016/j.neuroimage.2019.116324. <http://www.sciencedirect.com/science/article/pii/S1053811919039152>.
- Eskildsen, S.F., Coup, P., Fonov, V., Manjiv, J.V., Leung, K.K., Guizard, N., Wassef, S.N., stergaard, L.R., Collins, D.L., 2012. Beast: brain extraction based on nonlocal segmentation technique. *Neuroimage* 59 (3), 2362–2373. doi:10.1016/j.neuroimage.2011.09.012. <http://www.sciencedirect.com/science/article/pii/S1053811911010573>.
- Gholipour, A., Asl, A.A., Estroff, J., Warfield, S., 2012. Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly. *Neuroimage* 60, 1819–1831.
- Gholipour, A., Rollins, C., Velasco-Annis, C., Ouassal, A., Akhondi-Asl, A., Afacan, O., Ortinau, C., Clancy, S., Limperopoulos, C., Yang, E., Estroff, J., Warfield, S., 2017. A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Sci Rep* 7. doi:10.1038/s41598-017-00525-w.
- Guha Roy, A., Conjeti, S., Navab, N., Wachinger, C., 2018. Quicknat: a fully convolutional network for quick and accurate segmentation of neuroanatomy. *Neuroimage* 186. doi:10.1016/j.neuroimage.2018.11.042.
- Gutiérrez-Becker, B., Arámbula Cosío, F., Guzmán Huerta, M.E., Benavides-Serralle, J.A., Camargo-Marín, L., Medina Bañuelos, V., 2013. Automatic segmentation of the fetal cerebellum on ultrasound volumes, using a 3D statistical shape model. *Med. Biol. Eng. Comput.* 51 (9), 1021–1030. doi:10.1007/s11517-013-1082-1.
- Gutiérrez-Becker, B., Arámbula Cosío, F., Guzmán Huerta, M.E., Benavides-Serralle, J.A., Camargo-Marín, L., Medina Bañuelos, V., 2013. Automatic segmentation of the fetal cerebellum on ultrasound volumes, using a 3D statistical shape model. *Med. Biol. Eng. Comput.* 51 (9), 1021–1030. doi:10.1007/s11517-013-1082-1.
- Habas, P.A., Kim, K., Corbett-Deitz, J.M., Rousseau, F., Glenn, O.A., Barkovich, A.J., Studholme, C., 2010. A spatiotemporal atlas of mr intensity, tissue probability and shape of the fetal brain with application to segmentation. *Neuroimage* 53 (2), 460–470. doi:10.1016/j.neuroimage.2010.06.054. <http://www.sciencedirect.com/science/article/pii/S1053811910009122>.
- Han, X., Kwitt, R., Aylward, S., Bakas, S., Menze, B., Asturias, A., Vespa, P., Van Horn, J., Niethammer, M., 2018. Brain extraction from normal and pathological images: a joint pca/image-reconstruction approach. *Neuroimage* 176, 431–445. doi:10.1016/j.neuroimage.2018.04.073. <http://www.sciencedirect.com/science/article/pii/S1053811918303938>.
- Haratz, K.K., Lerman-Sagie, T., 2018. Prenatal diagnosis of brainstem anomalies. *European Journal of Paediatric Neurology* 22 (6), 1016–1026. doi:10.1016/j.ejpn.2018.06.011. <https://linkinghub.elsevier.com/retrieve/pii/S1090379818302290>.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. *Nature* 585 (7825), 357–362. doi:10.1038/s41586-020-2649-2.
- Hesse, L. S., Aliasi, M., Moser, F., Haak, M.C., Xie, W., Jenkinson, M., Namburete, A.I.L., 2022. Subcortical segmentation of the fetal brain in 3D ultrasound using deep learning. *NeuroImage* <https://www.sciencedirect.com/science/article/pii/S1053811922002452>.
- Hou, B., Khanal, B., Alansary, A., McDonagh, S., Davidson, A., Rutherford, M., Hajnal, J.V., Rueckert, D., Glocker, B., Kainz, B., 2018. 3-D reconstruction in canonical co-ordinate space from arbitrarily oriented 2D images. *IEEE Trans Med Imaging* 37 (8), 1737–1750. doi:10.1109/TMI.2018.2798801.
- Huang, R., Namburete, A.I.L., Yaqub, M., Noble, J.A., 2015. Automated mid-sagittal plane selection for corpus callosum visualization in 3D ultrasound images. Undefined. <https://www.semanticscholar.org/paper/Automated-Mid-sagittal-Plane-Selection-for-Corpus-Huang-Namburete/15d7207472ec91485e5d4d17934acada5d783>.
- Huang, R., Noble, J.A., Namburete, A.I., 2018. Omni-supervised learning: Scaling up to large unlabelled medical datasets. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing, Cham, pp. 572–580. doi:10.1007/978-3-030-00928-1_65. http://link.springer.com/10.1007/978-3-030-00928-1_65.
- Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. *Computing in Science & Engineering* 9 (3), 90–95. doi:10.1109/MCSE.2007.55.
- Huo, Y., Xu, Z., Xiong, Y., Aboud, K., Parvathaneni, P., Bao, S., Bermudez, C., Resnick, S.M., Cutting, L.E., Landman, B.A., 2019. 3D whole brain segmentation using spatially localized atlas network tiles. *Neuroimage* 194, 105–119. doi:10.1016/j.neuroimage.2019.03.041. <http://www.sciencedirect.com/science/article/pii/S1053811919302307>.
- Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 15 (9), 850–863.
- Inc., P. T., 2015. Collaborative data science. <https://plot.ly>.
- Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.-P., Heiland, S., Wick, W., et al., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum Brain Mapp* 40 (17), 4952–4964. <https://onlinelibrary.wiley.com/doi/full/10.1002/hbm.24750>.
- Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K., 2015. Spatial transformer networks. *arXiv preprint arXiv:1506.02025*.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. *Neuroimage* 62 (2), 782–790. doi:10.1016/j.neuroimage.2011.09.015. 20 YEARS OF fMRI.
- Kainz, B., Keraudren, K., Kyriakopoulou, V., Rutherford, M., Hajnal, J.V., Rueckert, D., 2014. Fast fully automatic brain detection in fetal mri using dense rotation invariant image descriptors. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1230–1233.
- Kamarainen, J., 2012. Gabor features in image analysis. In: 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 13–14. doi:10.1109/IPTA.2012.6469502.
- Keraudren, K., Kuklisova-Murgasova, M., Kyriakopoulou, V., Malamateniou, C., Rutherford, M.A., Kainz, B., Hajnal, J.V., Rueckert, D., 2014. Automated fetal brain segmentation from 2d mri slices for motion correction. *Neuroimage* 101, 633–643.
- Khalili, N., Moeskops, P., Claessens, N.H., Scherpenzeel, S., Turk, E., de Heus, R., Benders, M.J., Viergever, M.A., Pluijm, J.P., Išgum, I., 2017. Automatic segmentation of the intracranial volume in fetal MR images. In: *Fetal, Infant and Ophthalmic Medical Image Analysis*. Springer, pp. 42–51.
- Khalili, N., Turk, E., Benders, M., Moeskops, P., Claessens, N., de Heus, R., Franx, A., Wagenaar, N., Breur, J., Viergever, M., et al., 2019. Automatic extraction of the intracranial volume in fetal and neonatal MR scans using convolutional neural networks. *Neuroimage: Clinical* 24, 102061.
- Kim, M.S., Jeanty, P., Turner, C., Benoit, B., 2008. Three-Dimensional sonographic evaluations of embryonic brain development. *Journal of Ultrasound in Medicine* 27 (1), 119–124. doi:10.7863/jum.2008.27.1.119. <https://onlinelibrary.wiley.com/doi/abs/10.7863/jum.2008.27.1.119>.
- Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A., 2016. Deep mri brain extraction: a 3D convolutional neural network for skull stripping. *Neuroimage* 129, 460–469. doi:10.1016/j.neuroimage.2016.01.024. <http://www.sciencedirect.com/science/article/pii/S1053811916000306>.
- Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46 (3), 786–802. doi:10.1016/j.neuroimage.2008.12.037. <http://www.sciencedirect.com/science/article/pii/S1053811908012974>.
- Klein, A., Ghosh, S.S., Avants, B., Yeo, B.T., Fischl, B., Ardekani, B., Gee, J.C., Mann, J.J., Parsey, R.V., 2010. Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage* 51 (1), 214–220. doi:10.1016/j.neuroimage.2010.01.091. <http://www.ncbi.nlm.nih.gov/pubmed/20123029>.
- Kuklisova-Murgasova, M., Cifora, A., Napolitano, R., Papageorgiou, A., Quaghebeur, G., Rutherford, M.A., Hajnal, J.V., Noble, J.A., Schnabel, J.A., 2013. Registration of 3D fetal neurosonography and mri. *Med Image Anal* 17 (8), 1137–1150. doi:10.1016/j.media.2013.07.004. <http://www.sciencedirect.com/science/article/pii/S1361841513001114>.
- Li, J., Wang, Y., Lei, B., Cheng, J.-Z., Qin, J., Wang, T., Li, S., Ni, D., 2017. Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. *IEEE J Biomed Health Inform* PP. doi:10.1109/JBHI.2017.2703890. 1–1.
- Li, Y., Khanal, B., Hou, B., Alansary, A., Cerrolaza, J., Sinclair, M., Matthew, J., Gupta, C., Knight, C., Kainz, B., Rueckert, D., 2018. Standard Plane Detection in 3D Fetal Ultrasound Using an Iterative Transformation Network, pp. 392–400.
- Makropoulos, A., Counsell, S.J., Rueckert, D., 2018. A review on automatic fetal and neonatal brain MRI segmentation. *Neuroimage* 170, 231–248. doi:10.1016/j.neuroimage.2017.06.074. [files/144/Makropoulos et al. - 2018 - A review on automatic fetal and neonatal brain MRI.pdf](https://arxiv.org/abs/1404.0001) <http://www.ncbi.nlm.nih.gov/pubmed/28666878>.
- McKinney, W., et al., 2010. Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference*, Vol. 445, pp. 51–56.
- Miao, S., Wang, Z.J., Liao, R., 2016. A cnn regression approach for real-time 2D/3D registration. *IEEE Trans Med Imaging* 35 (5), 1352–1363. doi:10.1109/TMI.2016.2521800.
- Mohseni Salehi, S.S., Erdogmus, D., Gholipour, A., 2017. Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging. *IEEE Trans Med Imaging* 36 (11), 2319–2330. doi:10.1109/TMI.2017.2721362.

- Moser, F., Huang, R., Papageorgiou, A.T., Papież, B.W., Namburete, A.I.L., 2020. Automated fetal brain extraction from clinical ultrasound volumes using 3d convolutional neural networks. In: Zheng, Y., Williams, B.M., Chen, K. (Eds.), *Medical Image Understanding and Analysis*. Springer International Publishing, Cham, pp. 151–163.
- Namburete, A.I., Stebbing, R.V., Kemp, B., Yaqub, M., Papageorgiou, A.T., Alison Noble, J., 2015. Learning-based prediction of gestational age from ultrasound images of the fetal brain. *Med Image Anal* 21 (1), 72–86. doi:10.1016/j.media.2014.12.006. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4339204/>.
- Namburete, A.I., Xie, W., Noble, J.A., 2017. Robust regression of brain maturation from 3D fetal neurosonography using CRNs. In: Cardoso, M.J., Arbel, T., Melbourne, A., Bogunovic, H., Moeskops, P., Chen, X., Schwartz, E., Garvin, M., Robinson, E., Trucco, E., Ebner, M., Xu, Y., Makropoulos, A., Desjardin, A., Vercauteren, T. (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10554 LNCS. Springer International Publishing, Cham, pp. 73–80. doi:10.1007/978-3-319-67561-9_8. <http://link.springer.com/10.1007/978-3-319-67561-9>.
- Namburete, A.I., Xie, W., Yaqub, M., Zisserman, A., Noble, J.A., 2018. Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning. *Med Image Anal* 46, 1–14. doi:10.1016/j.media.2018.02.006. <http://www.sciencedirect.com/science/article/pii/S1361841518300306>.
- Nestares, O., Heeger, D., 2000. Robust multiresolution alignment of MRI brain volumes. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 43, 705–715.
- Paladini, D., Malinger, G., Birnbaum, R., Monteagudo, A., Pilu, G., Salomon, L., Timor-Tritsch, I., 2021. Isug practice guidelines (updated): sonographic examination of the fetal central nervous system. part 2: performance of targeted neurosonography. *Ultrasound in Obstetrics & Gynecology*.
- Paladini, D., Malinger, G., Monteagudo, A., Pilu, G., Timor-Tritsch, I., Toi, A., 2007. Sonographic examination of the fetal central nervous system: Guidelines for performing the 'basic examination' and the 'fetal neurosonogram'. <https://obgyn.onlinelibrary.wiley.com/doi/full/10.1002/uog.3909>.
- Papageorgiou, A.T., Ohuma, E.O., Altman, D.G., Todros, T., Ismail, L.C., Lambert, A., Jaffer, Y.A., Bertino, E., Gravett, M.G., Purwar, M., Noble, J.A., Pang, R., Victoria, C.G., Barros, F.C., Carvalho, M., Salomon, L.J., Bhutta, Z.A., Kennedy, S.H., Villar, J., 2014. International standards for fetal growth based on serial ultrasound measurements: the fetal growth longitudinal study of the INTERGROWTH-21st project. *The Lancet* 384 (9946), 869–879. doi:10.1016/S0140-6736(14)61490-2. <http://www.thelancet.com/article/S0140673614614902/fulltext>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pilu, G., Ghi, T., Carletti, A., Segata, M., Perolo, A., Rizzo, N., 2007. Three-dimensional ultrasound examination of the fetal central nervous system. *Ultrasound in Obstetrics & Gynecology* 30 (2), 233–245. doi:10.1002/uog.4072. <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.4072>.
- Prager, R.W., Ijaz, U.Z., Gee, A., Trece, G.M., 2010. Three-dimensional ultrasound imaging. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 224 (2), 193–223.
- Prayer, D., Kasprian, G., Krampl, E., Ulm, B., Witzani, L., Prayer, L., Brugger, P.C., 2006. MRI of normal fetal brain development. *Eur J Radiol* 57 (2), 199–216. doi:10.1016/j.ejrad.2005.11.020. *Fetal Imaging*.
- Prayer, D., Malinger, G., Brugger, P.C., Cassidy, C., De Catte, L., De Keersmaecker, B., Fernandes, G.L., Glanc, P., Gonçalves, L.F., Gruber, G.M., Laifer-Narin, S., Lee, W., Millischer, A.E., Molho, M., Neelavalli, J., Platt, L., Pugash, D., Ramaekers, P., Salomon, L.J., Sanz, M., Timor-Tritsch, I.E., Tutschek, B., Twickler, D., Weber, M., Ximenes, R., Raine-Fenning, N., 2017. ISUOG Practice guidelines: performance of fetal magnetic resonance imaging. *Ultrasound in Obstetrics & Gynecology* 49 (5), 671–680. doi:10.1002/uog.17412. <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.17412> <https://obgyn.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/uog.17412>.
- Rajchl, M., Lee, M.C., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M.A., Hajnal, J.V., Kainz, B., et al., 2016. Deepcut: object segmentation from bounding box annotations using convolutional neural networks. *IEEE Trans Med Imaging* 36 (2), 674–683.
- Rajchl, M., Lee, M.C., Schrans, F., Davidson, A., Passerat-Palmbach, J., Tarroni, G., Alansary, A., Oktay, O., Kainz, B., Rueckert, D., 2016. Learning under distributed weak supervision. *arXiv preprint arXiv:1606.01100*.
- Ranzini, M., Fidon, L., Ourselin, S., Modat, M., Vercauteren, T., 2021. Monai: Monai-based fetal brain MRI deep learning segmentation. *arXiv preprint arXiv:2103.13314*.
- Ray, J.G., Vermeulen, M.J., Bharatha, A., Montanera, W.J., Park, A.L., 2016. Association between MRI exposure during pregnancy and fetal and childhood outcomes. *JAMA - Journal of the American Medical Association* 316 (9), 952–961. doi:10.1001/jama.2016.12126. <http://www.ncbi.nlm.nih.gov/pubmed/27599330>.
- Rodríguez, G.V., Cosío, F.A., Huerta, M.C., Olivares, H.B., Ramírez, B.E., 2015. Automatic segmentation of the cerebellum in ultrasound volumes of the fetal brain. *Revista Mexicana de Ingeniería Biomédica* 36 (2), 121–129.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597*. <http://arxiv.org/abs/1505.04597>.
- Ruder, S., 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.
- Student, 1908. The probable error of a mean. *Biometrika* 1–25.
- Studholme, C., 2011. Mapping fetal brain development in utero using magnetic resonance imaging: the big bang of brain mapping. *Annu Rev Biomed Eng* 13 (1), 345–368. doi:10.1146/annurev-bioeng-071910-124654. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3682118/>.
- Suzuki, K., 2017. Overview of deep learning in medical imaging. *Radiol Phys Technol* 10 (3), 257–273. doi:10.1007/s12194-017-0406-5.
- Taimouri, V., Gholipour, A., Velasco-Annis, C., Estroff, J.A., Warfield, S.K., 2015. A template-to-slice block matching approach for automatic localization of brain in fetal MRI. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 144–147. doi:10.1109/ISBI.2015.7163836.
- Taleb, Y., Schweitzer, M., Studholme, C., Koob, M., Dietemann, J.-L., Rousseau, F., 2013. Automatic template-based brain extraction in fetal mr images. <https://hal.archives-ouvertes.fr/hal-00879089>.
- Torres-Barrena, J., Piella, G., Masoller, N., Gratacós, E., Eixarch, E., Ceresa, M., Ballester, M.A.G., 2019. Segmentation and classification in MRI and US fetal imaging: recent trends and future prospects. *Med Image Anal* 51, 61–88. doi:10.1016/j.media.2018.10.003. <http://www.ncbi.nlm.nih.gov/pubmed/30390513>.
- Tourbier, S., Velasco-Annis, C., Taimouri, V., Hagmann, P., Meuli, R., Warfield, S.K., Bach Cuadra, M., Gholipour, A., 2017. Automated template-based brain localization and extraction for fetal brain MRI reconstruction. *Neuroimage* 155, 460–472. doi:10.1016/j.neuroimage.2017.04.004. <http://www.sciencedirect.com/science/article/pii/S1053811917302938>.
- Velásquez-Rodríguez, G., Cosío, F.A., Ramírez, B.E., 2015. Automatic segmentation of the fetal cerebellum using spherical harmonics and gray level profiles. In: 11th International Symposium on Medical Information Processing and Analysis, Vol. 9681. International Society for Optics and Photonics, p. 968114. <https://doi.org/10.1117/12.2207833>.
- Velásquez-Rodríguez, G., Cosío, F.A., Ramírez, B.E., 2015. Automatic segmentation of the fetal cerebellum using spherical harmonics and gray level profiles. In: 11th International Symposium on Medical Information Processing and Analysis, Vol. 9681. International Society for Optics and Photonics, p. 968114.
- Villar, J., Altman, D., Purwar, M., Noble, J., Knight, H., Ruyan, P., Cheikh Ismail, L., Barros, F., Lambert, A., Papageorgiou, A., et al., 2013. The objectives, design and implementation of the intergrowth-21st project. *BJOG: An International Journal of Obstetrics & Gynaecology* 120, 9–26.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2.
- Wilcoxon, F., 1992. Individual Comparisons by Ranking Methods. In: *Breakthroughs in Statistics*. Springer, pp. 196–202.
- Wright, R., Khanal, B., Gomez, A., Skelton, E., Matthew, J., Hajnal, J.V., Rueckert, D., Schnabel, J.A., 2018. Lstm spatial co-transformer networks for registration of 3D fetal US and MR brain images. In: *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*. Springer, pp. 149–159.
- Wright, R., Kyriakopoulou, V., Ledig, C., Rutherford, M.A., Hajnal, J.V., Rueckert, D., Aljabar, P., 2014. Automatic quantification of normal cortical folding patterns from fetal brain MRI. *Neuroimage* 91, 21–32. <https://doi.org/10.1016/j.neuroimage.2014.01.034>.
- Wright, R., Toussaint, N., Gomez, A., Zimmer, V., Khanal, B., Matthew, J., Skelton, E., Kainz, B., Rueckert, D., Hajnal, J.V., et al., 2019. Complete fetal head compounding from multi-view 3d ultrasound. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 384–392. https://link.springer.com/chapter/10.1007/978-3-030-32248-9_43.
- Wyburd, M. K., Hesse, L. S., Alias, M., Jenkinson, M., Papageorgiou, A. T., Haak, M. C., Namburete, A. I. L., 2021. Assessment of Regional Cortical Development Through Fissure Based Gestational Age Estimation in 3D Fetal Ultrasound. Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis. Springer. https://link.springer.com/chapter/10.1007/978-3-030-87735-4_23.
- Yaqub, M., Cuingnet, R., Napolitano, R., Roundhill, D., Papageorgiou, A., Ardon, R., Noble, J.A., 2013. Volumetric segmentation of key fetal brain structures in 3d ultrasound. In: *International Workshop on Machine Learning in Medical Imaging*. Springer, pp. 25–32.
- Yaqub, M., Kelly, B., Papageorgiou, A.T., Noble, J.A., 2015. Guided random forests for identification of key fetal anatomy and image categorization in Ultrasound Scans. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 687–694. https://link.springer.com/chapter/10.1007/978-3-319-24574-4_82.
- Yaqub, M., Kelly, B., Papageorgiou, A.T., Noble, J.A., 2017. A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. *Ultrasound in Medicine and Biology* 43 (12), 2925–2933. doi:10.1016/j.ultrasmedbio.2017.07.013. <http://www.ncbi.nlm.nih.gov/pubmed/28958729>.
- Yaqub, M., Napolitano, R., Ioannou, C., Papageorgiou, A., Noble, J.A., 2012. Automatic detection of local fetal brain structures in ultrasound images. In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 1555–1558.