



3D Fetal Skull Reconstruction from 2DUS via Deep Conditional Generative Networks

Juan J. Cerrolaza¹(✉), Yuanwei Li¹, Carlo Biffi¹, Alberto Gomez²,
Matthew Sinclair¹, Jacqueline Matthew², Caronline Knight², Bernhard Kainz¹,
and Daniel Rueckert¹

¹ Biomedical Image Analysis Group, Imperial College London, London, UK
j.cerrolaza-martinez@imperial.ac.uk

² Division of Imaging Sciences and Biomedical Engineering,
King's College London, London, UK

Abstract. 2D ultrasound (US) is the primary imaging modality in antenatal healthcare. Despite the limitations of traditional 2D biometrics to characterize the true 3D anatomy of the fetus, the adoption of 3DUS is still very limited. This is particularly significant in developing countries and remote areas, due to the lack of experienced sonographers and the limited access to 3D technology. In this paper, we present a new deep conditional generative network for the 3D reconstruction of the fetal skull from 2DUS standard planes of the head routinely acquired during the fetal screening process. Based on the generative properties of conditional variational autoencoders (CVAE), our reconstruction architecture (REC-CVAE) directly integrates the three US standard planes as conditional variables to generate a unified latent space of the skull. Additionally, we propose HiREC-CVAE, a hierarchical generative network based on the different clinical relevance of each predictive view. The hierarchical structure of HiREC-CVAE allows the network to learn a sequence of nested latent spaces, providing superior predictive capabilities even in the absence of some of the 2DUS scans. The performance of the proposed architectures was evaluated on a dataset of 72 cases, showing accurate reconstruction capabilities from standard non-registered 2DUS images.

Keywords: Generative model · Variational autoencoder
Fetal ultrasound

1 Introduction

Ultrasound (US) screening is the primary imaging modality for the prenatal evaluation of growth, gestational age estimation, and early structural abnormalities detection. Thanks to its non-ionizing nature, relative low-cost, and real-time visualization, a detailed mid-trimester morphology US scan is routinely performed at 18–22 weeks of gestation in most countries. As part of the examination, the quantification of 2D biometrics is extensively used to evaluate the

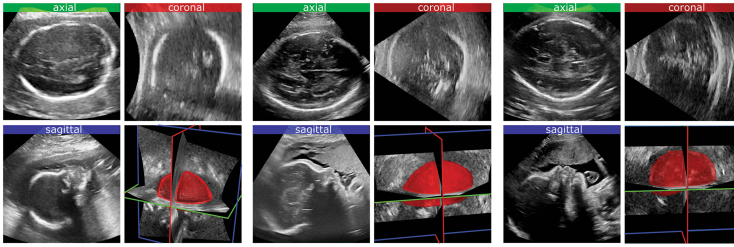


Fig. 1. Fetal standard US scans of the head. Example of the axial (green), sagittal (blue), and coronal (red) standard planes of three patients acquired in freehand 2DUS during the routine mid-trimester US examination. The image also shows a 3D representation of the skull manually segmented from the corresponding 3DUS volume.

growth and well-being of the fetus. However, the detection rates of fetal abnormalities often remain below the recommended values [1], showing significant differences between industrialized and developing countries. Recently, new technological solutions have been proposed to assist in the acquisition of standard views in freehand 2DUS scans [2], improving reproducibility and reducing operator bias. However, the screening process is still constrained by the inherent limitation of 2D-based biometry to accurately characterize the true 3D anatomy of the fetus. Recent studies have reported on the advantages of 3DUS in the evaluation of fetal anatomy [3], and the superior diagnostic potential of 3D shape analysis over traditional 2D biometrics for the early detection and characterization of cranial deformations [4] (e.g., dolichocephaly, or craniosynostosis). Moreover, studies on 3DUS perception by medical professionals showed their interest in having access to 3D-based information during prenatal screening [5]. However, despite the reported advantages, there are three main factors that have notably hampered the adoption of 3D-based biometry by the obstetric community: (1) the lack of experience with 3DUS often slows down the acquisition process as compared to traditional 2DUS; (2) the need for new image processing solutions that enable efficient real-time analysis, visualization and reconstruction of volumetric information [6]; and (3) the limited access to 3D transducers, especially in developing countries [7]. Aware of these limitations, this paper presents a new practical approach for 3D head biometry, the 3D reconstruction of the fetal skull from standard planes in 2DUS, the current gold standard in obstetric radiology (see Fig. 1).

The reconstruction of 3D anatomical structures from a limited number of 2D views was previously studied as a strategy to reduce cost and radiation exposure to patients: e.g. reconstruction of the femur, pelvic bone, or vertebrae from X-ray images as alternative to 3D computed tomography [8,9]. Typically, these methods rely on deformable statistical models to incorporate *a priori* anatomical constraints to the generation process. Using contour- or registration-based strategies, these approaches often require complex density models of the bones to create virtual X-ray images that guide the generation process [8]. However,

these simulation-based strategies are impractical in the less controlled environment of fetal sonography. Fetal US is arguably one of the most challenging imaging modalities, suffering from low signal-to-noise ratio, signal attenuation and dropout, as well as random shadows and occlusions frequently caused by unpredictable movements of the fetus.

With the advent of deep learning-based techniques, convolutional neural networks (CNNs) have become the current state-of-the-art for many medical imaging tasks, including semantic segmentation, and object recognition [2]. However, the prediction of 3D structures remains a relatively unexplored area, with only a few works using deep networks to address the mapping from 2D images to 3D volumes. Of special interest to our work is the TL-embedding network (TL-net) [10]. In this architecture, the authors use a 3D convolutional autoencoder (ConvAE) to generate a latent space of the 3D structures, using a second CNN to map 2D images to the corresponding coordinates in the latent space. Similarly, Wang and Fang [11] also used a common latent vector space between the 2D and the 3D image domains, combining ConvAE with adversarial learning to control the matching process in an unsupervised manner. Alternatively, Choy et al. [12] proposed a recurrent network to incrementally refine the reconstruction as more views of the object are provided. However, the proposed architectures normally rely on the availability of large-scale datasets, and/or the artificial generation of realistic 2D images from 3D synthetic models, both an important limitation when working with medical images.

In this paper, we present a new architecture to address the problem of fetal skull reconstruction from multiple 2DUS standard views of the head. First we propose a deep generative network using the conditional variational autoencoder (CVAE) formulation. Additionally, we present an alternative hierarchical framework, based on the different clinical relevance of each view. Imposing a specific hierarchy on the 2DUS standard views, the model learns a sequence of nested latent spaces, which allows the network to operate effectively even in the absence of any of the predictive views.

2 Method

In this paper, we formulate the 3D reconstruction problem in the form of a conditional manifold learning task. We use the CVAE framework [13] to create deep generative networks able to reconstruct the fetal skull, using freehand 2DUS standard views as predictive variables. These predictors are incorporated in the optimization model in the form of conditional variables, thus modulating the latent space of 3D skulls learned by the network.

Suppose \mathbf{Y} represents a 3D parameterization of the fetal skull (e.g., a binary voxel map of the skull), with \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{X}_3 representing the corresponding 2DUS standard views acquired in the coronal, sagittal, and axial plane, respectively (see Fig. 1). For simplicity of notation we denote $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$ as $\mathbf{X}_{1,2,3}$. We seek a generative model that learns the conditional distribution $P(\mathbf{Y}|\mathbf{X}_{1,2,3})$, so it produces a close approximation of \mathbf{Y}^i , for a given observation $\mathbf{X}_{1,2,3}^i$. In the context of variational autoencoders, the generative process

is modeled by means of a latent d -dimensional variable, \mathbf{z} , with some known simple distribution (typically $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$). Thanks to this latent variable, the model can generate new instances of the target structure (i.e., the fetal skull) by randomly sampling values of \mathbf{z} . However, it would be very difficult in practice to directly infer $P(\mathbf{Y}|\mathbf{X}_{1,2,3})$ without sampling a large number of \mathbf{z} values. Alternatively, we introduce a new function Q (e.g., a high-capacity function here parameterized in the form of a CNN), which can generate values of \mathbf{z} likely to produce \mathbf{Y} s. Using Bayes' rule, we have $E_{\mathbf{z} \sim Q} [\log P(\mathbf{Y}|\mathbf{z}, \mathbf{X}_{1,2,3})] = E_{\mathbf{z} \sim Q} [\log P(\mathbf{z}|\mathbf{Y}, \mathbf{X}_{1,2,3}) - \log P(\mathbf{z}|\mathbf{X}_{1,2,3}) + \log P(\mathbf{Y}|\mathbf{X}_{1,2,3})]$. Rearranging and subtracting $E_{\mathbf{z} \sim Q} [\log Q(\mathbf{z})]$ from both sides yields

$$\begin{aligned} \log P(\mathbf{Y}|\mathbf{X}_{1,2,3}) - \mathcal{D}_{KL}[Q(\mathbf{z})\|P(\mathbf{z}|\mathbf{Y}, \mathbf{X}_{1,2,3})] = \\ E_{\mathbf{z} \sim Q} [\log P(\mathbf{Y}|\mathbf{z}, \mathbf{X}_{1,2,3})] - \mathcal{D}_{KL}[Q(\mathbf{z})\|P(\mathbf{z}|\mathbf{X}_{1,2,3})], \end{aligned} \quad (1)$$

where $\mathcal{D}_{KL}[a\|b] = E_{\mathbf{z} \sim Q} [\log(a) - \log(b)]$ represents the Kullback-Leibler (KL) divergence. Typically, the function Q is defined as $Q(\mathbf{z}|\mathbf{Y}, \mathbf{X}_{1,2,3}) = \mathcal{N}(\mathbf{z}|\mu(\mathbf{Y}, \mathbf{X}_{1,2,3}), \Sigma(\mathbf{Y}, \mathbf{X}_{1,2,3}))$ where μ and Σ are arbitrary, deterministic functions learned from the data, and parameterized in the form of CNNs (Σ is constrained to be a diagonal matrix). Since $P(\mathbf{z}|\mathbf{X}_{1,2,3})$ is still $\sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (i.e., assuming \mathbf{z} is sampled independently of $\mathbf{X}_{1,2,3}$ at test time), this choice of Q allows us to compute $\mathcal{D}_{KL}[Q(\mathbf{z})\|P(\mathbf{z}|\mathbf{X}_{1,2,3})]$ as the KL-divergence between two Gaussians, which has a closed-form solution [13]. Optimizing the right hand side via stochastic gradient descent, and assuming Q is a high-capacity function which can approximate $P(\mathbf{z}|\mathbf{Y}, \mathbf{X}_{1,2,3})$, the KL-term on the left hand side of (1) will tend to 0. That is, we will be directly optimizing $P(\mathbf{Y}|\mathbf{X}_{1,2,3})$. At training time, we make the sampling of \mathbf{z} differentiable with respect to μ and Σ by using the “reparameterization trick” [13], and defining $\mathbf{z}^i = \mu(\mathbf{Y}^i, \mathbf{X}_{1,2,3}^i) + \eta * \Sigma(\mathbf{Y}^i, \mathbf{X}_{1,2,3}^i)$, where $\eta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Based on equation (1), the reconstruction network can be implemented using CNNs, whose structure, at training time, resembles a traditional ConvAE. The function Q takes the form of the encoder, “encoding” \mathbf{Y} and $\mathbf{X}_{1,2,3}$ into a d -dimensional latent space \mathbf{z} , via μ and Σ . In the proposed architecture, we use a multi-branch CNN to model Q , using 3D convolutional filters for \mathbf{Y} , and a separate view-specific bank of 2D filters for each standard view. The outputs of each branch are concatenated and mapped to two separate fully-connected layers to generate $\mu(\mathbf{Y}, \mathbf{X}_{1,2,3})$ and $\Sigma(\mathbf{Y}, \mathbf{X}_{1,2,3})$, which will be combined with η to create \mathbf{z} . Finally, the decoder of the network, modeled also as a CNN, reconstructs \mathbf{Y} given \mathbf{z} and $\mathbf{X}_{1,2,3}$. The conditional dependency on $\mathbf{X}_{1,2,3}$ is explicitly modeled by the concatenation of \mathbf{z} with the vector representation of $\mathbf{X}_{1,2,3}$ (see Fig. 2(a)). At test time, the decoder operates as a generative reconstruction network given the coronal, \mathbf{X}_1 , sagittal, \mathbf{X}_2 , and axial, \mathbf{X}_3 2DUS views, generating valid 3D skulls by sampling $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. In particular, we generate the highest-confidence prediction with $\mathbf{z} = \mathbf{0}$.

With this configuration, the reconstruction network requires the three standard views to approximate the 3D fetal skull. However, it is common in clinical practice that not all the standard views of the head are routinely acquired,

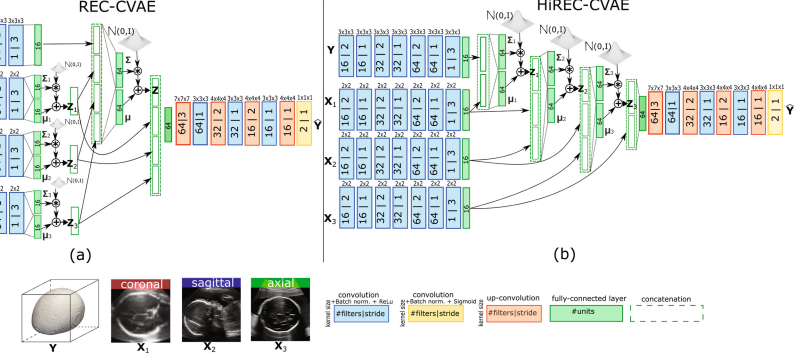


Fig. 2. Deep generative networks for the reconstruction of the fetal skull. (a) REC-CVAE: Reconstruction network based on the conditional variational autoencoder framework. (b) HiREC-CVAE: Hierarchical reconstruction network.

thus limiting the potential utility of our model in retrospective studies. During the mid-trimester examination, axial views of the head (i.e., \mathbf{X}_3) are routinely acquired and used for 2D biometrics measurements (e.g., head circumference and biparietal diameter). Additionally, a sagittal view (\mathbf{X}_2) is also normally acquired to ensure there is a normal face/head shape. However, coronal planes (\mathbf{X}_1) are usually only included as part of a dedicated scan, used to clarify suspicious findings. To make the reconstruction network more flexible and operative in the absence of some of the standard views, we propose two alternative architectures. In the first model, we define the conditional variables as three multidimensional Gaussians $\mathcal{N}(\mathbf{0}, \mathbf{I})$, \mathbf{z}_1 , \mathbf{z}_2 , and \mathbf{z}_3 . Thus, if \mathbf{X}_1 , or \mathbf{X}_2 are missing at test time, we still can approximate \mathbf{Y} with the same network, by sampling \mathbf{z}_1 , or \mathbf{z}_2 . The resulting objective function is $CE(\mathbf{Y}, \hat{\mathbf{Y}}) - \nu(\mathcal{D}_{KL}[\mathbf{z} \parallel \mathcal{N}(\mathbf{0}, \mathbf{I})] + \sum_{i=1,2,3} \mathcal{D}_{KL}[\mathbf{z}_i \parallel \mathcal{N}(\mathbf{0}, \mathbf{I})])$, where the first term represents the cross-entropy (CE) between \mathbf{Y} and the reconstructed $\hat{\mathbf{Y}}$, and ν is a constant set to 0.01. See Fig. 2(a) for a detailed description of the proposed reconstruction network using CVAE (REC-CVAE). While REC-CVAE represents a more direct implementation of the CVAE formulation, we propose a second configuration that explicitly incorporates the predefined hierarchy of the conditional variables as a cascade of conditional blocks, HiREC-CVAE (see Fig. 2(b)). In the first block, only \mathbf{X}_1 is used as conditional variable for \mathbf{Y} , thus defining a latent space, \mathbf{z}_1 , for $\mathbf{Y}|\mathbf{X}_1$. The sagittal and axial planes are incorporated in the second and third blocks, producing \mathbf{z}_2 , and \mathbf{z}_3 , respectively. Unlike REC-CVAE, where a single generative latent space is defined, $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ can be interpreted as a set of nested latent spaces. Now, in the absence of one of the views, we are sampling from the corresponding latent-spaces, that effectively integrate the missing components into a manifold of fetal skulls. The resulting objective function is $CE(\mathbf{Y}, \hat{\mathbf{Y}}) - \nu(\sum_{i=1,2,3} \mathcal{D}_{KL}[\mathbf{z}_i \parallel \mathcal{N}(\mathbf{0}, \mathbf{I})])$.

3 Results

Both approaches for fetal skull reconstruction were evaluated on a dataset of 72 cases. For each case, one 3DUS volume of the head, and at least one image from each of the standard views in the coronal, sagittal, and transventricular axial planes were acquired by an experienced obstetric sonographer during routine mid-trimester examination (i.e., more than one image per standard view were available for some cases). The mean gestational age was 24.7 weeks, ranging from 20 to 36 weeks. The images were acquired using a Philips Epiq7G US system, with a X6-1 xMatrix array transducer. The data were preprocessed using non-local means filtering, and resampled to isotropic size of 0.50 mm per dimension; the 3D volumes and the 2D standard planes were resized to $96 \times 96 \times 96$ voxels, and to 96×96 pixels, respectively, using cropping and zero-padding if needed. For each volume, the skull was manually delineated under the supervision of an expert radiologist. In this study, we consider a smooth reconstruction of the cranial region located above the transthalamic plane, including the parietal and frontal bones, and excluding the facial bones, sutures and fontanels. The set of manual segmentations were aligned and used as ground-truth for skull reconstruction. No registration was used for the standard planes, using only flipping or mirroring to provide orientation consistency (e.g., the fetus is approximately looking up in the sagittal views as shown in Fig. 1). The patients were randomly divided in two groups, using 58 cases for training and 14 for testing. This process was repeated three times, always using a different subset for testing. During training, data augmentation was applied to the ground-truth volumes, applying random anisotropic scaling in the three orthogonal axes. This strategy allowed us, not only to expand the training set, but also to replicate potential fetal skull anomalies, such as dolichocephaly or trigonocephaly. One image for each standard view was randomly selected, if more than one scan were available, and deformed independently using the corresponding anisotropic scaling, and random translation and rotation. The reconstruction capability of REC-CVAE and HiREC-CVAE was compared with the TL-net, a CNNs-based state-of-the-art architecture for 2D-to-3D reconstruction [10] (see Sect. 1). The TL-net configuration is depicted in Fig. 3(a). All the networks were trained for 1000 epochs on an NVIDIA® GeForce® 1080 Ti (approx. 12 hours per network), using stochastic gradient descent with momentum (Adam with learning rate = 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.995$) in Theano, using a small batch size of 1.

Table 1. The table presents the average and standard deviation for the Dice’s coefficient (DC), sensitivity (SEN.), and precision (PPV) of the reconstruction of the fetal skull, and the effect of using three, two, or one standard US views as predictors.

	axial + sagittal + coronal			axial + sagittal			axial		
	DC	SEN.	PPV	DC	SEN.	PPV	DC	SEN.	PPV
REC-CVAE	0.91 ± 0.02	0.91 ± 0.05	0.91 ± 0.06	0.86 ± 0.05	0.88 ± 0.13	0.87 ± 0.09	0.83 ± 0.06	0.86 ± 0.15	0.84 ± 0.13
HiREC-CVAE	0.91 ± 0.04	0.89 ± 0.06	0.93 ± 0.06	0.89 ± 0.05	0.90 ± 0.10	0.91 ± 0.08	0.86 ± 0.05	0.86 ± 0.11	0.90 ± 0.08
TL-net	0.89 ± 0.03	0.93 ± 0.05	0.86 ± 0.07	0.89 ± 0.05	0.90 ± 0.11	0.90 ± 0.06	0.85 ± 0.04	0.88 ± 0.05	0.80 ± 0.09

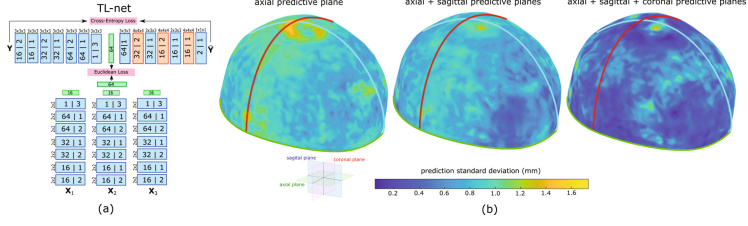


Fig. 3. (a) TL-net architecture (see Fig. 2 for a description of the constituent blocks). The network assumes the three views are used for the reconstruction. A separate network with two (\mathbf{X}_2 , \mathbf{X}_3) or one (\mathbf{X}_3) predictors is defined if any of the views are missing. (b) Prediction uncertainty in HiREC-CVAE for a varying number of predictors. The images represent the standard deviation of $N = 50$ different predictions randomly generated from the latent spaces.

Table 1 shows the reconstruction accuracy for the three architectures when using three (coronal, sagittal, and axial), two (sagittal and axial) or one (axial) US standard views as predictors. During testing, the 3D encoder branch was disabled, setting the latent variables \mathbf{z} and \mathbf{z}_1 to $\mathbf{0}$, in REC-CVAE and HiREC-CVAE, respectively. In the absence of the coronal or also the sagittal view, the corresponding latent variables, \mathbf{z}_1 and \mathbf{z}_2 in REC-CVAE, and \mathbf{z}_2 and \mathbf{z}_3 in HiREC-CVAE, were also set to $\mathbf{0}$ (see Fig. 2). For TL-net, three different case-specific configurations of the network were trained in order to deal with a varying number of available standard views. When using the three planes, REC-CVAE and HiREC-CVAE showed slightly higher (although not statistically significant) performance than the TL-net (e.g., $DC_{REC-CVAE} = 0.91 \pm 0.02$, $DC_{HiREC-CVAE} = 0.91 \pm 0.04$, and $DC_{TL-net} = 0.89 \pm 0.03$). However, the performance of REC-CVAE was significantly affected when one or two of the views were missing ($DC_{REC-CVAE} = 0.86 \pm 0.05$, and 0.83 ± 0.06 , respectively). One important limitation in REC-CVAE is the independent Gaussian encoding of the predictors. While this allows the network to operate in the absence of any of the views by automatically generating a valid input from the predefined distributions, the resulting semi-optimal code passed to the decoder (e.g., $(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{z}_3)$) can produce an inaccurate reconstruction of the skull. In HiREC-CAE, the potential correlation between the predictors is effectively encoded through a three-level nested space of latent variables, showing good reconstruction capabilities even when the coronal, or also the sagittal planes are absent ($DC_{HiREC-CVAE} = 0.89 \pm 0.05$, and 0.86 ± 0.05 , respectively). Similar performance was obtained with the dedicated case-specific TL-nets ($DC_{TL-net} = 0.89 \pm 0.05$ and 0.85 ± 0.04), although a separate case-specific network is needed for each scenario. Moreover, the TL-nets require a three-stage training process [10], instead of the end-to-end approach used in REC-CVAE and HiREC-CVAE.

Finally, the generative capability of HiREC-CVAE can be exploited to generate a confidence map of the reconstructed skull. Here, we define the confidence maps as the standard deviation of N ($N = 50$ in Fig. 3(b)) different predic-

tions, randomly generated by sampling \mathbf{z}_1 , \mathbf{z}_2 , or \mathbf{z}_3 from $\mathcal{N}(\mathbf{0}, \mathbf{I})$. These maps can be generated in real-time (each prediction is generated in ~ 0.04 sec.), and used as an indirect indicator of the reconstruction accuracy, thus informing the sonographer about the need of additional views for a more accurate result.

4 Discussion and Conclusion

This paper presents the first deep conditional generative network for the 3D reconstruction of the fetal skull from freehand non-aligned 2D scans of the head. We propose two different models, REC-CVAE, based on the CVAE formulation, and HiREC-CVAE, an alternative configuration that effectively encodes a predefined hierarchy of the predictive variables. Both networks learn a low-dimensional embedding representation of the skull, which guarantees the anatomical consistency of the reconstructions. Moreover, the use of a predefined distribution model for the latent space allows the networks to operate even when some of the 2DUS images are missing. The results demonstrate the potential of the networks for the 3D reconstruction and characterization of the fetal skull from 2DUS standard planes. This framework can contribute significantly to the popularization of 3D-based fetal screening, allowing for large-scale 3D-based biometrics studies that include a wide and varied demographic representation, including cases from developing countries with limited access to 3D transducer technology. In the near future, we will continue exploring the clinical value of the proposed framework and its potential for the early detection and characterization of congenital deformations. We also plan to explore the potential of deep conditional generative networks for 3D cardiac reconstruction from cardiac cine MRI.

Acknowledgement. This research was supported in part by the Marie Skłodowska-Curie Actions of the EU Framework Programme for Research and Innovation, under REA grant agreement 706372.

References

1. Springett, A., et al.: Congenital Anomaly Statistics, 2012. England and Wales. Technical report, British Isles Network of Congenital Anomaly Registers (2014)
2. Baumgartner, C., et al.: Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE TMI* **36**, 2204–2215 (2017)
3. Lima, J., et al.: Biometry and fetal weight estimation by two-dimensional and three-dimensional ultrasonography: an intraobserver and interobserver reliability and agreement study. *Ultrasound Obstet. Gynecol.* **40**, 186–93 (2012)
4. Matthew, J., et al. Novel 3D ultrasound-based metric to assess the fetal skull: a pilot study. In: *BMUS Annual Meeting* (2017)
5. Lee, S., et al.: Prenatal three-dimensional ultrasound: perception of sonographers, sonologists and undergraduate students. *Ultrasound Obstet. Gynecol.* **1**(30), 77–80 (2007)
6. Cerrolaza, J.J.: Fetal skull segmentation in 3D ultrasound via structured geodesic random forest. In: Cardoso, M.J. (ed.) *FIFI/OMIA -2017. LNCS*, vol. 10554, pp. 25–32. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67561-9_3

7. Shah, S., et al.: Perceived barriers in the use of ultrasound in developing countries. *Crit. Ultrasound J.* **7**(1), 7–11 (2015)
8. Whitmarsh, T., et al.: Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy x-ray absorptiometry. *IEEE TMI* **12**(30), 2101–2114 (2011)
9. Ehlke, M., et al.: Fast generation of virtual x-ray images for reconstruction of 3D anatomy. *IEEE TVCG* **19**(12), 2673–2682 (2013)
10. Girdhar, R., Fouhey, D.F., Rodriguez, M., Gupta, A.: Learning a predictable and generative vector representation for objects. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9910, pp. 484–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_29
11. Wang, L., Fang, Y.: Unsupervised 3D reconstruction from a single image via adversarial learning (2017). CoRR, abs/1711.09312
12. Choy, C.B., Xu, D., Gwak, J.Y., Chen, K., Savarese, S.: 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9912, pp. 628–644. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_38
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: *ICLR 2014* (2014)