2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

DEEP LEARNING WITH ULTRASOUND PHYSICS FOR FETAL SKULL SEGMENTATION

Juan J. Cerrolaza[1], Matthew Sinclair[1], Yuanwei Li[1], Alberto Gomez[2], Enzo Ferrante[3], Jaqueline Matthew[2], Chandni Gupta[2], Caroline L. Knight[4,5], Daniel Rueckert[1]

[1]Biomedical Image Analysis Group, Imperial College London, UK
[2]Division of Imaging Sciences and Biomedical Engineering, King's College London, UK
[3]Universidad Nacional del Litoral / CONICET, Santa Fe, Argentina
[4]Department of Women and Children's Health, King's College London, UK
[5]Fetal Medicine Unit, Guy's and St Thomas' NHS Foundation Trust, London, UK

ABSTRACT

2D ultrasound (US) is still the preferred imaging method for fetal screening. However, 2D biometrics are significantly affected by the inter/intra-observer variability and operator dependence of a traditionally manual procedure. 3DUS is an alternative emerging modality with the potential to alleviate many of these problems. This paper presents a new automatic framework for skull segmentation in fetal 3DUS. We propose a two-stage convolutional neural network (CNN) able to incorporate additional contextual and structural information into the segmentation process. In the first stage of the CNN, a partial reconstruction of the skull is obtained, segmenting only those regions visible in the original US volume. From this initial segmentation, two additional channels of information are computed inspired by the underlying physics of US image acquisition: an angle incidence map and a shadow casting map. These additional information channels are combined in the second stage of the CNN to provide a complete segmentation of the skull, able to compensate for the fading and shadowing artefacts observed in the original US image. The performance of the new segmentation architecture was evaluated on a dataset of 66 cases, obtaining an average Dice coefficient of 0.83 ± 0.06. Finally, we also evaluated the clinical potential of the new 3DUS-based analysis framework for the assessment of cranial deformation, significantly outperforming traditional 2D biometrics (100% vs. 50% specificity, respectively).

Index Terms— Segmentation, deep learning, fully convolutional network, fetal imaging, ultrasound, skull.

1. INTRODUCTION

Two-dimensional ultrasound (2DUS) is the primary screening modality for fetal screening, thanks to the lack of harmful effects on the fetus and mother, relatively low cost, and possibility of real time imaging and reporting. Nowadays, a second trimester US examination is routinely performed at 18-20 weeks of gestation in most countries, including a fetal survey, and a complete anatomical screening examination. As part of this comprehensive analysis, 2DUS-based biometry (i.e., sonographic measurements of the fetal anatomy) has been extensively used to indirectly assess the growth and well-being of the fetus, and estimating fetal weight [1]. The assessment of the fetal skull is an essential part of routine sonographic examination, including head circumference (HC), biparietal diameter (BPD), and occipitofrontal diameter (OFD). However, these biometrics are prone to errors, still relying on 2D measurements manually extracted from a specific anatomical planes (i.e., the transthalamic or transventricular planes). This subjectivity and operator dependence can affect the diagnostic capability of US-based fetal screening, limiting reproducibility, and may hinder the early detection of malformations [2]. Moreover, the early detection of cranial malformations, such as dolichocephaly, or brachycephaly, requires the detail structural analysis of the skull, its curvilinear bones, and boundaries, which may be difficult to visualize and quantify in a single 2D plane.

Despite its limitations, manual 2DUS-based biometry remains the current gold standard in obstetric sonography [3]. 3DUS has the potential to mitigate many of these drawbacks, whose superior diagnostic power, higher reproducibility, and consistency has been reported by numerous studies [4]. The possibility to scan and analyze volumetric data provides a more comprehensive, and anatomically consistent view of the complex anatomy of the fetus, as well as the definition of a new generation of objective and more accurate volumetric biometrics. In this context, the development of automatic segmentation methods for 3DUS is paramount to alleviate the tedious and subjective manual delineation process. Despite the growing interest of the image analysis community in the development of new segmentation strategies for sonographic images, the complete automatic segmentation of the skull in fetal 3DUS has not yet been satisfactorily addressed.

US-imaging remains, arguably, the most difficult imaging modality upon which to perform segmentation, suffering from low signal-to-noise ratio, signal attenuation and dropout, and missing boundaries. The use of predefined parametric curves was exploited by early works on 2DUS [5][6] as strategy to deal with fuzzy and incomplete skull boundaries, approximating the head contour to an ellipse. Shape priors were also used by the first skull segmentation

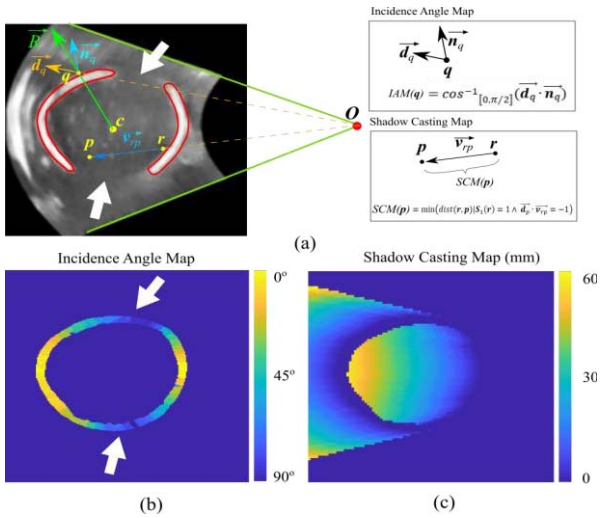

Figure 1. US-based information channels. (a) US axial view of a fetal skull (marked with a red contour) and geometric description of the incidence angle and shadow casting maps (see Section 2.2). White arrows indicate missing regions suffering from fading effects, corresponding to regions tangent to the US wavefront in (b). (b) Incidence angle map. The incidence angles corresponding to the whole skull are shown for illustration purposes only. (c) Shadow casting map corresponding to the complete skull shown in (a).

framework for 3DUS presented by Chen et al. [7]. In the proposed registration-based framework, the authors used a fetal phantom to incorporate shape constraints into a registration-based method. However, the use of restrictive shape priors limits the flexibility of these models, and their capacity to capture the local shape abnormalities of skulls presenting mild or severe cranial deformity. A more general approach was presented by Namburete and Noble [8], using a classic random forest classifier to identify cranial pixels in 2DUS images. Recently, Cerrolaza et al. [9] presented the first automatic framework for the segmentation of the skull in 3DUS, using a random forest classifier that combined contextual features and structured semantic labels. However, despite the promising results, competitive with the state-of-the-art deep learning-based methods, only a partial reconstruction of the skull was provided.

Frequently, the contours of fetal skulls appear discontinuous and irregular in clinical sonographic images, due to fading and shadowing artefacts, both inherent to the imaging physics of US. In US imaging, it is known that those edges tangent to the propagation direction of the wavefront can be affected by low contrast and fading effects, due to the dependence of the echo with the angle of incidence. On the other hand, acoustic shadows casted by the section of the skull closer to the US probe can also obscure and hamper the detection of more distant regions of the cranium (see Fig. 1(a)). In this paper, we present the first fully automatic framework able to accurately segment the whole skull in fetal 3DUS. Inspired by the underlying physics of US imaging, we propose a two-stage cascade deep convolutional neural network (2S-CNN) architecture, tailored to deal with the specific challenges raised by the segmentation of fetal US images.

2. METHODOLOGY

Recently, CNN-based architectures have become the current state-of-the-art for many image segmentation tasks, including organ segmentation in medical imaging. Unlike most common implementations, where a single CNN is used, we propose a new two-stage cascade approach as alternative architecture to incorporate additional information channels into the segmentation process. These new channels are specifically tailored to deal with the inherent challenges and limitations of fetal 3DUS (i.e., shadowing and fading effects), providing relevant information for the complete reconstruction of the skull. The overall flow diagram of the proposed framework is depicted in Figure 2(a).

2.1. Two-Stage Cascade CNN Architecture

The proposed 2S-CNN separate the segmentation process into two stages, using a 3D convolutional U-Net [11] (i.e., a fully convolutional network which includes shortcut connections between a contracting encoder and a successive expanding decoder; see Fig. 2(b)) as basic structural element of the architecture. Taking the original 3DUS volume as single input channel, the aim of the first block (B1) is to generate a partial reconstruction of the skull by segmenting the cranial bone visible in the US data. Despite the numerous artifacts and low signal-to-noise ratio of US imaging, bone tissue can be partially identified by its hyperechoic nature, and its characteristic structural curved pattern. Thanks to the concatenation of multiple convolutional and pooling layers, the proposed CNN-based architecture learns specific filters at multiple resolutions able to identify these distinctive features, thus providing an accurate, yet incomplete, initial estimation of the skull.

Based on this initial estimation, the aim of the second block (B2) is to generate a full reconstruction of the skull, completing those gaps and missing regions generated by the combination of fading and shadowing effects. Inspired by the auto-context architecture [10], the output probability map from B1 is used as additional input channel for B2, which provides valuable contextual and structural image information. Additionally, the input to B2 is completed with two additional channels, the incidence angle map (IAM; see Fig. 1(b)) and the shadow casting map (SCM; see Fig. 1(c)), both derived from the initial partial segmentation obtained in B1 (see Section 2.2 for details). The IAM and SCM, provide relevant complementary information regarding the underlying physics of US imaging, the angle of incidence of the US wavefront, and the shadowing effect caused by the bone structures closer to the probe, respectively. The impact of IAM and SCM in the segmentation process is analyzed in Section 3.

2.2. Additional Information Channels

2.2.1. Incidence Angle Map

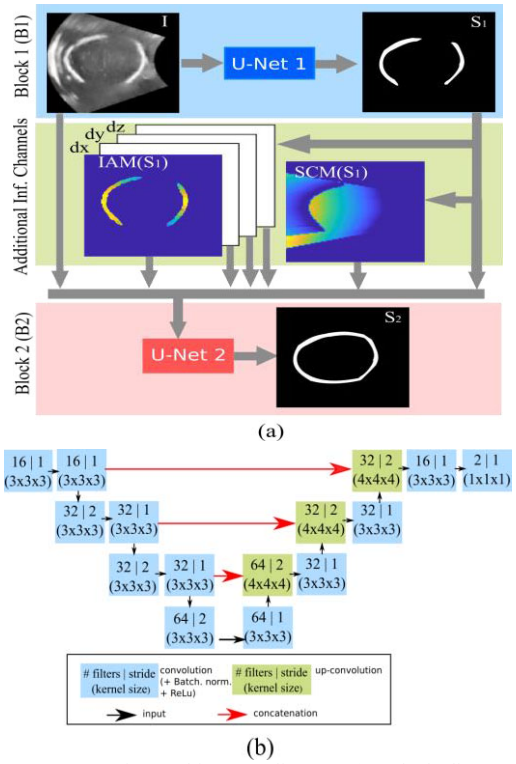

Figure 2. Segmentation architecture diagram. (a) Block diagram of the two-stage convolutional network. (b) Detail of the 3DU-Net architecture used in the proposed 2S-CNN segmentation framework.

Suppose $\boldsymbol{I}$ is the original US volume, and $\boldsymbol{S}_1$ represents the partial segmentation obtained in B1 (i.e., the cranial tissue visible in $\boldsymbol{I}$), with $\boldsymbol{S}_1(\boldsymbol{p}) = 1$ if the voxel $\boldsymbol{p} \in \mathbb{R}^3$ belongs to the foreground, and $\boldsymbol{S}_1(\boldsymbol{p}) = 0$ otherwise. For every voxel $\boldsymbol{p}|\boldsymbol{S}_1(\boldsymbol{p}) = 1$, the value of $IAM(\boldsymbol{p})$ is approximated by the angle of incidence at $\boldsymbol{q} \in \mathbb{R}^3$, the closest point to $\boldsymbol{p}$ located on the outer surface of the skull. Let the center of mass of $\boldsymbol{S}_1$, $\boldsymbol{C} = (\sum \boldsymbol{p}|\boldsymbol{S}_1(\boldsymbol{p}) = 1)/(\sum \boldsymbol{S}_1(\boldsymbol{p}))$ approximate the center of the skull. Assuming the skull is an approximately convex structure, the points on the surface of the skull $\{\boldsymbol{q}\}$, can be easily identified via ray casting, using $\boldsymbol{C}$ as origin. In particular, given a ray $\vec{R}$, $\boldsymbol{q} = argmax_{\boldsymbol{p}} dist(\boldsymbol{C}, \boldsymbol{p})|\boldsymbol{S}_1(\boldsymbol{p}) = 1 \wedge \boldsymbol{p} \in \vec{R}$, where $dist(\cdot,\cdot)$ is the Euclidean distance. Suppose now $\overrightarrow{\boldsymbol{d}_q}$ is a unitary direction vector that represents the propagation direction of the wavefront at $\boldsymbol{q}$, and that $\overrightarrow{\boldsymbol{n}_q}$ is the unitary vector normal to the skull at that point. Defining the center of the US probe as $\boldsymbol{O}$ (which can be easily interpolated from the US image cone), $\overrightarrow{\boldsymbol{d}_q}$ can be approximated as $\overrightarrow{\boldsymbol{d}_q} \approx \overrightarrow{\boldsymbol{O},\boldsymbol{q}}/\|\overrightarrow{\boldsymbol{O},\boldsymbol{q}}\|$, where $\|\cdot\|$ represents the Euclidean norm. Finally, for every surface point $\boldsymbol{q}$, $IAM(\boldsymbol{q}) = cos^{-1}{}_{[0,\pi/2]}(\overrightarrow{\boldsymbol{d}_q} \cdot \overrightarrow{\boldsymbol{n}_q})$. The IAM is finally completed with three additional channels containing the x, y, and z components of the unitary directional vector at every voxel of the image (i.e., dx_p, dy_p, and dz_p).

2.2.2. Shadow Casting Map

TABLE I
SKULL SEGMENTATION ERROR

	DC	JI	SSD (mm)
Single-Channel	0.78 ± 0.07	0.65 ± 0.09	1.11 ± 0.74
SCM	0.77 ± 0.01	0.63 ± 0.11	1.02 ± 0.83
IAM	0.83 ± 0.07	0.71 ± 0.08	0.99 ± 0.22
IAM+SCM	0.83 ± 0.06	0.70 ± 0.08	0.98 ± 0.52
IUA	0.84 ± 0.01	0.72 ± 0.02	0.75 ± 0.12

Segmentation accuracy. The table presents the average error and standard deviation for the Dice coefficient (DC), Jaccard index (JI), and symmetric surface distance (SSD) for different configurations of the architecture: single-channels (i.e., not including SCM and IAM), using only SCM, using only IAM, and combining both additional information channels, SCM and IAM. The table also includes the intra-user accuracy (IUA).

The SCM is defined by a distance map where, for every location $\boldsymbol{p}$, $SCM(\boldsymbol{p})$ represents the minimum distance from $\boldsymbol{p}$ to the section of the skull detected in B1, in the opposite direction to the wavefront propagation in $\boldsymbol{p}$ (see Fig.1(a)). Let $\overrightarrow{\boldsymbol{v}_{rp}}$ represents the unitary vector pointing from $\boldsymbol{r}$ to $\boldsymbol{p}$, i.e., $\overrightarrow{\boldsymbol{v}_{rp}} = \overrightarrow{\boldsymbol{r},\boldsymbol{p}}/\|\overrightarrow{\boldsymbol{r},\boldsymbol{p}}\|$. Thus, $SCM(\boldsymbol{p})$ can be mathematically defined as $SCM(\boldsymbol{p}) = \min(dist(\boldsymbol{r},\boldsymbol{p})|\boldsymbol{S}_1(\boldsymbol{r}) = 1 \wedge \overrightarrow{\boldsymbol{d}_p} \cdot \overrightarrow{\boldsymbol{v}_{rp}} = -1)$.

3. EXPERIMENTAL RESULTS AND DISCUSION

In the experiments we used a database of 66 fetal 3DUS images. The mean gestational age (GA) was 24.7 weeks, ranging from 20 to 36 weeks. The volumes were acquired from an axial transventricular plane using a Philips Epiq7G scanner with a X6-1 xMatrix array transducer. All the volumes were preprocessed using non-local means filtering, and resized to 96 × 96 × 96 voxels. For each image, the skull was delineated manually by experienced members of the team, and under the supervision of an expert radiologist. The set of images was randomly divided into two groups, using 52 images for training and 14 for testing. We included training data augmentation applying random affine transformations on-the-fly, thus resulting in as many different images as training iterations. The networks (B1 and B2) were trained separately using cross-entropy loss, and stochastic gradient descent with momentum (Adam with learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.995$) on Theano. We ran a total of 500 epoch on an NVIDIA® GeForce® GTX 1080 Ti (which took a total of 10 hours).

3.1. Segmentation Accuracy

Table I shows the segmentation accuracy for different configurations of the 2S-CNN architecture, including a single-channel implementation (i.e., a classic auto-context approach where only the original US volume and the output probability map from B1 are used in B2). To analyze the contribution of each new information channel (SCM and IAM), we also tested the 2S-CNN when incorporating only SCM, only IAM, or both together (SCM+IAM). The intra-user accuracy (IUA) is also included. It can be observed how the best performance is provided by the combination of both channels, SCM+IAM, with a DC of 0.83 ± 0.06. However, no

statistically significant difference was observed (p-value < 0.05 using a Wilcoxon paired signed non-parametric test) when using only the IAM (DC 0.83 ± 0.07). This suggests that only the IAM provides additional relevant information to complete the skull in B2. Moreover, no difference was observed when comparing the implementation using only the SCM (DC 0.77 ± 0.01) and the single-channel approach (DC 0.78 ± 0.07). The incorporation of IAM into the 2S-CNN model provides similar results in terms of DC (0.83 ± 0.07) and JI (0.71 ± 0.08) to the intra-user accuracy (DC 0.84 ± 0.01 and JI 0.72 ± 0.02). The average segmentation time was 0.08 sec. per volume.

3.2. Volumetric Biometrics for Skull Shape Assessment: A pilot study

In this section we present a pilot study to evaluate the potential of the proposed 3D segmentation method for fetal skull shape assessment. We analyze the value of 3D-based shape analysis for the identification of cranial deformities, such as dolichocephaly, a condition where the head becomes disproportionately long and narrow, for example due to mechanical forces associated with breech positioning in utero. The ground truth was established by an experienced obstetric sonographer and a fetal medicine specialist, identifying two cases with dolichocephaly in our database. The traditional 2D cephalic index CI is calculated by BPD/OFD. We proposed an alternative 3DCI derived from the volumetric analysis of the skull. Once the complete skulls were segmented, we used principal component analysis to create a statistical model of the cranium, defining the 3DCI as the distance to the mean shape of the skull normalized to the patient's gestational age. The 5th percentile threshold was used to identify potentially abnormal cases in both metrics, CI and 3DCI. The accuracy, specificity, and sensitivity for the abnormal shape identification were 98%, 100% and 98% for 3DCI, significantly outperforming the CI with 90%, 50%, and 92%, respectively (p-value < 0.005 using McNemar's test). Despite the limited number of cases with skull deformation in our database, this pilot study suggests the potential of 3D-based biometry to provide objective and accurate assessment of the fetal head. We plan to extend this study in the future.

4. CONCLUSIONS

This paper presents the first fully automatic framework for the complete segmentation and quantification of the skull in fetal 3DUS. We propose a new two-stage CNN-based architecture that integrate additional information channels in the segmentation process. The incorporation of these new maps, derived from the imaging physics of US, provides relevant contextual and structural information for the complete reconstruction of the skull. In particular, the incident angle map has proven particularly useful in regenerating the skull, allowing the network to compensate for the fading effects caused by the dependence of the contrast with the angle of incidence in US-imaging. The promising results in terms of segmentation accuracy and fetal skull assessment demonstrate the potential of the proposed framework to objectively assess fetal 3DUS and support better informed clinical decision making.

ACKNOWLEDGMENTS

This research was supported in part by the Marie Sklodoska-Curie Actions of the EU Framework Programme for Research and Innovation, under REA grant agreement 706372, and by the Wellcome Trust and EPSRC, Innovative Engineering for Health Award [102431]. The co-author Enzo Ferrante is also beneficiary of an AXA Research Fund postdoctoral grant.

5. REFERENCES

[1] U. Sovio, et al., "Screening for fetal growth restriction with universal third trimester ultrasonography in nulliparous women in the Pregnancy Outcome Prediction (POP) study: a prospective cohort study," *Lancet*, 386, pp. 2089-2097, 2015.

[2] I. Sarris, et al., "Intra- and interobserver variability in fetal ultrasound measurements," Ultrasound Obstet. Gynecol., 39(3), pp. 266-273, 2012.

[3] T. Kiserud, et al. "The WHO fetal growth charts: A multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight", PLoS Med. 14(1), pp. 2017.

[4] I. Lee, et al., "A review of three-dimensional ultrasound applications in fetal growth restriction", Journal of Medical Ultrasound, 20(3), pp. 142-149, 2012.

[5] W. Lu, et al., "Automated fetal head detection and measurement in ultrasound images by iterative randomized hough transform", Ultrasound in Medicine and Biology, 31(7), pp. 929-936, 2005.

[6] Y. Shenet al., "Fetal skull analysis in ultrasound images based on iterative randomized Hough transform". SPIE 7265, 2009.

[7] H.C. Chen, et al., "Registration-based segmentation of three-dimensional ultrasound images for quantitative measurement of fetal craniofacial structure". Ultrasound in Medicine and Biology, 38(5), pp. 811-823, 2012.

[8] A.I.L. Namburete, J.A. Nobel, "Fetal cranial segmentation in 2D ultrasound images using shape properties of pixel clusters", ISBI, pp. 720-723, 2013.

[9] J.J. Cerrolaza, et al.,"Fetal skull segmentation in 3D ultrasound via structured geodesic random forest", Int. Workshop on Fetal and Infant Image Analysis, pp. 25-32, 2017.

[10] Z. Tu and X. Bai, "Auto-context and its application to high-level vision task and 3D brain image segmentation", IEEE TPAMI, 32(10), pp.1744-1757, 2010.

[11] O. Cicek, A. Abdulkadir, S.S. Lienkamp, O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation", MICCAI, pp. 424-432, 2016.