Deep learning based automatic segmentation of the placenta and uterine cavity on prenatal MR images

James Huang[a,b], **Quyen N. Do**[c], **Maysam Shahed**[a,b], **Yin Xi**[c,d], **Matthew A. Lewis**[c], **Christina L. Herrera**[e], **David Owen**[e], **Catherine Y. Spong**[e], **Ananth J. Madhuranthakam**[c], **Diane M. Twickler**[c,e], **Baowei Fei**[a,b,c,*]

[a]Department of Bioengineering, The University of Texas at Dallas, TX

[b]Center for Imaging and Surgical Innovation, The University of Texas at Dallas, TX

[c]Department of Radiology, The University of Texas Southwestern Medical Center, Dallas, TX

[d]Department of Population and Data Sciences, The University of Texas Southwestern Medical Center, Dallas, TX

[e]Department of Obstetrics and Gynecology, The University of Texas Southwestern Medical Center, Dallas, TX

Abstract

Magnetic resonance imaging (MRI) has potential benefits in understanding fetal and placental complications in pregnancy. An accurate segmentation of the uterine cavity and placenta can help facilitate fast and automated analyses of placenta accreta spectrum and other pregnancy complications. In this study, we trained a deep neural network for fully automatic segmentation of the uterine cavity and placenta from MR images of pregnant women with and without placental abnormalities. The two datasets were axial MRI data of 241 pregnant women, among whom, 101 patients also had sagittal MRI data. Our trained model was able to perform fully automatic 3D segmentation of MR image volumes and achieved an average Dice similarity coefficient (DSC) of 92% for uterine cavity and of 82% for placenta on the sagittal dataset and an average DSC of 87% for uterine cavity and of 82% for placenta on the axial dataset. Use of our automatic segmentation method is the first step in designing an analytics tool for to assess the risk of pregnant women with placenta accreta spectrum.

Keywords

Placenta; uterus; magnetic resonance imaging (MRI); image segmentation; deep learning; neural network; placenta accreta spectrum

INTRODUCTION

During pregnancy, the human placenta serves as the lifeline of the fetus, providing endocrine, metabolic, and immune functions during fetal development [1]. It has a major

*bfei@utdallas.edu, Website: https://fei-lab.org.

role in fetal and maternal health. Proper placental development can have significant impact on future wellbeing [2]. Abnormal placenta growth and development results in adverse pregnancy complications, such as preeclampsia, fetal growth restriction, and stillbirth [3]. A growing placental condition is placenta accreta spectrum (PAS), an abnormal invasion of the placenta into the uterus, which can result in hysterectomy, multisystem organ failure, and death due to the placenta not detaching from the uterus following delivery [4]. The current standard for assessing the placenta is 2D ultrasound, which requires the physician to create 3D images in their mind and suffers from several limitations including approximation of the shape of the placenta and localizing features. An addition of 3D magnetic resonance imaging (MRI) may provide improved soft tissue contrast and wider field of view needed for placental investigation [5]. Studies have shown that MRI can highlight features such as intraplacental bands which correspond to abnormal placentation [6,7].

Various segmentation methods have been developed for the placenta. Shahedi *et al.* proposed a minimally interactive semi-automatic segmentation tool which combined deep learning and region growing to segment the uterine cavity and placenta in MR images [10]. Shahedi *et al.* further improved their network and performed fully automatic segmentation of the uterine cavity and placenta in MR images using a 3D U-Net model [8,9]. Han *et al.* utilized the U-Net for automatic placenta segmentation and modified it to be thinner and less computationally expensive using separable convolution [14]. Looney *et al.* developed a two-pathway, hybrid convolutional neural network model using transfer learning to automatically segment the placenta [15]. Hu *et al.* applied a novel layer into the segmentation U-Net which was weighted by automated acoustic shadow detection to recognize artifacts specific to ultrasound [16]. Alansary *et al.* addressed the issue of motion corrupted prenatal MRI segmentation via a 3D multi-scale convolutional neural network and a conditional random field method [17]. Wang *et al.* developed a semi-automatic placenta segmentation tool called Slice-Seg, which used slice-by-slice propagation to provide a minimally interactive segmentation of the placenta from sparse and motion-corrupted prenatal MRI [11]. Recently, Wang *et al.* proposed another deep learning-based semi-automatic segmentation method called DeepIGeoS, which combined automatic segmentations with user interactions through geodesic distance transforms to further refine segmentation results [18]. Zhang *et al.* designed a multi-region saliency-aware learning method for cross-domain *ex vivo* placenta image segmentation, which guides the translation between domains by enforcing attention and saliency consistency constraints [19]. Most recently, Zimmer *et al.* proposed a multi-task learning approach that combines the classification of placental location and semantic placenta segmentation in a fully convolutional neural network for ultrasound to achieve a high-quality segmentation with reduced image artifacts [20].

This study is a continuation of our previous work in devising a fully automatic segmentation platform for delineating the placenta and uterine cavity in pregnant women using MRI. In previous work, we demonstrated the advantages of end-to-end deep neural networks for placenta and uterus segmentation over semi-automatic methods and computerized algorithms [8–12]. Deep learning brings the advantages of speed of computation time and improved segmentation accuracy [13]. In this study, our objective is to present a fully automatic deep learning segmentation model for fast, accurate, and repeatable 3D segmentation of the placenta and uterine cavity in sagittal and axial T2-weighted MRI

of pregnant women with and without placental complications. We used a 3D U-Net 3+ architecture to implement our end-to-end prediction model and achieved an average Dice similarity coefficient (DSC) of 95.3% for uterine cavity and of 87.7% for placenta on the sagittal dataset and an average DSC of 91.8% for uterine cavity and of 82.7% for placenta on the axial dataset. A clinical obstetrics application of the automatic segmentation method can provide a powerful analytical tool to assess the risk of pregnant women with placenta accreta spectrum.

METHODS

Data

Our dataset consisted of 101 normal women with 7-mm axial placental acquisitions as well as 5-mm sagittal acquisitions and 143 women with suspected PAS, all who had 7-mm axial acquisitions. The MR images were taken in an interleaved fashion. In total there were 244 axial datasets. Each axial image volume contained between 28 and 62 2D axial slices spaced 7 mm apart. The slices were sized at 256×256 pixels except for three cases with smaller sizes. The physical sizes of the image pixels in the axial dataset ranged from 1.051×1.051 mm^2 to 1.953×1.953 mm^2. Each sagittal volume had 27–80 2D sagittal slices scanned with the patient lying on her side, with each slice sized at 256×256 pixels. Corresponding physical sizes of the image pixels ranged from 1.172×1.172 mm^2 to 1.758×1.758 mm^2. For both datasets, manually segmented labels of the uterine cavity and the placenta were provided by an expert radiologist.

Data preparation and preprocessing

To prepare the data, cases in the axial dataset with MR image slice sizes of less than 256×256 pixels were first zero-padded across the volume to be 256×256 pixels. Following this, the data preparation and preprocessing steps for the sagittal and axial datasets were the same. The MR image volumes and labels were cropped along the imaging axis based on the manual segmentations to contain only those slices containing uterine cavity. This step narrowed the minimum and maximum number of MR image volume slices in the sagittal dataset from (26 to 80) to (24 to 61). Similarly, the range of the slices in the axial dataset was reduced from (28 to 62) to (24 to 61). A 2D median filter with a kernel size of 3×3 was applied to each image slice to reduce image noise. Additionally, intensity-level slicing and normalization were applied to the MR image volumes according to the following equation:

$$\hat{I}_i(x, y, z) = \begin{cases} 0 \; if \; I_i(x, y, z) < p_5(I_i) \\ 1 \; if \; I_i(x, y, z) > p_{99}(I_i) \\ \dfrac{I_i(x, y, z) - p_5(I_i)}{p_{99}(I_i) - p_5(I_i)} \; otherwise \end{cases}$$

where $I_i(x, y, z)$ represents the original intensity of a given voxel in an image volume, $\hat{I}_i(x, y, z)$ represents the new intensity of the voxel, and $p_5(I_i)$ and $p_{99}(I_i)$ represent the 5th and 99th intensity percentiles across the image volume, respectively.

Because the placenta was always entirely within the uterine cavity, there was a complete overlap between the manual segmentation label of the placenta and the uterine cavity. To

convert the label volumes into one -hot encoded tensors, the placenta label was subtracted from the uterine cavity label to create a new uterus channel without the intersection. The placenta channel was not modified, and the background channel was generated by inverting the original uterine cavity label. Finally, the channels were stacked along the fourth axis to create image volumes of shape $256 \times 256 \times N \times 3$. To keep the network inputs homogeneous, blocks of five consecutive slices were extracted from the training image volumes, with a separate $256 \times 256 \times 5$ sized contiguous block for each possible position within the volume for a total of N-4 training cases per image. The image blocking procedure is detailed in Figure 1, where the identification for each block extracted from an MR image volume is given as B_i^l, with i representing the number of the initial slice of the block as part of the volume sequence, and l representing the number of the last slice in the block.

Both datasets were divided randomly using a 70-10-20 training, validation, testing split. The 101 sagittal image data were randomly split into groups of 71 training, 10 validation, and 20 testing images. The 241 axial image data gave 169 training, 24 validation, and 48 testing images. The two datasets were used to train and test two different neural networks as described below. When feeding image blocks into the network, we applied a simple data augmentation of left-right flipping across the block at a 50% chance.

Neural network architecture

For the automatic segmentation task, we used the state-of-the-art U-Net 3+ architecture, which makes an improvement on the normal U-Net and U-Net++ by employing full-scale skip connections to increase positional and boundary awareness in segmentation while keeping the model size relatively small [21]. U-Net 3+ adds skip connections from each encoder layer to each decoder layer of equal or deeper depth, as well as intra-skip connections between all decoder layers. These full-scale skip connections are used to integrate low-level details with high-level positional information from feature maps in different scales, which is where the U-Net and U-Net++ fall short by failing to explicitly learn positional and bounding details from organs in segmentation tasks. The U-Net 3+ combines this holistic hierarchical data from the low- and high-level features of the image via max pooling each low-level encoded layer with a larger resolution than the output decoder layer and bilinear up-sampling each high-level encoded layer with a smaller resolution than the output decoder layer [21]. Huang *et al.* also proposed a module called a classification-guided module (CGM), which applies an extra classification task for predicting whether the input image has organ or not to reduce the number of false-positives, but this was not utilized in our work due to the preprocessing steps making the CGM unnecessary.

We hoped the incorporation of the U-Net 3+ would yield improvements in segmentation performance as compared to our previous U-Net model. The original U-Net 3+ proposed by Huang *et al.* is a 2D network with five resolution levels. We modified the U-Net 3+ to accept 3D inputs of size $256 \times 256 \times 5 \times 1$ and output segmentation predictions of size $256 \times 256 \times 5 \times 3$ with the background, placenta, and uterine cavity as the three channels. Additionally, we reduced the number of resolution levels from five to four to greatly lighten the memory load. The convolutional neural network (CNN) architecture is detailed in Figure

2. As an example of data flow in the 3D U-Net 3+, the decoder *D2* as shown in Figure 2 was fed through the concatenation of all four encoder layer outputs. Encoder layer *E1*, having a two times larger resolution than *D2*, was max pooled by a factor of two and passed through 64 filters. Encoder layer *E2* with the same resolution as *D2* was simply filtered and concatenated. Encoder layer outputs *E3* and *E4* were up-sampled using bilinear interpolation at two and four times the resolutions, respectively, before filtering and concatenation. The filters at the decoder input at each encoder scale provide crucial learning parameters for enhancing the organ segmentation performance of the U-Net 3+ model.

The chosen hybrid loss function was given as:

$$L = \alpha L_{pl} + \beta L_{ut}$$

where L_{pl} was the placenta label Dice similarity coefficient (DSC) and L_{ut} was the uterine cavity label DSC. The loss weights were chosen as $\alpha = 0.6$ and $\beta = 0.4$ to give slightly greater weight to the placenta.

Implementation details

The 3D U-Net 3+ architecture was written in Python using TensorFlow and Keras. We trained the network for up to 500 epochs and saved weights based on best validation loss. Our machine was a high-performance computer with four NVIDIA RTX A6000 GPUs, each with 48 GB of memory. Due to the large size of the network, the batch size was initially restricted to eight 3D input blocks on one GPU. We artificially expanded this to 32 blocks per batch by employing the TensorFlow mirrored distribution strategy to duplicate the model onto the four available GPUs. Training was done using the Adam optimizer with an initial learning rate of 1e-4.

Postprocessing and evaluation

Using the 20 testing images from the sagittal dataset and 48 from the axial dataset, each testing image set was fed into the trained network with a single five-slice block at a time with a four-slice overlap with the next block. The blocks were also flipped from left-right and separate predictions were made on the original and mirrored blocks. The mirrored predicted label was flipped again and was averaged with the original predicted label. With a SoftMax output layer, the network output probabilities for each voxel across the three channels summing to 1. All calculated voxel probabilities were added to an empty array with the size of the label volume with block overlap, which was finally collapsed into a ternary output label using TensorFlow's *argmax* function, which collapses a tensor along a specified axis by replacing the values across the axis with the index of the largest value.

Resultant predicted uterine cavity and placenta segmentation labels were further postprocessed in MATLAB using a four-step segmentation smoothing and filling process proposed by [23]. First, a 3D erode operator was applied to the predicted labels with a cubic structuring element of size $3 \times 3 \times 3$ to remove small protrusions from the label. Next, the islands in the label volumes were removed by keeping only the largest 26-connected component of the labels. In the third step, the labels were dilated using a cubic structuring

element of size $3 \times 3 \times 3$. Lastly, the holes, or regions in the segmentations labeled as *0* unreachable from the surface of the volume, were filled in.

We compared the predicted labels to the manually segmented labels using several error metrics: Dice similarity coefficient (DSC), Hausdorff distance (HD), signed volume difference, and percent signed volume difference. The Dice similarity coefficient can be calculated as follows:

$$Dice = \frac{(2 * Area\ of\ Overlap)}{Total\ pixels\ combined}$$

The Hausdorff distance is calculated as the greatest of all the distance from a point in one image to the closest point in the other image [22]. The volume difference and percent volume difference can be calculated as follows:

$$\Delta V = V_{pred} - V_{ref}$$
$$\Delta \% V = \frac{V_{pred} - V_{ref}}{V_{ref}} * 100$$

where V in represents the raw volume difference scaled to the real world using the corresponding physical sizes of the image pixels, V_{pred} represents the total 3D volume of the uterine cavity or placenta predicted label, and V_{ref} represents the total 3D volume of the uterine cavity or placenta ground truth label.

RESULTS

Figure 3 shows the training and validation curves for the sagittal and axial 3D U-Net 3+ models. The best validation Dice score was achieved at epoch 169 for the sagittal model and epoch 157 for the axial model.

The hold-out test data of the sagittal and axial datasets were segmented using the best training models for each dataset. The end-to-end process of forming a prediction label on an entire image volume, including all postprocessing steps, took approximately thirty seconds on average. Figure 4 displays a boxplot for the DSC and percent volume difference on all test data segmentations as compared to the ground truth labels for the sagittal and axial models.

In the sagittal dataset, we achieved a DSC of 0.953 ± 0.014 and 0.877 ± 0.031 for the uterine cavity and placenta, respectively. In the axial dataset, we achieved a DSC of 0.918 ± 0.033 and 0.827 ± 0.051 for the uterine cavity and placenta. Tables 1 and 2 show the full results of testing the best models on the sagittal and axial data using the four selected error metrics. We trained our old 3D U-Net network on the data with the same hyperparameters and postprocessing steps to compare performances. Each evaluation metric is a comparison between the predicted label and manually segmented label from the same MR image.

Sample slices of predicted labels of the placenta and uterine cavity for four cases overlaid onto the original test image slices are shown in Figure 5. Another four example cases of 3D

surface overlaps of predicted placenta and uterine cavity labels with the ground truth labels are shown in Figure 6.

On the sagittal dataset consisting of MR images in women with suspected PAS, we achieved an improvement of 3.1% in Dice score on the placenta (81.5% to 84.6%) and 1% on the uterine cavity (92.1% to 93.1%). The 3D U-Net 3+ model demonstrated superior performance to the 3D U-Net model for all evaluation metrics. The network seemed to predict the uterine cavity label quite accurately but under-segmented the placenta.

On the axial dataset consisting of MR images of both normal and suspected abnormal pregnancies, we achieved an improvement of 1.2% in Dice score on the placenta (80.3% to 81.5%) and 2% on the uterine cavity (86.5% to 88.5%). As with the sagittal data, the network under-segmented the placenta but tended towards over-segmentation of the uterine cavity. This could be attributed to the large variability between patients with placental complications.

In contrast with the sagittal dataset, the network under-segmented the placenta but tended towards over-segmentation of the uterine cavity. Figure 7 depicts an example of a poor segmentation in an axial image case. In Figure 7, the network severely under-segmented both the placenta and the uterine cavity. Because over a third of the axial dataset consisted of patients who eventually underwent a hysterectomy, the underperformance of the axial network compared to the sagittal network could possibly be attributed to the large variability in placenta shape between patient s with placental complications.

The 3D U-Net 3+ model demonstrated superior performance to the 3D U-Net model for all evaluation metrics. On the sagittal dataset consisting of MR images of only normal pregnancies, we achieved an improvement of 4.3% in Dice score on the placenta (83.4% to 87.7%) and 6.7% on the uterine cavity (88.6% to 95.3%). The network segments the uterine cavity and placenta accurately, with high average DSC and low average SD for both. On the axial dataset consisting of MR images of both normal and abnormal pregnancies, we achieved an improvement of 1.7% in Dice score on the placenta (81.0% to 82.7%) and 4.6% on the uterine cavity (87.2% to 91.8%).

DISCISSION AND CONCLUSION

We proposed a new 3D U-Net 3+ architecture which can automatically segment the placenta and uterine cavity on MR image volumes with relatively high accuracy in a rapid way. Performance of automatic segmentation algorithms remains limited heavily by the amount of data, especially in abnormal placenta cases, resulting in inter-patient variability in uterus and placenta shapes. The results are promising for comparative performance to current MRI segmentation methods on the axial slices and brings automated processing of the MRI data to facilitate diagnosis of maternal and fetal health in clinical obstetrics.

Training of the deep neural network was hindered by memory limitations and the amount of data. With the goal of 3D segmentation of the placenta and uterine cavity in mind, the network accepted image blocks as inputs of size $256 \times 256 \times 5$. Allowing for training on larger input blocks than five slices could further improve segmentation performance by

widening the field of view. However, with the size of the 3D U-Net 3+ model and number of parameters, the amount of available computer memory limited the size of the 3D blocks and posed a challenge for training optimally. As with other deep learning-based medical image segmentation tasks, the network would benefit from additional data.

In the future, we plan specific, quantifiable insights on the advantages or disadvantages of sagittal MRI over axial MRI for prenatal imaging, as well as apply additional augmentations, post -processing techniques, and binning methods to improve upon our current segmentation results.

ACKNOWLEDGMENTS

REFERENCES

[1]. Burton G and Jauniaux E, "What is the placenta?," American Journal of Obstetrics and Gynecology, vol. 213, no. 4, pp. S6.e1–S6.e4, October 2015.

[2]. Turco M and Moffett A, "Development of the human placenta," Development, vol. 146, no. 22, p. dev163428, 2019.

[3]. Brosens I et al. , "The "Great Obstretrical Syndromes" are associated with disorders of deep placentation," Am. J. Obstet. Gynecol, vol. 204, no. 3, pp. 193–201, 2011. [PubMed: 21094932]

[4]. Silver R and Branch DW, "Placenta Accreta Spectrum," N Engl J Med, vol. 378, no. 16, p. 1529:1536, 2018. [PubMed: 29669225]

[5]. Khan A, Do Q, Xi Y, Spong C, Happe S, Dashe J, Twickler D, "Inter-reader agreement of multivariable MR evaluation of placenta accreta spectrum (PAS) and association with cesarean hysterectomy," Placenta, vol. 126, pp. 196–201, 2022. [PubMed: 35868245]

[6]. Clark H, Ng T, Khan A, Happe S, Dashe J, Xi Y, Twickler D, "Placenta accreta spectrum: correlation of MRI parameters with pathologic and surgical outcomes of high-risk pregnancies," Am J Roentgenol, vol. 214, no. 6, pp. 1417–1423, 2020. [PubMed: 32208011]

[7]. Do Q, Lewis M, Xi Y, Madhuranthakam A, Happe S, Dashe J, Lenkinski R, Khan A, Twickler D, "MRI of the placenta accreta spectrum (PAS) disorder: raiodmics analysis correlates with surgical and pathological outcome, J Magn Reson Imaging, vol. 51, no. 3, pp. 936–946, 2020. [PubMed: 31397528]

[8]. Shahedi M et al. , "Deep learning-based segmentation of the placenta and uterus on MR images," Journal of Medical Imaging, vol. 8, no. 5, p. 054001, 2021. [PubMed: 34589556]

[9]. Shahedi M et al. , "Automatic segmentation of uterine cavity and placenta on MR Images using deep learning," Medical Imaging 2022: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 12036, pp. 287–293, 2022.

[10]. Shahedi M et al. , "Segmentation of uterus and placenta in MR images using a fully convolutional neural network," Medical Imaging 2020: Computer-Aided Diagnosis, vol. 11314, pp. 411–418, 2020.

[11]. Wang G et al. , "Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views," Medical Image Analysis, vol. 34, no. 1, pp. 137–147, 2016. [PubMed: 27179367]

[12]. Namias R et al. , "Uterus segmentation in dynamic MRI using LBP texture descriptors," SPIE Medical Imaging, vol. 9034, no. 1, p. 90343W, 2014.

[13]. Hesamian M et al. , "Deep learning techniques for medical image segmentation: achievements and challenges," Journal of Digital Imaging, vol. 32, no. 1, pp. 582–596, 2019. [PubMed: 31144149]

[14]. Han M et al. , "Automatic segmentation of human placenta images with U-Net," IEEE Access, vol. 7, pp. 180083–180092, 2019.

[15]. Looney P et al. , "Fully automated 3-D ultrasound segmentation of the placenta, amniotic fluid, and fetus for early pregnancy assessment," IEEE Transactions on Ultrasonics, Ferroeletrics, and Frequency Control, vol. 68, no. 6, pp. 2038–2047, 2021.

[16]. Hu R et al. , "Automated placenta segmentation with a convolutional neural network weighted by acoustic shadow detection," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2019, pp. 6718–6723, 2019.

[17]. Alansary A et al. , "Fast fully automatic segmentation fo the human placenta from motion corrupted MRI," International conference on medical image computing and computer-assisted intervention, vol. 2016, pp. 589–597, 2016.

[18]. Wang G et al. , "DeepIGeoS: A deep interactive geodesic framework for medical image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp. 1559–1572, 2018. [PubMed: 29993532]

[19]. Zhang Z et al. , "Multi-region saliency-aware learning for cross-domain placenta image segmentation," Pattern Recognition Letters, vol. 140, pp. 165–171, 2020. [PubMed: 33324026]

[20]. Zimmer V et al. , "Placenta segmentation in ultrasound imaging: Addressing sources of uncertainty and limited field-of-view," arXiv, vol. 2206, no. 14746, 2022.

[21]. Huang H et al. , "UNet 3+: A full-scale connected UNet for medical image segmentation," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 2020, pp. 1055–1059, 2020.

[22]. Hutteonlocher D, Klanderman G and Rucklidge W, "Comparing images using the Hausdorff distance," IEEE Transactions on pattern analysis and machine intelligence, vol. 15, no. 9, pp. 850–863, 1993.

[23]. Furtado P, "Improving deep segmentation of abdominal organs MRI by post-processing," Biomedinformatics, vol. 1, pp.88–105, 2021.
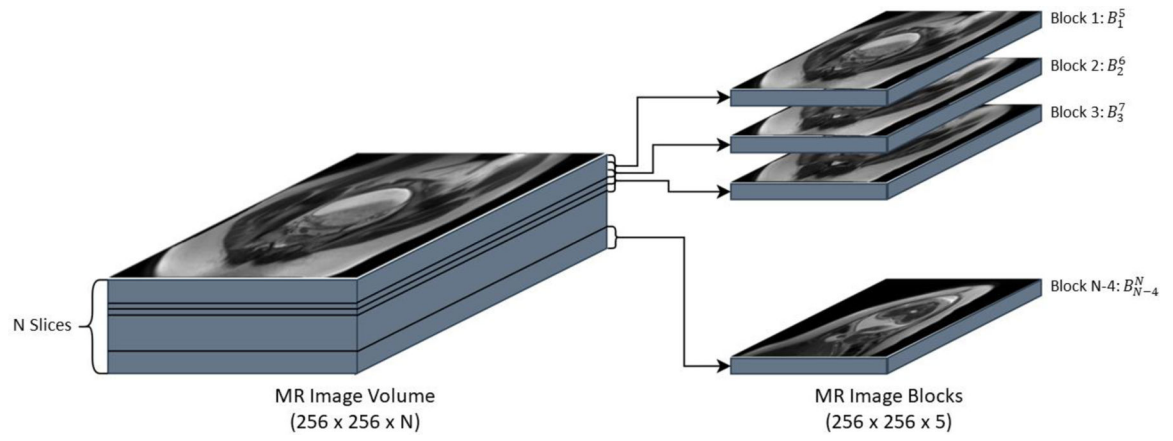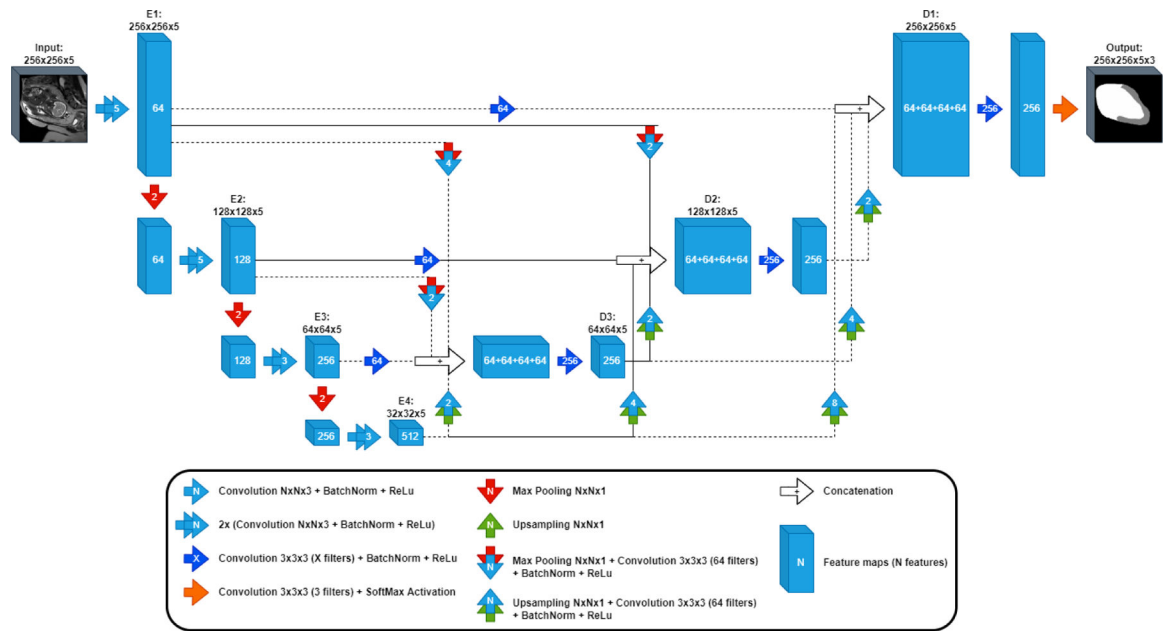
Figure 1.
Process to turn 3D MR image volumes into blocks of five slices each for inputs into the neural network.

Figure 2.
Convolutional neural network architecture (3D U-Net 3+). The network accepts inputs of size 256×256×5 and outputs channels corresponding to background, placenta, and uterine cavity for each voxel for a total output size of 256×256×5×3.

Figure 3.
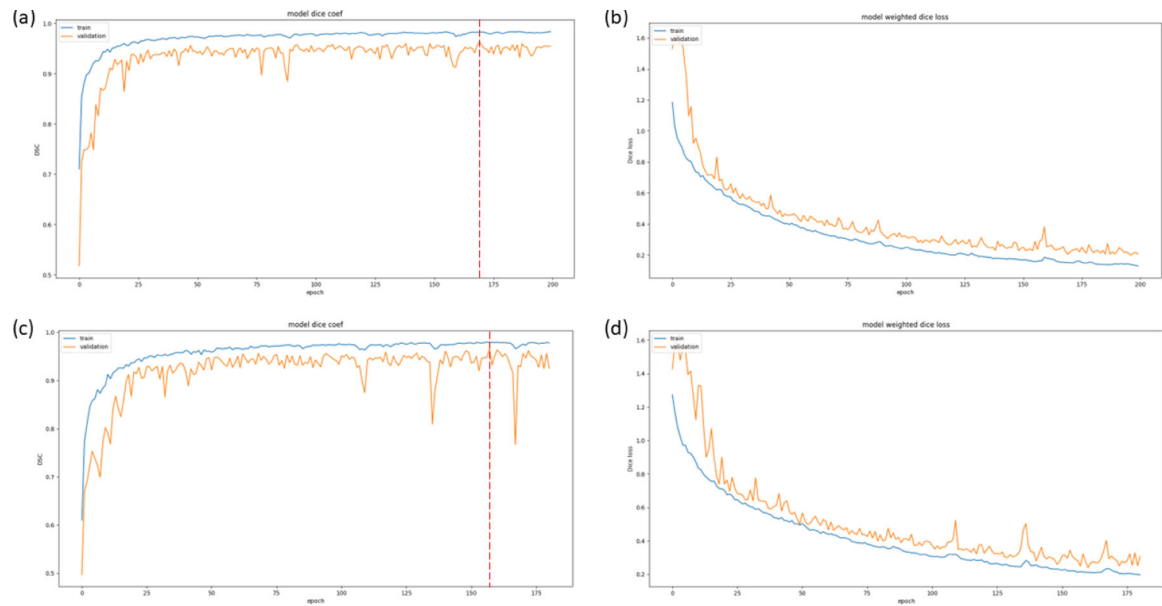Training curves given as a) overall Dice coefficient of the sagittal model, b) weighted Dice loss of sagittal model, c) overall Dice coefficient of the axial model, and d) weighted Dice loss of the axial model. Red dotted lines indicate the epoch at which the best training weights were achieved.
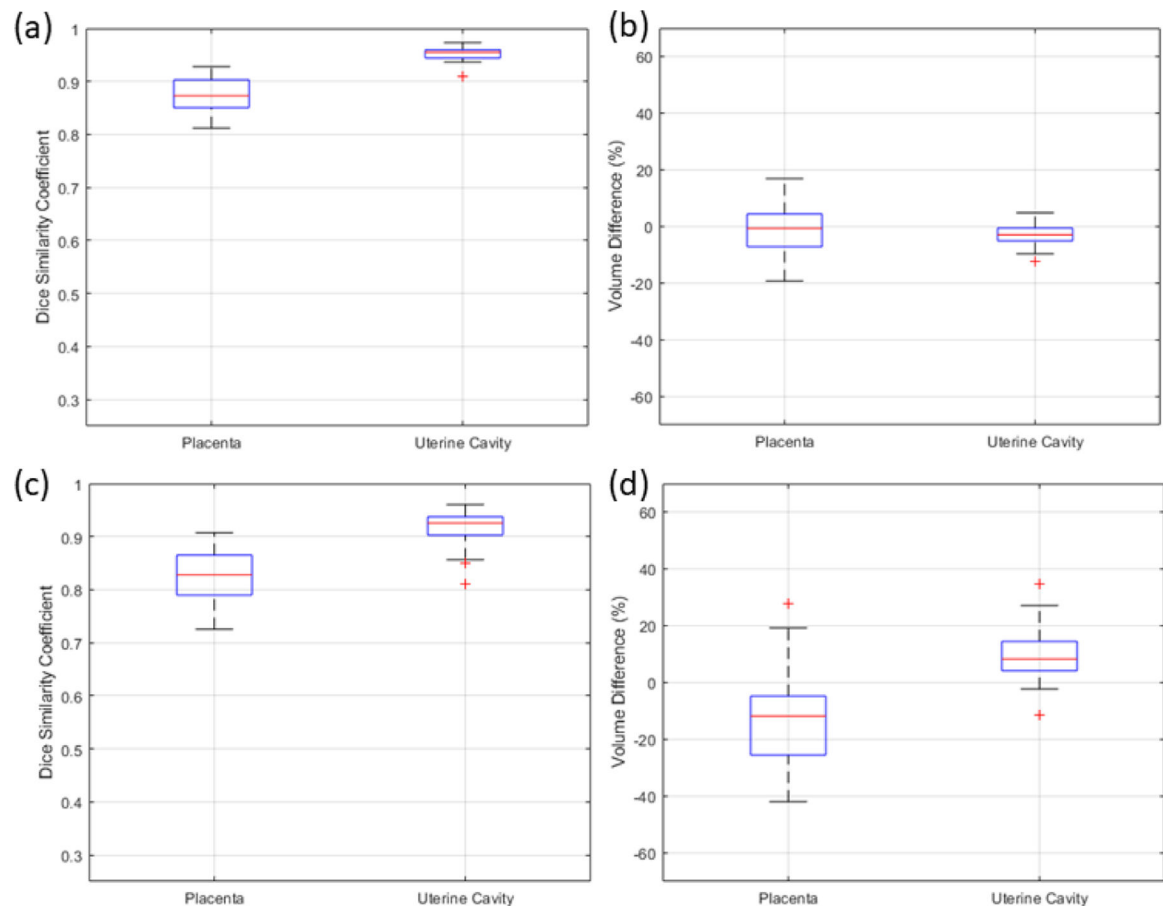
Figure 4.
Boxplots for placenta and uterine cavity segmentation performance represented as a) DSC for the sagittal model, b) percent volume difference for the sagittal model, c) DSC for the axial model, and d) percent volume difference for the axial model. The sagittal model was evaluated on hold-out test data with *n=20* and the axial model was evaluated on hold-out test data with *n=48*.
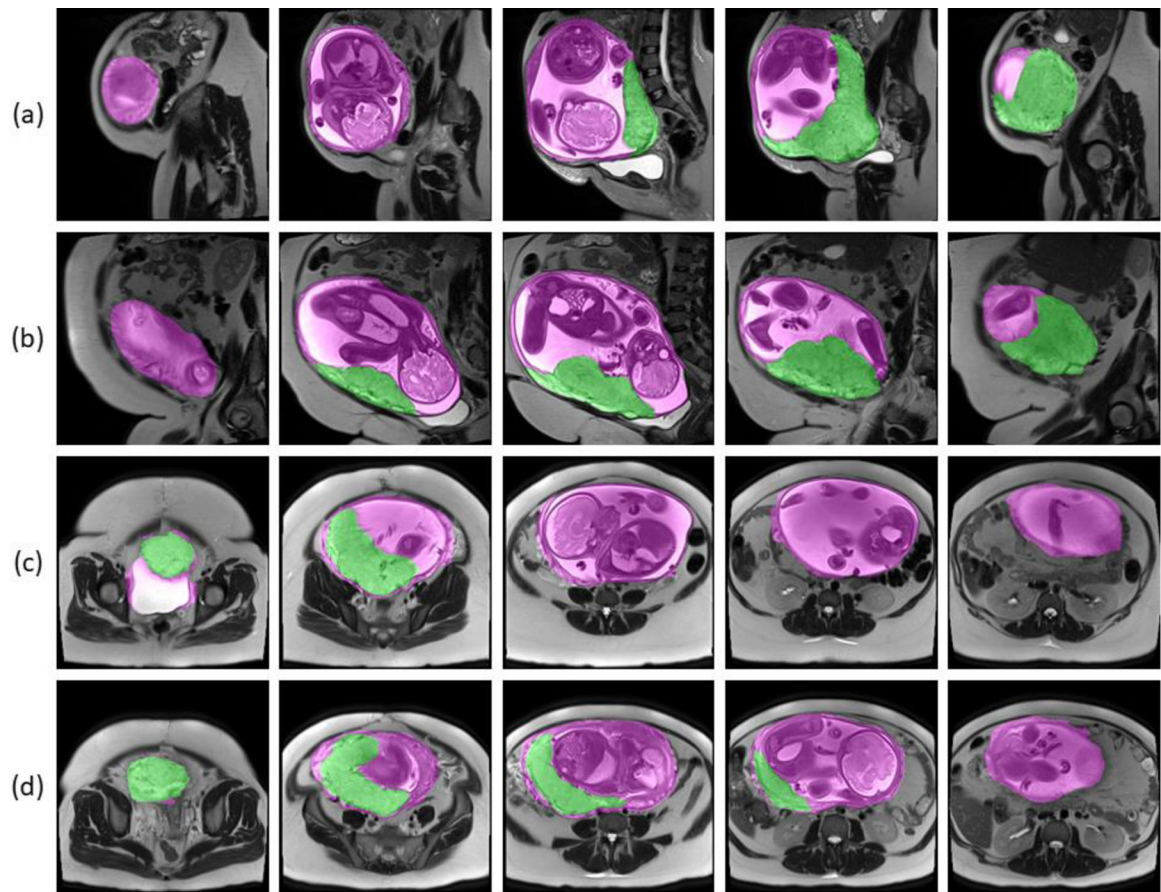
Figure 5.
Predicted labels for four patient's MR images (rows a, b, c, and d) with five selected 2D slices across the volume for each. a) and b) are sagittal images. c) and d) are axial images. Labels are overlaid over the original image slices in magenta and green for uterine cavity and placenta, respectively.
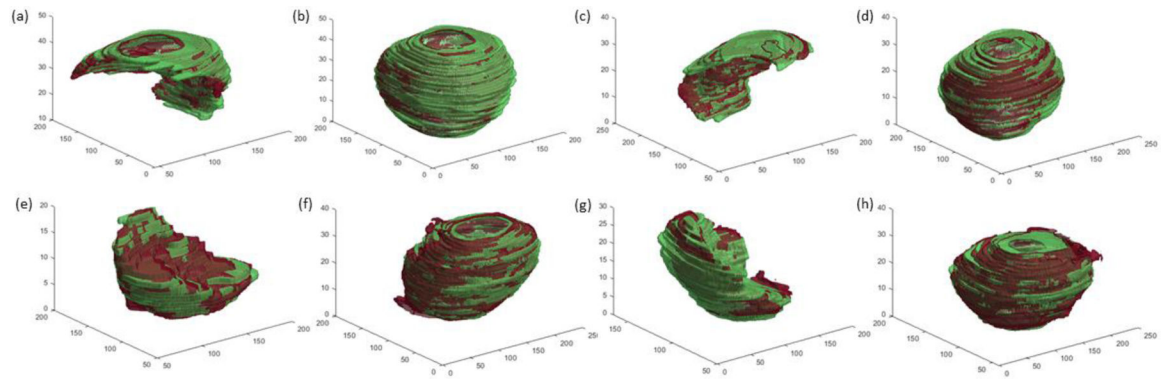
Figure 6.
Superimposed 3D surfaces of predicted placenta (left column) and uterine cavity (right column) segmentation labels and ground truth labels for four test cases, with each row of images from the same patient. a), b), c) and d) are sagittal image cases. e), f), g), and h) are axial image cases. Predicted labels are shown in maroon and ground truths are shown in green.
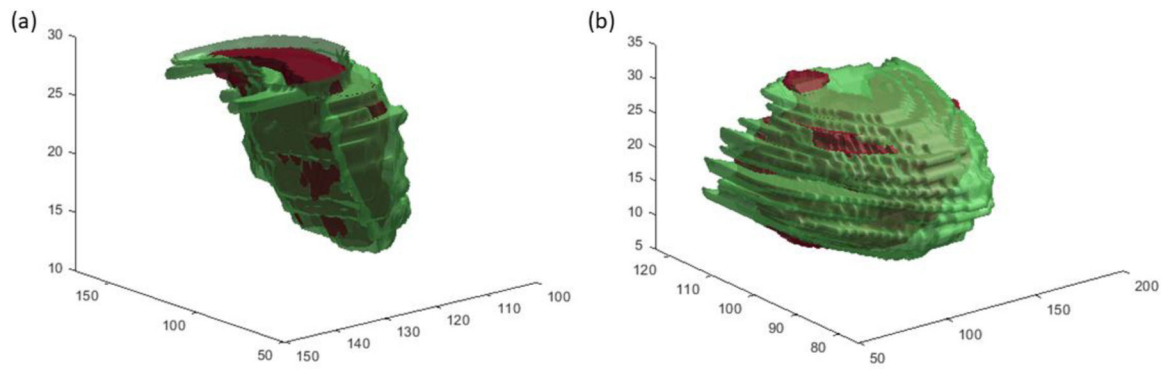
Figure 7.
Poor segmentation performance case in an axial image, shown as superimposed 3D surfaces of predicted and ground truth labels of the a) placenta and b) uterine cavity. Predicted labels are shown in maroon and ground truths are shown in green.

Table 1.

Automatic segmentation performance of the proposed 3D U-Net 3+ network and previous 3D U-Net network on the 20 test images of the sagittal MRI dataset. Error metrics are presented as mean ± std for the placenta and uterine cavity labels.

	Placenta				Uterine cavity			
	DSC (%)	HD (mm)	VD (cm^3)	VD (%)	DSC (%)	HD (mm)	VD (cm^3)	VD (%)
U-Net 3+ (new)	87.7 ±3.1	20.6 ± 10.6	−16.6 ± 109	−1.24 ±9.25	95.3 ± 1.4	13.0 ± 11.4	−146 ±226	−2.91 ±4.50
U-Net (previous)	83.4 ±5.7	24.7 ±21.5	−23.4 ± 151	−1.75 ± 12.8	88.6 ±3.4	25.2 ± 10.1	−172 ± 419	−3.43 ± 8.34

Table 2.

Automatic segmentation performance of the proposed 3D U-Net 3+ network and previous 3D U-Net network on the 48 test images of the axial MRI dataset with *n=16* for number of hysterectomy patients and *n=32* for non-hysterectomy patients. Error metrics are presented as mean ± std for the placenta and uterine cavity labels.

	Placenta				Uterine cavity			
	DSC (%)	HD (mm)	VD (cm³)	VD (%)	DSC (%)	HD (mm)	VD (cm³)	VD (%)
U-Net 3+ (new)	82.7 ±5.1	19.3 ±9.2	−90.5 ± 159	−11.2 ± 17.9	91.8 ±3.3	17.2 ± 14.2	374 ±337	10.3 ±9.3
U-Net (previous)	81.0 ±7.2	25.5 ±20	−84.2 ±207	−10.4 ±23.3	87.2 ±5.6	26.1 ±9.8	−391 ±433	−10.8 ± 12