

Towards Whole Placenta Segmentation At Late Gestation Using Multi-View Ultrasound Images

Veronika A. Zimmer^a, Alberto Gomez^a, Emily Skelton^a, Nicolas Toussaint^a,
Tong Zhang^a, Bishesh Khanal^{a,b}, Robert Wright^a, Yohan Noh^{a,c}, Alison Ho^d,
Jacqueline Matthew^a, Joseph V. Hajnal^a, and Julia A. Schnabel^a

^aSchool of Biomedical Engineering and Imaging Sciences
King’s College London, London, UK

^bNepal Applied Mathematics and Informatics Institute for Research (NAAMII)

^cDepartment of Mechanical and Aerospace Engineering, Brunel University London,
Uxbridge, UK

^dWomens Health Academic Centre, Kings College London, London, UK

Abstract. We propose a method to extract the human placenta at late gestation using multi-view 3D US images. This is the first step towards automatic quantification of placental volume and morphology from US images along the whole pregnancy beyond early stages (where the entire placenta can be captured with a single 3D US image). Our method uses 3D US images from different views acquired with a multi-probe system. A whole placenta segmentation is obtained from these images by using a novel technique based on 3D convolutional neural networks. We demonstrate the performance of our method on 3D US images of the placenta in the last trimester. We achieve a high Dice overlap of up to 0.8 with respect to manual annotations, and the derived placental volumes are comparable to corresponding volumes extracted from MR.

1 Introduction

Fetal ultrasound (US) is the primary imaging modality to monitor fetal development. While the fetal body, especially the fetal brain, are subjects of intensive research, only few methods exist to study the placenta in utero [11]. Placental development and function influence fetal health yet only placental side and cord insertion are routinely assessed using US. Limiting factors are the large size of the placenta at late gestation, high variation in shape and position, and the limited field-of-view (FoV) and lack of contrast in US. Placenta magnetic resonance image (MRI) acquisition overcomes some of those challenges as it provides a large FoV and an excellent soft-tissue contrast. Recently, the first method to assess the placenta in utero in a standardized way was presented using fetal MRI [5]. However, fetal MRI is corrupted by motion artifacts due to fetal motion and maternal breathing and fetal MRI reconstruction is an active field of research [10]. US is still the standard screening tool because, in contrast to MRI, it is performed in real-time, is relatively inexpensive, and widely available. MRI is generally only used upon referral from the US clinic to gain insight into specific

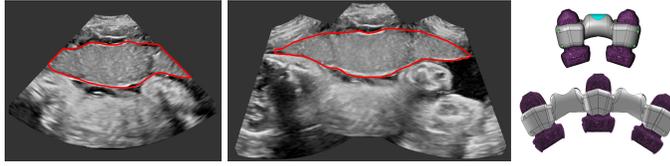


Fig. 1. Ultrasound placenta imaging using a multi-probe system. Left: Slice of 3D US image covering only part of the placenta; Middle: Slice of 3D multi-view image of the whole placenta; Right: Physical multi-probe holder for two and three US probes.

conditions. US based placenta segmentation could therefore lead to automatic volume quantification, morphology and function, in clinical routine scans. In [9], a semi-automatic method based on the random walker algorithm was proposed to segment the placenta. State-of-the-art segmentation methods using convolutional neural networks (CNNs) have been used in [7, 4] and in [13] additionally for the fetus and the gestational sac. These methods focused on early pregnancies between 10-14 weeks of gestational age (GA), when the placenta is small enough to fit in the limited FoV of US. Routine anomaly screening is performed at 20 weeks GA but placental volume at later gestations may be of benefit in predicting/monitoring fetal development. Therefore a larger field of view is required to capture the whole placenta by US.

Multi-view imaging can be used to extend the FoV of a single image. For example, the entire placenta can be captured by acquiring, aligning and fusing multiple 3D US images (Fig. 1). In previous works [12, 6, 1], registration algorithm and/or tracker information were employed to align the images and provide multi-view US. The resulting image has an extended FoV, and view-dependent artifacts such as shadows can be minimized through the additional signal information from multiple views [14]. Aligning US placenta remains however challenging, due to the lack of salient features to drive the registration process. External tracking, on the other hand, can provide position information of the US probe but is oblivious to maternal and fetal motion.

In this work, we introduce, for the first time, a pipeline to extract the whole placenta at late gestation. The approach consist of three stages: first, multi-view image acquisition, second, multi-view image fusion, and third, multi-view placenta segmentation. The multi-view US images are acquired using a time-interleaved multi-probe US system without the need of image registration. We present a voxel-wise image fusion method to combine the images and to reduce view-dependent artifacts, and compare four approaches based on CNNs to extract the whole placenta from the multi-view images.

We test our pipeline on a dataset of 3D US images to estimate placental volume in the last trimester of pregnancy. We successfully fuse multi-view images to get an extended FoV and are able to extract the placenta. The derived placental volumes are comparable to corresponding volumes extracted from MR. To the best of our knowledge, this is the first time the placenta is segmented at late gestation from US imaging.

2 Methods and Materials

The three stages of the proposed pipeline for whole placenta extraction (multi-view acquisition, fusion and segmentation) are described below.

2.1 Multi-probe ultrasound imaging

We acquire multiple US images using an in-house US signal multiplexer which allows to connect multiple Philips X6-1 probes to a Philips EPIQ V7 US system. The multiplexer switches rapidly between up to three probes so that images from each probe are acquired in a time-interleaved fashion. The manual operation of the transducers is very slow compared to the acquisition frame rate. Therefore, for the purpose of data processing, consecutive images are assumed to have been acquired simultaneously over a small time window. We designed a physical device that fixes the probes in an angle of 30° to each other, which ensures a large overlap between the images, and allows easy and comfortable operation (see Fig. 1).

2.2 Multi-view image fusion

In our application, the goal is to combine multiple images of the placenta to extend the FoV while minimizing view-dependent artifacts. The multi-view fusion method proposed in [14] produces high quality fusion but is computationally expensive. We propose a simplification of that method by replacing the B-spline based fusion for a voxel-wise fusion as follows.

Let I_1, \dots, I_V , with $I_v : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}, i = 1, \dots, V$ be images of the same object taken from V views. Their spatial correspondences are known through spatial transformations $\phi_v : \Omega_v \subset \mathbb{R}^d \rightarrow \Omega \subset \mathbb{R}^d$. The fused image $I_F : \Omega_F \rightarrow \mathbb{R}$ with $\Omega_1, \dots, \Omega_V \subset \Omega_F$ at point $\mathbf{x} \in \Omega_F$ is obtained by

$$I_F(\mathbf{x}) = \frac{\sum_{v=1}^V w_v(\mathbf{x}) \cdot \mathbb{1}_{\mathbf{x} \in \Omega_v} \cdot (I_v \circ \phi_v)(\mathbf{x})}{\sum_{v=1}^V w_v(\mathbf{x})}$$

with weight function $w : \mathbb{R}^d \rightarrow \mathbb{R}$. In other words, the intensity of a point \mathbf{x} of the fused image is calculated by the weighted mean of corresponding points in the single images. The weight of a (transformed) data point \mathbf{x} are formulated as a function of the depth in the US image with respect to the probe position $\mathbf{b} \in \mathbb{R}^d$ and the beam angle $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ in the same way as in [14]. In effect, image points with a strong signal (to correct for shadow artifacts) and at a position close to the center of the US frustum (where the quality of the image is typically the best) will receive higher weights.

2.3 Whole placental segmentation

Semantic segmentation using neural networks In recent years, convolutional neural networks (CNNs) have shown excellent segmentation results, outperforming conventional methods in quality of the segmentation and in speed

[8, 3]. In a supervised CNN approach, the segmentation of an object is learned only driven by the data from a training dataset $\mathcal{T} = \{(I_n, S_n), n = 1, \dots, N\}$, with images $I_n : \mathbb{R}^d \rightarrow \mathbb{R}$ and ground truth segmentations $S_n : \mathbb{R}^d \rightarrow \{0, 1\}$ (in our case US images with $d = 3$ and manual annotations of the placenta). The model f estimates for an input image I the segmentation map S : $S = f(I, \Theta)$. During the training process of model $f(I, \Theta)$ with the training set \mathcal{T} , the parameters of the network $\Theta \in \mathbb{R}^P$ are optimized to minimize a loss function \mathcal{L} , which measures the agreement between the ground truth S_n and the estimated segmentation of the model. The parameters Θ include the connection weights, biases and convolutional kernel weights.

Our segmentation network is based on the widely used U-net architecture [8] which is a fully convolutional neural network with an encoder-decoder architecture, a bottleneck layer in between and skip connections from encoder layers to decoder. We use ReLUs, strided convolution using max pooling for down-sampling, and zero padding. We set [32, 64, 128, 256] feature maps per layer where all the convolutional kernels and feature maps are in 3D.

For training, we resample all images to size $128 \times 128 \times 128$. During training, we minimize the Dice loss using the Adam optimizer and a learning rate of 0.001. We augment our dataset by image flipping in x- and z-axis (avoid flipping the image upside down to keep the correct positioning of the US frustum), intensity rescaling by $\pm 10\%$ and small random translation (up to five pixels in all directions). Rotations are avoided because that would produce non-realistic view direction-dependent image features.

Multi-view image segmentation At later gestation, the field-of-view of US is too small to capture the whole placenta. Multiple probes, as described above, or placenta sweeps using an appropriate registration or tracking method to align the images, can be used to visualize the whole placenta in one US volume. Those images differ not only in the view of the placenta, but also in view-dependent artifacts, such as shadows or attenuation. To provide a consistent segmentation of the whole placenta across multiple images, we propose four CNN-based variants, which make use of the multi-view information in different ways (see Fig. 2).

- (S1) The model f_1 is trained on N single US images I_n^s with manual annotations $S_n^s, n = 1, \dots, N$ of the placenta, without using any information of correspondences between different views of the same placenta. The resulting segmentations for the individual images are aligned and fused using maximum intensity voting to obtain the segmentation of the whole placenta.
- (S2) The model f_2 is trained on $M < N$ fused multi-view images I_m^{mv} with manual annotations $S_m^{mv}, m = 1, \dots, M$. The training set is smaller compared to approach (S1) since one image I_m^{mv} is the fusion of two or three images I_n^s . The fused images are resampled to size $192 \times 128 \times 128$. The larger size in the first dimension is to account for the larger FoV in the fused images.
- (S3) The model f_1 , trained on individual images, is re-trained as model f_3 using fused multi-view images (re-sampled to the size of individual images) which have been separately annotated. Using pre-trained weights from f_1 as initialization allows to train on the smaller dataset of fused images.

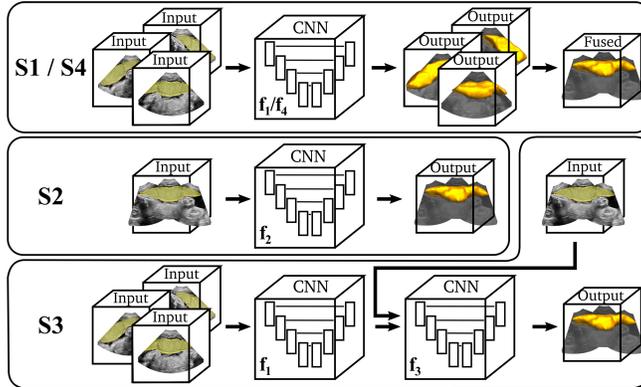


Fig. 2. Illustration of strategies to obtain multi-view placenta segmentation using 3D convolutional networks.

(S4) The last model f_4 is trained in a similar manner as f_1 , except that the individual annotations are obtained from the manual segmentations of the fused multi-view images by mapping them from the fused image space back to the single image space. This reduces the amount of manual segmentations to carry out for the same amount of training data. The manual segmentation task is easier since fused images have better image quality and larger FoV.

3 Experimental Results

3.1 Dataset

We used a dataset of 127 3D US images to test our pipeline, which were selected from 4D (3D+time) image streams from 30 patients covering different parts of the placenta. A subset of 94 images were acquired with a two-transducer (64) or a three-transducer (30) holder device, and the rest were acquired using a single transducer. Two patients were in the second trimester (24 and 25 weeks GA) and the others in the third trimester (29 – 34 weeks GA). We split the dataset into training (85 images), validation (16 images) and test set (26 images). When trained only on multi-view images (approaches (S2) and (S3)), the sets reduce to 27, 3 and 12 images for training, validation and testing, respectively.

3.2 Results

The results for placenta segmentation are shown in Table 1 and a representative example of whole placenta segmentations is shown in Fig. 3. The accuracy of the segmentation during training, validation and inference are calculated using the Dice overlap and the absolute volume difference relative to the ground truth segmentation. Methods (S1) and (S3) both achieve the best results with a mean Dice of 0.8 and a volume difference of around 16% for the multi-view images

Table 1. Segmentation accuracy measured using the Dice coefficient and relative volume difference (mean \pm std.).

| | | train | validate | test (single) | test (multi-view) |
|----|------------------|-----------------|-----------------|-----------------|-----------------------------------|
| S1 | # images | 85 | 16 | 26 | 12 |
| | Dice | 0.80 \pm 0.09 | 0.75 \pm 0.15 | 0.74 \pm 0.15 | 0.81 \pm 0.05 |
| | Δ Vol (%) | 16.1 \pm 17.9 | 27.9 \pm 22.5 | 26.2 \pm 22.1 | 16.5 \pm 12.2 |
| S2 | # images | 27 | 3 | – | 12 |
| | Dice | 0.75 \pm 0.12 | 0.84 \pm 0.10 | – | 0.76 \pm 0.17 |
| | Δ Vol (%) | 23.8 \pm 18.7 | 13.2 \pm 19.5 | – | 20.7 \pm 27.1 |
| S3 | # images | 85/27 | 16/3 | – | 12 |
| | Dice | 0.86 \pm 0.04 | 0.83 \pm 0.11 | – | 0.80 \pm 0.11 |
| | Δ Vol (%) | 11.9 \pm 7.7 | 16.8 \pm 21.3 | – | 15.9 \pm 19.9 |
| S4 | # images | 85 | 16 | 26 | 12 |
| | Dice | 0.70 \pm 0.15 | 0.61 \pm 0.24 | 0.65 \pm 0.18 | 0.72 \pm 0.12 |
| | Δ Vol (%) | 28.6 \pm 25.7 | 44.7 \pm 29.5 | 37.1 \pm 25.2 | 20.8 \pm 15.8 |

Table 2. Comparison of placental volume (PV, in mm³) extracted from US (manual and automatic) and MR (manual) segmentations (seg.). (S1) US: automatic seg. from US using method (S1); GT US: manual ground truth seg.; MR: manual MR seg.

| | GA | (S1) US PV | GT US PV | MR PV | (S1)/MR | GT/MR |
|---|------|------------|----------|-------|--------------|--------------|
| 1 | 32+2 | 549.6 | 583.9 | 588.8 | –6.66 | –0.83 |
| 2 | 30+1 | 505.5 | 539.9 | 534.1 | –5.35 | 1.09 |
| 3 | 33+1 | 616.6 | 605.5 | 664.1 | –7.15 | –8.82 |
| 4 | 31+2 | 584.8 | 614.5 | 578 | –1.18 | 6.31 |

of the whole placenta with the smallest volume difference obtained by (S3) at 15.9%. The additional CNN in (S3) did not improve the results from (S1) further. For (S2), the training was performed only on the multi-view images, and therefore fewer images were used for training. Unsurprisingly, this yields worse results than (S3), which was pre-trained. The worst results were obtained using (S4). This can be explained by the fact that although the ground truth segmentation incorporates information from all views, the network is trained on individual images and therefore not capable of learning the relation between different views of the same object. To test for significance, we performed a paired Wilcoxon signed-rank test on the Dice overlaps. The results for single images (only S1 and S4) are significantly different with $p < 0.001$. On multi-view images, we found a significant difference between S1 and S4 ($p = 0.018$) and S2 and S3 ($p = 0.02$).

Additionally, we compared the placental volume measured in MRI and US for a subset of the data. Table 2 shows the results using manual expert segmentations of the MRI data, the multi-view US data and automatic segmentation with method (S1). All US volumes differ less than 10% from the corresponding MR volume, suggesting that nearly the whole placenta is covered by the multi-view US images.

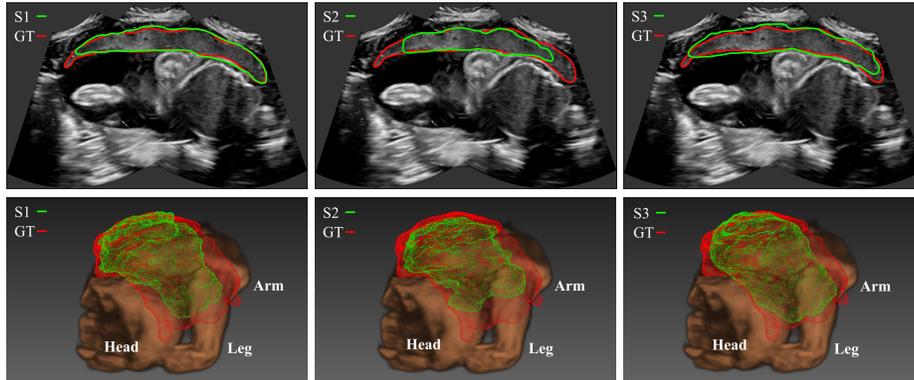


Fig. 3. Top row: 2D slice showing a cut of the 3D whole-placenta segmentation; Bottom row: 3D visualization of fetus and placenta. Automatic segmentations (in green) using (S1) (left), (S2) (centre) and (S3) (right). Ground truth is shown in red.

4 Conclusions

We proposed a fully automatic method for the extraction of whole placenta volumes at late gestation. The method consists of three stages, namely multi-view image acquisition, image fusion and placenta segmentation. We used CNNs with a standard architecture (U-net) for the segmentation procedure. We showed with four approaches how this can be adapted to multi-view placenta segmentation with a dice overlap of 0.8 and placental volumes comparable to MR.

The automatic segmentation underestimates the placental volume, both for the manual annotation in US (see Fig. 3) and the MR volume (see Table 2). This consistent negative bias could be explained by the poor contrast in US and requires further investigation. The manual annotation is challenging and thin parts of the placenta, especially at the boundaries, as shown in Fig. 3, are hard to delineate. Future work will include the refinement of the segmentation method. However, the negative bias with respect to the MR volume is also present in the manual annotations from US [2]. Apart from the poor contrast in US, some part of the placenta might still be missing in the multi-view images. In this work, the placental MR volume is extracted from the raw MR volume prior to motion correction and reconstruction, which explains also some of the differences between US and MR annotations. Our method enables future work for the comparison of US and MR placenta in more detail.

Our approach is suitable for real-time application, with an inference time of 40 ms. Enabling automatic placenta extraction from US in clinical routine can pave the way for placenta-based abnormality detection.

Acknowledgements. This work was supported by the Wellcome Trust IEH Award [102431], by the Wellcome/EPSRC Centre for Medical Engineering [WT203148/Z/16/Z] and by the National Institute for Health Research (NIHR)

Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Gomez, A., Bhatia, K., Tharin, S., Housden, J., Toussaint, N., Schnabel, J.A.: Fast registration of 3D fetal ultrasound images using learned corresponding salient points. In: Proc. FIFI@MICCAI, pp. 33–41 (2017)
2. Gomez, A., Oktay, O., Rueckert, D., Penney, G.P., Schnabel, J.A., Simpson, J.M., Pushparajah, K.: Regional differences in end-diastolic volumes between 3D echo and CMR in HLHS patients. *Frontiers in Pediatrics* **4**, 133 (2016)
3. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Med. Imag. Anal.* **42**, 60–88 (2017)
4. Looney, P., Stevenson, G.N., Nicolaidis, K.H., Plasencia, W., Molloholli, M., Natsis, S., Collins, S.L.: Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning. *JCI Insight* **3**(11) (2018)
5. Miao, H., Mistelbauer, G., Karimov, A., Alansary, A., Davidson, A., Lloyd, D.F., Damodaram, M., Story, L., Hutter, J., Hajnal, J.V., Rutherford, M., Preim, B., Kainz, B., Gröller, M.E.: Placenta maps: in utero placental health assessment of the human fetus. *IEEE Trans. Vis. & Comp. Graph.* **23**(6), 1612–1623 (2017)
6. Ni, D., Qu, Y., Yang, X., Chui, Y.P., Wong, T.T., Ho, S.S., Heng, P.A.: Volumetric ultrasound panorama based on 3D SIFT. In: Proc. MICCAI, pp. 52–60 (2008)
7. Oguz, B.U., Wang, J., Yushkevich, N., Pouch, A., Gee, J., Yushkevich, P.A., Schwartz, N., Oguz, I.: Combining deep learning and multi-atlas label fusion for automated placenta segmentation from 3D US. In: Proc. DATRA/PIPPi@MICCAI, pp. 138–148 (2018)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proc. MICCAI, pp. 234–241 (2015)
9. Stevenson, G.N., Collins, S.L., Ding, J., Impey, L., Noble, J.A.: 3-D ultrasound segmentation of the placenta using the random walker algorithm: reliability and agreement. *Ultrasound in Med. & Biol.* **41**(12), 3182–3193 (2015)
10. Torrents-Barrena, J., Piella, G., Masoller, N., Gratacós, E., Eixarch, E., Ceresa, M., González Ballester, M.Á.: Fully automatic 3D reconstruction of the placenta and its peripheral vasculature in intrauterine fetal MRI. *Med. Imag. Anal.* **54**, 263–279 (2019)
11. Torrents-Barrena, J., Piella, G., Masoller, N., Gratacós, E., Eixarch, E., Ceresa, M., González Ballester, M.Á.: Segmentation and classification in MRI and US fetal imaging: Recent trends and future prospects. *Med. Imag. Anal.* **51**, 61–88 (2019)
12. Wachinger, C., Wein, W., Navab, N.: Three-dimensional ultrasound mosaicing. In: Proc. MICCAI, pp. 327–335 (2007)
13. Yang, X., Yu, L., Li, S., Wen, H., Luo, D., Bian, C., Qin, J., Ni, D., Heng, P.A.: Towards automated semantic segmentation in prenatal volumetric ultrasound. *IEEE Trans. Med. Imag.* **38**(1), 180–193 (2018)
14. Zimmer, V.A., Gomez, A., Noh, Y., Toussaint, N., Khanal, B., Wright, R., Peralta, L., van Poppel, M., Skelton, E., Matthew, J., Schnabel, J.A.: Multi-view image reconstruction: Application to fetal ultrasound compounding. In: Proc. DATRA/PIPPi@MICCAI, pp. 107–116 (2018)