



Yuzhou Wu 🔍, Cheng Peng, Xuechen Chen, Xin Yao 🔍 and Zhigang Chen *

School of Computer Science and Engineering, Central South University, Changsha 410083, China * Correspondence: czg@csu.edu.cn

Abstract: Cavum septum pellucidum (CSP) is one of the most important physiologic structures that should be detected in Ultrasound (US) scanning for the normal development of the fetal central nervous system. However, manual measurement of CSP is still a difficult and time-consuming task due to the high noise of US images, even for experienced sonographers. Especially considering that maternal mortality remains high in many developing countries, a data-driven system with a medical diagnosis can help sonographers and obstetricians make decisions rapidly and improve their work efficiency. In this study, we propose a novel data-driven system based on deep learning for the diagnosis of CSP called CA-Unet, which consists of a channel attention network to segment the CSP and a post-processing module to measure and diagnose the anomalies of CSP. We collected the US data from three hospitals in China from 2012 to 2018 year to validate the effectiveness of our system. Experiments on a fetal US dataset demonstrated that our proposed system is able to help doctors make decisions and has achieved the highest precision of 79.5% and the largest Dice score of 77.5% in the segmentation of CSP.

Keywords: data-driven system; cavum septum pellucidum (CSP); segmentation; U-net; attention network; deep learning

MSC: 68T07; 68T09; 37N25

1. Introduction

Ultrasound (US) is a widely used imaging method for screening and diagnosis in prenatal diagnosis because it is relatively safe, non-destructive, and low cost. However, identifying US images manually is almost the only way used by mainstream medical institutions. It relies heavily on the experience and state of sonographers. Besides, there are some characteristics of US images which makes the detection more complicated. US acquisition, the quality of cut surface, and manual measurement techniques all lead to observed variability and different diagnostic results.

Although the global maternal mortality ratio is declining year by year, there are still almost 303,000 maternal deaths globally each year [1]. Maternal mortality remains high in many developing countries, such as South Africa, India, and China. This is mainly caused by the contradiction between excessive population and deficient medical resources [2]. In particular, the complexity and operator dependence of US imaging, lagging medical equipment and severe doctor–patient relationship brings about a high misdiagnosis rate of obstetricians in some underdeveloped regions. In China, excellent sonographers are mainly distributed in first-class hospitals in big cities. So, the level of sonographers is uneven and advanced medical devices are scarce in many towns. Many pregnant women need to go to higher-level hospitals and re-diagnose. Moreover, heavy workloads and mental stress have already put a heavy burden on obstetricians. After a long period of work, the accuracy of diagnosis may be affected [3].

In clinical practice, the cavum septum pellucidum (CSP) is one of the most important physiologic structures for the normal development of the fetal head and central nervous



Citation: Wu, Y.; Peng, C.; Chen, X.; Yao, X.; Chen, Z. A Data-Driven System Based on Deep Learning for Diagnosis Fetal Cavum Septum Pellucidum in Ultrasound Images. *Mathematics* 2022, 10, 4612. https:// doi.org/10.3390/math10234612

Academic Editor: Konstantin Kozlov

Received: 17 October 2022 Accepted: 28 November 2022 Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



system in a pregnant routine examination. According to the guidelines of the American Society of US, the CSP has become a standard that must be examined for assessing the development of the fetal central nervous system [4]. The CSP is defined as a liquid chamber located in two transparent compartments at the front of the midline of the brain [5]. When the gestation of a normal fetus is $18 \sim 37$ weeks [5], CSP should be detected in the transthalamic (TT) sagittal plane and TT axial plane during US scanning [6,7]. The standard TT axial plane shows the CSP, with both sides of the thalamus and midline clearly. The TT sagittal plane should detect the integrity of CSP and corpus callosum (CC). Because the clinical significance of CSP is very important, we mainly focus on the measurement of CSP. The pathological representation of CSP is mainly the absence of its normal appearances, such as the absence of the CSP, the presence of an enlarged CSP or narrow CSP. According to the method of Ref. [8], the CSP width (maximum transverse diameter) is measured at the axial position, and the CSP wide diameter normal value is 2 to 10 mm. If the CSP width is <2 mm, our diagnosis is narrow CSP. If the CSP width is >10 mm, our diagnosis is enlarged CSP. If there is no CSP, then the initial diagnosis is absent CSP. These performances may indicate an important marker for many associated brain abnormalities [8]. For example, the absence of CSP on fetal sonography is associated with other intracranial abnormalities, including septo-optic dysplasia, CC dysgenesis, and so on. However, because the feature of US images such as a low signal-to-noise ratio usually leads to fuzzy and discontinuous CSP boundaries, manual measurement of CSP is still a difficult and time-consuming task even for experienced sonographers. Besides, with the development of data-driven systems, many researchers combine it with medical decision analysis and implement segmentation of an entire measurement. However, they have ignored the real clinical application for fully automatic measurement and anomaly diagnosis. Few studies have been conducted on CSP, an important biometric for fetal neural maldevelopment.

On the basis above, a data-driven system can be built through the automatic diagnosis of fetal CSP using deep learning techniques as a basic treatment. Therefore, obstetricians can use the machine annotation of CSP and diagnosis results as a reference to guarantee a more accurate diagnosis and reliable treatment decision. This practice can improve the work efficiency of sonographers and lower the misdiagnosis rate.

In this work, we propose a novel data-driven system for segmenting and measuring fetal CSP. The whole model illustration is reported in Figure 1. Firstly, TT, CSP, midline in TT axial plane and CC, CSP in TT sagittal plane should be segmented. After detecting the integration of physiologic structure, we measure the width of CSP. The new architecture proposed for diagnosis is named CA-Unet based on U-net [9] with changes, which is an end-to-end deep learning architecture. To obtain fine-grained details and context information, VGG-11 is involved in the encoder of U-net [9] as a pre-trained structure. Besides, we bring up a channel attention module to decrease the redundant use of computational resources and model parameters, which can also enhance the robustness to noise of US [10]. Then we can analyze the integrity of the physiologic structure for the TT axial and sagittal plane and the width of CSP to get the first-step diagnosis results. Our major contributions to this study are summarized as follows:

- We propose a medical data-driven system for the segmentation and measurement of CSP using a channel-attention deep network, named CA-Unet. Our CA-Unet is the first combination of the measurement of fetal CSP and deep learning techniques;
- (2) Our system for segmentation is based on improved U-net [9], similar to an encoderdecoder model. To increase the receptive field and improve the efficiency, we initialized the encoder with weights obtained from an ImageNet pre-trained VGG11 [11];
- (3) Moreover, we introduce a new channel attention module, guiding the network to focus on meaningful information at different channels. By sharing the weights of the parallel convolutions, we keep the size of the network unchanged and only add a small number of parameters. This module can be easily added to CNN architectures, increasing prediction accuracy.

(4) Our data-driven system considers the measurement process of clinicians to tackle the automated problem of fetal biometrics measurement. Post-processing is proposed to measure and diagnose the anomalies of CSP. For validation, we have collected fetal US datasets in Xiangya hospital, on which our method achieves a Dice score (DSC) of 77.5% and precision of 79.5%. Our CA-Unet achieves the best segmentation performance among other state-of-art models, Experimental results demonstrate that it could effectively enhance obstetricians' working efficiency and reduce their misdiagnosis rate.

2. Related Works

Our model draws on experience from multiple areas, including medical data-driven systems, deep learning for image segmentation, and the application of attention modules. Next, a detailed introduction will be given to the research related to our works.

2.1. US Data-Driven System

US examination is commonly used for pregnancy diagnosis since US screening is relatively safe, non-destructive, and radiation-free [12,13]. With the development of data-driven segmentation systems, these systems can segment and carry out decision analysis, which is helpful to doctors and other decision makers [14,15]. However, US image segmentation is greatly affected by the quality of the data. There are some properties that complicate the segmentation tasks, including artifact, attenuation, speckle, and missing boundaries [16–18]. With the widespread use of machine learning, the application of diagnosis, computer-aided detection (CAD), and US image segmentation is increasing. Luo et al. [19] utilized robust graph-based (RGB) segmentation and particle swarm optimization (PSO) algorithms to achieve Breast US image segmentation. Baumgartner et al. [20] presented a new framework for automatically detecting 13 fetal standard views on the basis of CNN (convolutional neural network) in 2D US images. The automatic segmentation methods in fetal US imaging usually developed from using segmentation as an intermediate step for measuring physiologic structure. Both Heuvel et al. [21] and Jing et al. [22] proposed an automatic measurement of fetal HC (head circumference) using a random forest and the rapid ellipse fitting method (ElliFit). Some researchers focus on the fetal standard plane recognition; Yu et al. [23] used a deep convolutional neural network architecture to automatically recognize FFSP (fetal facial standard plane), which includes the axial, coronal, and sagittal plane. See also Chen et al. [24] for detecting fetal standard planes by using a novel knowledge-transferred recurrent neural network.

2.2. Deep Learning for Image Segmentation

Fully Convolutional Networks (FCN) [25] is a pioneering work of semantic segmentation, which applied convolutional and upsampling segmented methods and merged information of different scales. There are many model variants based on FCNs for improvement. Firstly, the U-Net [9] utilized the encoder-decoder structure fusing low-level and high-level features to obtain different scale contexts. Then, Vijay et al. [26] proposed the decoder using pooling indices computed in the max-pooling layer to perform nonlinear upsampling. Furthermore, PSPNet [27] developed a pyramid pooling module to embed difficult scenery context features in an FCN-based pixel prediction framework. The most widely used deep learning networks for medical image segmentation is still UNet [9] and improvements based on it [28–30]. Fu et al. [31] presented an architecture called M-Net that solved optic disc and optic cup segmentation joint problems. It mainly consists of the multi-scale input layer, U-shape convolutional network, and a side-output layer. Unet++ [32] utilized a deeply-supervised encoder-decoder network and re-designed skip pathways to decrease the semantic gap between the feature maps of the encoder and decoder sub-networks, which showed better performance in medical image segmentation. Some researchers have proposed a similar idea to us but the detailed method and target are different. Trinh et al. combined spatial and channel attention gate by element-wise

multiplication and added them into skip connection for abnormal tissue segmentation [33]. SCAU-Net introduced spatial attention and channel attention as modules in the encoderdecoder structure for gland segmentation [34]. CA-Net presents channel attention and spatial attention in the decoder for Lesion Segmentation [35]. However, we propose a novel channel attention module at skip connections with an improved U-net structure for CSP segmentation. The automatic segmentation methods in fetal US imaging usually developed from using segmentation as an intermediate step for measuring physiologic structure. Both Heuvel et al. [21] and Jing et al. [22] proposed an automatic measurement of fetal head circumference using a random forest and the rapid ellipse fitting method. Our work is the first combination of the diagnosis of fetal CSP and deep learning method.

2.3. Application of Attention Module

The attention mechanism is widely used in many fields such as machine translation, natural language processing and so on [35]. Attention mechanisms can be divided into soft attention and hard attention. Hard attention is usually used in reinforcement learning, but soft attention is normally used in computer vision. Recently, there have been more applications combining deep learning with computer vision, and most of them focus on using a mask to form an attention mechanism in terms of classification and object detection [36,37]. Spatial Transformer Networks (STN) [38] employed a new module, the Spatial Transformer, which can manipulate the spatial data within the network to maintain the key information. SENet [39] adopted a new strategy of feature re-calibration to acquire the importance of each feature channel automatically, which included two operations, "Squeeze" and "Excitation". CBAM [40] sequentially applied attention maps along two separate dimensions, channel and spatial, improving the ability to capture features of a network model without increasing computations and parameters. Rao et al. [41] extracted information from the image space and the feature space as the input to capture useful information and discard useless features in the feature learning process.

3. Method

Our diagnosis system is based on U-net [9], which is one of the most successful methods of image segmentation, especially for medical image segmentation. The encoderdecoder structure and skip connection used are very classic design methods. There are many new modified convolutional neural network designs based on U-net, such as M-net, U-net++ [32], and so on. However, the upsampling and downsampling process of U-net causes a big loss of information [42]. Besides, the depth of the U-NET network is slightly insufficient to get a big receptive field and context information. It is especially easy to cause misidentification and mis-segmentation for US images due to high noise.

To solve this problem, we proposed CA-Unet. The whole model is shown in Figure 1. Firstly, TT, CSP, midline in TT axial plane and CC, CSP in TT sagittal plane should be segmented. After detecting the integration of physiologic structure, we measure the width of CSP. The main structure is developed from U-net using an encoder and decoder style of network structure. It consists of a Convolutional layer(CONV), rectified linear unit (ReLU), Batch Normalization layer(BN), max-pooling layer, transposed Convolution and Attention module, which can be divided into three major parts: the feature encoder module, the attention module, and the feature decoder module, as shown in Figure 2. There are several important modifications in our architecture to compare with the origin U-net. We can see the detailed design as follows.

3.1. Improved U-net Architecture

The improved U-net includes two main encoding and decoding paths. The encoder path can be seen as a feature extractor. It resembles a VGG11 [11] barring the terminal fully connection layer, which consists of 11 sequential layers, as seen in Figure 2. Each layer of the encoder path is firstly passed through two repeated blocks of 3×3 CONV, BN, and ReLU. A dropout layer (with a probability of 0.1) is included between two blocks. The

original U-Net [9] did not use batch normalization and dropout. BN makes the distribution of input data for each layer in the network relatively stable and accelerates the model learning rate. Dropout can effectively alleviate the occurrence of over-fitting, and to some extent achieve the effect of regularization. The output through the CONV-BN-ReLU is then downsampled and operated through a 2×2 max-pooling with strides of 2. Every time we downsample, the size of the feature map becomes 1/2, and the number of channels in the output feature map is doubled.



Figure 1. Model illustration. The model firstly proposed a new medical image segmentation method by jointly training the U-net (VGG 11) and the channel attention module. After detecting the integration of the physiologic structure, we measured the width of CSP to give the auxiliary diagnosis.



Figure 2. The diagram of the proposed CA-Unet (Channel Attention U-net) model. The solid yellow boxes represent the output of CONV-BN-ReLU block. The hollow yellow boxes represent the copied feature maps. Channel Attention Module filters the propagated features through the skip connections.

The decoding path is similar to the encoding path with one exception: max-pooling is replaced by transpose convolution with 2×2 up-convolution to double the size of the

feature map and half the number of feature channels each time. Simultaneously, skip connections are also implemented between cropped feature maps and decoding paths to ensure that semantic information is added at the top level of the network. Each upsampling contains two 3×3 convolutional layers, one BN, and ReLU, just like the encoding path. Finally, the 1×1 convolution network with a sigmoid activation function is used to determine the probability of each pixel belonging to different classes to achieve segmentation.

3.2. Channel Attention Module

The attention model has many applications in natural language processing and image saliency detection. It aims to find the target area from a large amount of information quickly. The attention mechanism learning method is divided into hard attention and soft attention, and most of them are used in multiple space scales. This paper proposes a different network design of attention mechanism based on the channel scale. We introduce a new architectural unit, which we call Channel Attention Module. Through this module, it can learn global information to enhance meaningful features and suppress meaningless information from a channel perspective. The structure of the Channel Attention Module is illustrated in Figure 3.

The Channel Attention Module process mainly consists of three parts, Global Average Pooling process, attention weighting, and generating context vector function, respectively. Firstly, we use Global average pooling [43] to turn all points in space into a single value. This is mainly because we want to exploit the correlation between the channels. GAP (Global Average Pooling) is able to mask the spatial distribution information to promote the calculation of scale more accurately. The input is a feature map (the dimension is $C \times H \times W$, where C, H, and W represent the number of channels, length , and width of the feature map, respectively), which processes GAP (Global Average Pooling) to get a $C \times 1 \times 1$ matrix , the *l*th layer of channels is calculated as follow:

$$e^{l} = F_{GAP}\left(a^{l}(i,j)\right) = \frac{1}{H \times W} \sum_{j=1}^{w} \sum_{i=1}^{h} a^{l}(i,j)$$
(1)

where (i, j) represents the horizontal and vertical coordinates of point $a \in a$ in feature map with the size of $H \times W$, where l ranges from $\{1, 2, ..., C\}$, which represents the number of feature maps. The so-called activation is to model the degree of correlation between channels.

After squeezing the channels from global information into a group of values, we would like to calculate the score function to find what input information should be focused on. It is also called attention weighting, and aims to model the degree of correlation between channels. Therefore, the activation function we choose should study the nonlinear relationships between the channels. The results we got from Equation (1) can form a vector *e*, which represents $\{e^l, l \in \{1, 2, ..., C\}\}$. As shown in Figure 3.

$$\boldsymbol{s} = \xi \{ W_s[\gamma(W_a \cdot \boldsymbol{e} + \boldsymbol{b}_a)] + \boldsymbol{b}_s \}$$
⁽²⁾

$$\gamma(z) = max(0, z) \tag{3}$$

$$\xi(x) = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
(4)

The dimension of W_a is $\frac{C}{sc} \times C$, and *sc* represents the scaling parameter, which can cut down the number of channels so as to reduce the amount of calculation. Where the weight parameter $W_a \in R^{\frac{C}{sc} \times C}$ and $W_s \in R^{C \times \frac{C}{sc}}$, bias term $b_a \in R^{\frac{C}{sc} \times 1}$ and $b_s \in R^{C \times 1}$.

The γ refers to the RELU function, through which the dimension will not be changed. We then multiply with W_s and add bias term b_s , which is also a fully connected layer as above. We finally pass by a *Tanh* activation function to get the output of *s*. This is the score function of each attention feature map. The attention weighting reflects the importance of channels at location *i*, *j* throughout the network.

Lastly, the operation is the element-wise multiplication of input feature maps and attention weighting:

$$x = s \cdot a \tag{5}$$

We replace the original input *a* with the *x* obtained by the Channel Attention Module, and introduce it into the original U-net network structure for locating the target area.



Figure 3. Schematic diagram of the Channel Attention Module: Input features are scaled with the score function to model the degree of correlation between the channels.

3.3. Post-Processing

After segmenting the fetal biometrics, we propose post-processing to measure the width of CSP and give a diagnosis of the anomalies. This combines the segmentation methods with real clinical applications. Firstly, we diagnose the width of CSP, which is measured at the axial position with a maximum transverse diameter in the TT axial plane. The width of CSP can be calculated by:

$$\operatorname{csp}_{\text{width}} = \frac{h_{\text{width}} \cdot e_{\text{depth}}}{c_{\text{width}}}$$
(6)

where h_{width} represents the height of pixels for CSP, the e_{depth} is the equipment we use to correspond with the real depth of images to the frequency, and c_{width} denotes the width of the images. Moreover, if the width is greater 10 mm, diagnosis is enlarged CSP. If width < 2 mm, it is narrow CSP. Finally, if any biometric such as CC, CSP, and TT cannot be detected, the biometric absent is added to the diagnosis.

4. Experimental Design

4.1. Experimental Data Collection

In this paper, data collection, arranging, and integration come from the department of Obstetrics and Gynecology of Xiangya Hospital, Xiangya Second Hospital, and Xiangya Third Hospital in Hunan Province. Informed consent was obtained from all research subjects. Table 1 demonstrates the HIS, EMR, and LIS systems used by the three hospitals to collect data. Table 1 shows the departmental information, transmitted statistics, and collection time of the three hospitals. As shown in Table 2, large amounts of medical data associated with Obstetrics from 2012 to 2018 are strictly recorded, pre-processed, and classified by different systems in the three hospitals. To ensure the accuracy of our experiment, we extracted pregnant women in the prenatal testing system with gestational age during 18–37 weeks from 2013–2018. In this experimental design, the age for all patients was 22–48 years. All images were collected using the Philips iU22, GE Voluson 730 Doppler US system by two-dimensional convex array probe, with a frequency from 2.5 to 5 MHz.

We extracted 224 patients for analysis, so we selected 224 TT axial planes and 224 sagittal planes of fetal US images for training and testing. Because we should detect the CSP from these two standard sections in clinical medicine, and if the fetal gestation between 18 and 37 weeks cannot be detected TT (trans-thalamic), CSP, midline in TT axial plane or CC and CSP in TT sagittal plane, the fetus should be diagnosed as lacking this structure [44]. The image size was 1136×852 pixels. The biological structure ground truth was drawn and measured by a collaboration between three radiologists with extensive experience in obstetric US images. One radiologist was responsible for manually extracting the area and measuring the width of CSP, and the other two radiologists refined the results.

4.2. Implementation Details

We have experimentally implemented our method on a PC with Intel Core E5-2690v4 CPU, 128 GB of system RAM, and 2X Nvidia RTX2080Ti GPU. The codes are implemented in Python 3.6 and PyTorch 0.4 on Ubuntu 16.04 LTS OS. In our experiments, fetal US image segmentation is realized by our new architecture U-net with channel attention by using VGG11 as an encoder path. The input is an image block of size 1136×852 , which includes the standard TT axial plane and sagittal plane. The output is the segmentation of the thalamus, midline, and CSP in the TT axial plane. The loss function we used is a combination of binary cross-entropy loss and dice loss, which is widely used in medical image segmentation, and can be calculated by:

$$L_{\text{detail}} = L_{\text{bce}} + L_{\text{dice}} \tag{7}$$

$$L_{\text{bce}} = -\frac{1}{n} \sum y * \log(y') + (1 - y) * \log(1 - y')$$
(8)

$$L_{\text{dice}} = \frac{2\sum(y'*y)}{\sum y' + \sum y}$$
(9)

where *y* is the ground truth of each pixel, and *y'* is model prediction. The segmentation of CSP and CC is in the TT sagittal plane to ensure the integrity of these structures and the standard of the selected plane. We then calculate the width of CSP. The CSP measurement method is as follows: The thalamus section shows CSP as the liquid opacity between the first 1/3 of the midline of the brain and the pair of parallel linear echoes in the anterior horn of the lateral ventricle. The shape of it is a rectangle or triangle. According to the method of Ref. [8], the CSP width (maximum transverse diameter) is measured at the axial position, and the CSP wide diameter normal value is 2 to 10 mm. If the CSP width is >10 mm, our diagnosis is enlarged CSP. If the CSP width is <2 mm, our diagnosis is narrow CSP. If there is no CSP, then the initial diagnosis is absent CSP. The equation can be measured by:

$$csp_{width} = \frac{h_{width} \cdot e_{depth}}{c_{width}}$$
(10)

where h_{width} represents the height of pixels for CSP, the e_{depth} is the equipment we use to correspond with the real depth of each photo, which changes with different frequencies, and c_{width} denotes the width of images with 852 pixels. Through this calculation we can give the diagnosis for CSP. The main segmentation method is trained using the Adam optimizer [45] with a learning rate of 0.001, and batch size of 10. Besides, we also use some common data augmentation such as random crops axial flips.

4.3. Evaluation Metric

In order to assess the image segmentation performance, we use four standard metrics: Precision, Recall, Dice coefficient (DC) [46], and Hausdorff distance. Dice coefficient is a function of evaluating similarity and is usually used to calculate the similarity or overlap of two samples. In medical image segmentation, it is the most common evaluation criteria. According to Table 3, they can be defined as follows:

$$Precision_{bio} = \frac{TP}{TP + FP}$$
(11)

$$Recall_{bio} = \frac{TP}{TP + FN} \tag{12}$$

$$DC_{bio} = \frac{2|A \cap B|}{A+B} = \frac{2TP}{2TP+FP+FN}$$
(13)

Hospital System Start Time **Finish Time** HIS 1 January 2014 7 July 2018 Xiangya Hospital EMR 1 December 2013 1 November 2018 HIS 1 September 2012 5 November 2018 EMR 25 September 2013 27 May 2018 EMR document file The Second 1 January 2014 10 May 2018 31 May 2017 Xiangya Hospital LIS 1 January 2015 RIS 17 December 2018 1 February 2016 PACS 18 December 2018 1 January 2015 HIS 1 April 2005 5 December 2018 The Third EMR 1 April 2005 5 December 2018 Xiangya Hospitall EMR document base 1 May 2017 9 December 2018

Table 1. Medical data collection with start and finish time from the three hospitals in China.

Table 2. Medical data classification for obstetrics from the three hospitals in China.

Data Category	Amount		
Medical information	1,933,535 items		
Outpatient service	591,237 people		
Doctors' device in outpatient	24,021,296 items		
Be hospitalized	1,149,184 people		
Diagnosis	1,089,321 items		
Electronic medical records	4,855,618 items		
Doctors' device in clinical	25,757,698 items		
Inspection records	157,426 items		
Medical laboratory records	8,725,584 items		
Routine inspection records	22,358,871 items		
Operation records	218,022 items		

Table 3. Definition of TP, FP, FN, and TN.

Unit (Pixels)	Ground Truth	Not the Physiologic Structure
Predicted the physiologic structure	True Positive (TP)	False Positive (FP)
Predicted Not the physiologic structure	False Negative (FN)	True Negative (TN)

Hausdorff distance [47] evaluates the symmetry distance between two samples. Dice coefficient (DC) is only sensitive to the internal padding of the mask, while Hausdorff distance is sensitive to the segmented boundaries. The definition is as follows:

$$d_H(G,S) = \max\{d_{G,S}, d_{S,G}\}$$

=
$$\max\left\{\max_{g \in G} \min_{s \in S} d(x, y), \max_{s \in S} \min_{g \in G} d(x, y)\right\}$$
(14)

where d(x, y) represents the Euclidean distance between pixels x and y. *G* is the ground truth CSP area, and S is the segmented area for CSP by our system. A smaller value of HD generally indicates better segmentation performance.

4.4. Experimental Results

In our experiment, we divide the Ultrasound (US) images of TT axial plane and TT sagittal plane into a training set and testing set, respectively. According to the method proposed in Section 3, we trained the model using the training set and the corresponding segmented ground truth. In the test set, the automated segmentation results were compared to the ground truth segmentations by radiologists for evaluation. Then, the precision, recall, Dice coefficient and Hausdorff distance were computed on the segmented results.

In Figure 4, we also visualized the segmentation results of the trained model on our dataset. The rows from the top to the bottom (a, c) represent the TT axial plane, (b, d) represent the TT sagittal plane. (a, b) and (c, d) are from different tangent planes of the same pregnant woman, respectively. In the TT axial plane of fetal head, we segmented the structure of TT, midline, and CSP. In the TT sagittal plane, CSP and CC are segmented by our model. From left and right, one can see the input image, ground truth marked by sonographers, segmentation by our model, and the comparison between ground truth and our model. As we can see from Figure 4, Red represents the ground truths, and yellow denotes the predictions. The Orange section is the ground truth part, but we did not recognize it, and the Green section is misidentification. The red section for midline, yellow for TT, blue for CSP, and yellow for CC are the overlap between ground truth and prediction. Visual inspection suggests that our model performed well on this dataset. In general, the performances on TT sagittal plane are better than TT axial plane, which can be caused by the low signal-to-noise ratio of TT axial plane. Next, we compare the performance of our algorithm with previous state-of-the-art methods including U-net, TernausNet [48], and U-net++ [32]. Qualitative comparisons of multi-class physiologic structure segmentation are presented in Table 4 in terms of Precision, Recall, DC (Dice score), and Hausdorff Distance. The best results for each physiologic structure segmentation are highlighted in bold font. As shown in the table, our model attains a precision of 79.5%, outperforming U-net and U-net++ by 9.2% and 5%, respectively. Our network shows better performance than others in terms of Precision, Recall, Dice coefficient, and Hausdorff distance. Through our experiments, it can be seen that U-net with VGG 11 encoder outperforms the U-net architecture. Besides, we found that our proposed Channel Attention module exploits the context information on the channel dimensions, so our proposed model performs better than the U-net baseline by 7.1% in terms of Dice coefficient. We also outline the number of trainable parameters and inference times of each model. By increasing the 1.38 M parameters in the standard U-net, the performance can be improved by 10% in terms of DC. As shown in Table 4, the segmentation results of our CA-Unet network significantly outperforms traditional U-net, TernausNet, and U-net++.

From Figure 5–10, the results of our automatic segmentation model were quantitatively evaluated by applying the evaluation metrics including Precision, Recall, and Dice. The results were reported by histograms with error bars and demonstrated performance on different biological structures in the TT-sagittal plane and TT-axial plane. The columns show the average performance of the segmentation, and the error bars indicate the stability of the models. Specifically, it can be easy to see that our model presents good performance on Precision and DC among all different physiologic structures, but such an improvement is not obvious for TT segmentation. This is because the boundary of TT is not obvious in the US images. Figures 11 and 12 present the ROC curves over the different planes in comparison with different state-of-art methods. For each ROC curve, the area under the curve (AUC) was computed. An average AUC of 80.3% and 81.6% were reached by our model and it showed that our proposed method achieved the highest AUC on these two planes.



(1) input image

Figure 4. From top to bottom, TT axial (a, c) and TT sagittal plane (b, d) of fetal US scan, The ground-truth segmentation (2) is highlighted in red line. Similarly, our model marks the contours of physiologic structure in yellow lines (3). The comparison of ground truth and our model (4).

Table 4. Quantitative comparison of multi-class physiologic structure segmentation on the TT-axial and TT-sagittal of fetal head US images in terms of Dice score (DC), Hausoff distance, Inference Time, and Parameters with different network structures.

Method	U-net [9]	Our CA-Unet	U-net (VGG11) [48]	U-net++ [32]
Precision	0.703 ± 0.116	0.795 ± 0.122	0.743 ± 0.121	0.745 ± 0.113
Recall	0.711 ± 0.204	0.742 ± 0.126	0.721 ± 0.127	0.731 ± 0.118
DC	0.704 ± 0.125	0.775 ± 0.123	0.709 ± 0.116	0.723 ± 0.129
Hausoff Distance	0.672 ± 0.116	0.781 ± 0.124	0.724 ± 0.128	0.701 ± 0.112
Inference Time (s)	0.13	0.18	0.36	0.31
Params (M)	7.76	9.14	13.7	9.04



Figure 5. Precision for physiologic structure segmentation based on the TT-sagittal plane of US images, including the comparison of three other state-of-the-art methods.



Figure 6. Precision for physiologic structure segmentation based on the TT-axial plane of US images, including the comparison of three other state-of-the-art methods.



Figure 7. Recall for physiologic structure segmentation based on the TT-sagittal plane of US images, including the comparison of three other state-of-the-art methods.



Figure 8. Recall for physiologic structure segmentation based on the TT-axial plane of US images, including the comparison of three other state-of-the-art methods.



Figure 9. Dice for physiologic structure segmentation based on the TT-sagittal plane of US images, including the comparison of three other state-of-the-art methods.



Figure 10. Dice for physiologic structure segmentation based on the TT-axial plane of US images, including the comparison of three other state-of-the-art methods.

According to the method of Ref. [8], the CSP width (maximum transverse diameter) is measured at the axial position at last. Table 5 is the evaluation of measuring the width of CSP. We used Precision, Recall, and F1-measure to evaluate the performance of our classification, which includes normal CSP, narrow CSP, and enlarged CSP. The F1-score can be calculated by F1-score = $\frac{P \times R}{2(P+R)}$. As shown in Table 5, the F1-score can achieve at least 77%. Automatic measurement of the width of the CSP can save the mark time of sonographers, which can improve their work efficiency.

The Width of CSP Precision Recall F1-Measure 0.910 Absent 0.667 0.770 Narrow, 0.765 0.929 0.839 width $\leq 2 \text{ mm}$ Normal, 0.738 0.788 0.845 $2 \leq width \leq 10 \text{ mm}$ Enlarged, 0.812 0.798 0.805 width $\geq 10 \text{ mm}$





Figure 11. The ROC curves with AUC scores for physiologic structure segmentation based on the TT-sagittal plane US images, including the comparison of three other state-of-the-art methods.



Figure 12. The ROC curves with AUC scores for physiologic structure segmentation based on the TT-axial plane US images, including the comparison of three other state-of-the-art methods.

5. Discussion

In this study, we propose a novel medical data-driven segmentation system: CA-Unet for segmenting and measuring fetal CSP. This is the first combination of the measurement of fetal CSP and deep learning techniques. More specifically, our model is based on improved U-net, which is the first end-to-end deep learning architecture for segmenting CSP. Secondly, we introduce a new channel attention module to decrease the redundant use of computational resources and guide the network to focus on meaningful information at different channels. Lastly, considering the measurement process of clinicians to tackle the automated problem of fetal biometrics measurement, post-processing is proposed to measure and diagnose the anomalies of CSP. The experimental results demonstrate that the proposed CA-Unet outperforms the state-of-art models and achieves the DSC of 77.5% and precision of 79.5% on fetal US datasets. However, our work only achieves initial success in enhancing obstetricians' working efficiency and giving a diagnosis of CSP. Our strategy, which is based on U-net, is limited by image quality and transferability since the selection of the standard image on the TT axial and TT sagittal plane also relies on the experience of clinicians. So, studying image registration in the selection of specifically US images before our segmentation is also an interesting and meaningful research direction. Besides, our work is mainly based on US images, and we do not consider the differences between the source domain and target domain, which limited us to promote our work to other segmentation systems. In future work, we plan to explore more effective data augmentation methods for limited data. Besides, we plan to consider using transfer learning to solve the transfer and robust problem of our strategy.

Author Contributions: Conceptualization, Y.W. and Z.C.; methodology, Y.W.; validation, Y.W., C.P., and X.C.; writing—original draft preparation, Y.W. and C.P.; writing—review and editing, X.Y.; supervision, Z.C. and X.C.; funding acquisition, Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Intelligent software and hardware system of the medical process assistant and its application, which belongs to "2030 Innovation Megaprojects"—New Generation Artificial Intelligence (Project no. 2020AAA0109605), and Major special project of Changsha science and technology plan (Project no. kh2103016).

Data Availability Statement: All medical data was collected from Xiangya hospital, the second Xiangya hospital, and the third Xiangya hospital of Central South University. The data used to support the findings of this study are currently under embargo while the research findings are commercialized.

Conflicts of Interest: There is no conflict of interest.

References

- Alkema, L.; Chou, D.; Hogan, D.; Zhang, S.; Moller, A.B.; Gemmill, A.; Fat, D.M.; Boerma, T.; Temmerman, M.; Mathers, C.; et al. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. *Lancet* 2016, 387, 462–474. [CrossRef]
- 2. Wu, J.; Tan, Y.; Chen, Z.; Zhao, M. Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country. *Comput. Methods Programs Biomed.* **2018**, *159*, 87. [CrossRef]
- 3. Li, Y.; Zhang, Y.; Fang, S.; Liu, S.; Liu, X.; Li, M.; Liang, H.; Fu, H. Analysis of inequality in maternal and child health outcomes and mortality from 2000 to 2013 in China. *Int. J. Equity Health* **2017**, *16*, 66. [CrossRef]
- 4. Medicine, A.I.O.U.I. AIUM practice guideline for the performance of obstetric ultrasound examinations. J. Ultrasound Med. Off. J. Am. Inst. Ultrasound Med. 2013, 32, 1083.
- Cooper, S.; Katorza, E.; Berkenstadt, M.; Hoffmann, C.; Achiron, R.; Bar-Yosef, O. Prenatal abnormal width of the cavum septum pellucidum—MRI features and neurodevelopmental outcome. J. Matern. Fetal Neonatal Med. Off. J. Eur. Assoc. Perinat. Med. Fed. Asia Ocean. Perinat. Soc. Int. Soc. Perinat. Obstet. 2017, 31, 1–16. [CrossRef]
- Zorila, G.L.; Tudorache, S.; Barbu, E.M.; Comanescu, M.C.; Capitanescu, R.G.; Marinas, M.C.; Florea, M.; Cernea, N.; Iliescu, D.G. Outcome of fetuses with abnormal cavum septi pellucidi: Experience of a tertiary center. J. Clin. Gynecol. Obstet. 2017, 5, 112–116. [CrossRef]
- Ali, H.; Rada, L.; Badshah, N. Image Segmentation for Intensity Inhomogeneity in Presence of High Noise. *IEEE Trans. Image Process.* 2018, 27, 3729–3738. [CrossRef]
- 8. Nagaraj, U.D.; Calvo-Garcia, M.A.; Kline-Fath, B.M. Abnormalities Associated With the Cavum Septi Pellucidi on Fetal MRI: What Radiologists Need to Know. *Am. J. Roentgenol.* **2018**, *210*, 989–997. [CrossRef]
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing & Computer-assisted Intervention, Munich, Germany, 5–9 October 2015.
- Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 October 2018; pp. 286–301.
- 11. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 2014, arXiv:1409.1556.
- 12. Baiying, L.; Ee-Leng, T.; Siping, C.; Liu, Z.; Shengli, L.; Dong, N.; Tianfu, W. Automatic Recognition of Fetal Facial Standard Plane in Ultrasound Image via Fisher Vector. *PLoS ONE* **2015**, *10*, e0121838.
- Santos, C.A.N.; Martins, D.L.N.; Mascarenhas, N.D.A. Ultrasound Image Despeckling Using Stochastic Distance-Based BM3D. IEEE Trans. Image Process. 2017, 26, 2632–2643. [CrossRef] [PubMed]
- 14. Chang, L.; Wu, J.; Moustafa, N.; Bashir, A.K.; Yu, K. AI-driven synthetic biology for non-small cell lung cancer drug effectivenesscost analysis in intelligent assisted medical systems. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 5055–5066. [CrossRef]
- 15. Ling, Z.; Yang, S.; Gou, F.; Dai, Z.; Wu, J. Intelligent Assistant Diagnosis System of Osteosarcoma MRI Image Based on Transformer and Convolution in Developing Countries. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5563–5574. [CrossRef] [PubMed]
- 16. Mozaffari, M.H.; Lee, W.S. 3D Ultrasound image segmentation: A Survey. arXiv 2016, arXiv:1611.09811.
- 17. Zhao, Z.; Zeng, Z.; Xu, K.; Chen, C.; Guan, C. DSAL: Deeply Supervised Active Learning from Strong and Weak Labelers for Biomedical Image Segmentation. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3744–3751. [CrossRef] [PubMed]
- Zhao, Z.; Lu, W.; Zeng, Z.; Xu, K.; Veeravalli, B.; Guan, C. Self-supervised Assisted Active Learning for Skin Lesion Segmentation. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Scotland, UK, 11–15 July 2022.
- Luo, Y.; Han, S.; Huang, Q. A Novel Graph-Based Segmentation Method for Breast Ultrasound Images. In Proceedings of the International Conference on Digital Image Computing: Techniques & Applications, Gold Coast, Australia, 30 November–2 December 2016.
- Baumgartner, C.F.; Kamnitsas, K.; Matthew, J.; Fletcher, T.P.; Smith, S.; Koch, L.M.; Kainz, B.; Rueckert, D. SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound. *IEEE Trans. Med. Imaging* 2017, 36, 2204–2215. [CrossRef] [PubMed]
- 21. Heuvel, T.L.A.V.D.; Bruijn, D.D.; Korte, C.L.D.; Ginneken, B.V. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS ONE* 2018, *13*, e0200412.
- 22. Li, J.; Wang, Y.; Lei, B.; Cheng, J.Z.; Qin, J.; Wang, T.; Li, S.; Ni, D. Automatic Fetal Head Circumference Measurement in Ultrasound using Random Forest and Fast Ellipse Fitting. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 215–223. [CrossRef]
- Yu, Z.; Tan, E.L.; Ni, D.; Qin, J.; Chen, S.; Li, S.; Lei, B.; Wang, T. A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. *IEEE J. Biomed. Health Inform.* 2017, 22, 874–885. [CrossRef]

- 24. Hao, C.; Dou, Q.; Ni, D.; Cheng, J.Z.; Qin, J.; Li, S.; Heng, P.A. Automatic Fetal Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks; Springer: Cham, Switzerland, 2015.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2014, 39, 640–651.
- Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2481–2495. [CrossRef] [PubMed]
- 27. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. arXiv 2017, arXiv:1612.01105.
- Wu, J.; Xiao, P.; Huang, H.; Gou, F.; Zhou, Z.; Dai, Z. An artificial intelligence multiprocessing scheme for the diagnosis of osteosarcoma MRI images. *IEEE J. Biomed. Health Inform.* 2022, 26, 4656–4667. [CrossRef]
- Wu, J.; Guo, Y.; Gou, F.; Dai, Z. A medical assistant segmentation method for MRI images of osteosarcoma based on DecoupleSeg-Net. Int. J. Intell. Syst. 2022, 37, 8436–8461. [CrossRef]
- Wu, J.; Zhou, L.; Gou, F.; Tan, Y. A Residual Fusion Network for Osteosarcoma MRI Image Segmentation in Developing Countries. Comput. Intell. Neurosci. 2022, 2022, 7285600. [CrossRef] [PubMed]
- Fu, H.; Cheng, J.; Xu, Y.; Wong, D.W.K.; Jiang, L.; Cao, X. Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation. *IEEE Trans. Med. Imaging* 2018, 37, 1597–1605. [CrossRef] [PubMed]
- Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
- 33. Khanh, T.L.B.; Dao, D.P.; Ho, N.H.; Yang, H.J.; Baek, E.T.; Lee, G.; Kim, S.H.; Yoo, S.B. Enhancing U-Net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging. *Appl. Sci.* **2020**, *10*, 5729. [CrossRef]
- Zhao, P.; Zhang, J.; Fang, W.; Deng, S. SCAU-net: Spatial-channel attention U-net for gland segmentation. *Front. Bioeng. Biotechnol.* 2020, 8, 670. [CrossRef] [PubMed]
- Gu, R.; Wang, G.; Song, T.; Huang, R.; Aertsen, M.; Deprest, J.; Ourselin, S.; Vercauteren, T.; Zhang, S. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation. *IEEE Trans. Med. Imaging* 2021, 40, 699–711. [CrossRef]
- Yoo, D.; Park, S.; Lee, J.Y.; Paek, A.S.; So Kweon, I. Attentionnet: Aggregating weak directions for accurate object detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2659–2667.
- Liu, F.; Gou, F.; Wu, J. An Attention-Preserving Network-Based Method for Assisted Segmentation of Osteosarcoma MRI Images. Mathematics 2022, 10, 1665. [CrossRef]
- Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 2017–2025.
- Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
- 40. Woo, S.; Park, J.; Lee, J.Y.; So Kweon, I. CBAM: Convolutional Block Attention Module. In Proceedings of the The European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
- 41. Rao, Y.; Lu, J.; Zhou, J. Attention-Aware Deep Reinforcement Learning for Video Face Recognition. In Proceedings of the The IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
- Erdil, E.; Ghani, M.U.; Rada, L.; Argunsah, A.O.; Unay, D.; Tasdizen, T.; Cetin, M. Nonparametric Joint Shape and Feature Priors for Image Segmentation. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* 2017, 26, 5312–5323. [CrossRef] [PubMed]
- 43. Lin, M.; Chen, Q.; Yan, S. Network In Network. arXiv 2014, arXiv:1312.4400.
- Pylypjuk, C.L.; Memon, S.F.; Chodirker, B.N. Utility of Measuring Fetal Cavum Septum Pellucidum (CSP) Width During Routine Obstetrical Ultrasound for Improving Diagnosis of 22q11. 2 Deletion Syndrome: A Case-Control Study. *Appl. Clin. Genet.* 2022, 15, 87. [CrossRef] [PubMed]
- 45. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv 2017, arXiv:1412.6980.
- 46. Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; Davidson, B.R.; Pereira, S.P.; Clarkson, M.J.; Barratt, D.C. Towards Image-Guided Pancreas and Biliary Endoscopy: Automatic Multi-organ Segmentation on Abdominal CT with Dense Dilated Networks. In *Proceedings of the International Conference on Medical Image Computing & Computer-assisted Intervention*; Springer: Cham, Switzerland, 2017.
- Ma, J.; Wang, A.; Feng, L.; Wesarg, S.; Erdt, M. Nonlinear Statistical Shape Modeling for Ankle Bone Segmentation Using a Novel Kernelized Robust PCA. In Proceedings of the International Conference on Medical Image Computing & Computer-assisted Intervention, Quebec City, QC, Canada, 10–14 September 2017.
- Iglovikov, V.; Shvets, A. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv* 2018, arXiv:1801.05746.