Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Towards Automated Prenatal Care: Attention-Based Deep Learning for Fetal Head Circumference Measurement

Seyed Vahab Shojaedini[1*], Mohammad Momenian[2]

1. Biomedical Engineering Department, Iranian Research Organization for Science & Technology, Tehran, Iran.
2. Department of Computer Engineering, Faculty of Engineering, Azad University, E-Campus, Tehran, Iran.

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: May 13, 2025
Accepted: Dec 04, 2025

Keywords:
Fetal Head Circumference
Ultrasonic Image
Deep Learning
Attention Based Semantic Segmentation

ABSTRACT

Introduction: Accurate fetal head circumference (HC) estimation from ultrasound images is critical for prenatal assessment, yet current deep learning approaches are limited by scarce annotated training data and inherently low image contrast. These limitations compromise the model's capacity to reliably delineate fetal head boundaries from surrounding uterine structures, directly impacting clinical utility.

Material and Methods: This study introduces an attention-based deep learning framework designed to optimize feature extraction by selectively emphasizing diagnostically relevant regions within ultrasound images. The attention mechanism guides the network to prioritize fetal head boundaries while suppressing irrelevant background information, thereby enhancing segmentation precision and feature discrimination during training.

Results: Comprehensive evaluation on benchmark ultrasound datasets validates the clinical effectiveness of our approach. The proposed attention-based model achieves a 2% improvement in fetal head detection accuracy compared to current state-of-the-art methods, while simultaneously reducing overfitting probability by 50%. These gains translate to more robust and reliable HC measurements across diverse imaging conditions.

Conclusion: Integration of attention mechanisms into deep neural networks substantially advances automated fetal biometry by addressing two critical challenges: measurement accuracy and model generalization. The demonstrated improvements in both detection performance and overfitting mitigation establish attention-guided learning as a viable pathway toward clinically deployable ultrasound analysis systems, with potential to enhance prenatal care quality and consistency.

►*Please cite this article as:*
Shojaedini SV, Momenian M. Towards Automated Prenatal Care: Attention-Based Deep Learning for Fetal Head Circumference Measurement. Iran J Med Phys 2025; 22 (6): 411-422. 10.22038/ijmp.2026.87180.2530.

Introduction

Ultrasound imaging has emerged as the gold standard for prenatal monitoring due to its real-time visualization capabilities and proven safety profile, unlike ionizing radiation-based modalities such as CT (Computed Tomography) or potential contraindications associated with MRI (Magnetic Resonance Imaging). This non-invasive nature makes ultrasound indispensable in obstetric practice, where repeated examinations throughout pregnancy are essential for monitoring fetal development.

Accurate measurement of fetal Head Circumference (HC) represents a critical clinical parameter with direct implications for prenatal care and outcomes. HC assessment serves as a primary diagnostic indicator for detecting growth abnormalities, including microcephaly, macrocephaly, and intrauterine growth restriction—conditions that require timely intervention to optimize neonatal outcomes [1, 2]. During the first trimester, crown-rump length provides initial gestational age estimation, but as pregnancy progresses, HC becomes increasingly vital for assessing fetal wellbeing and detecting developmental anomalies [3].

The clinical challenge lies in obtaining precise and reproducible HC measurements, as measurement variability directly affects gestational age estimation, detection of growth disorders, and clinical decision-making. Manual HC measurement by obstetricians is time-intensive, subject to inter-observer variability, and requires substantial expertise—factors that contribute to diagnostic inconsistencies, particularly in resource-limited settings. These limitations underscore the urgent need for automated, accurate, and reliable HC estimation methods that can standardize measurements, reduce clinical workload, and improve early detection of fetal abnormalities across diverse healthcare environments.

Fetal development is considered normal if the estimated gestational age falls within a standard deviation of one to ten days. Otherwise, that could be characterized as an unusual case if the dissimilarity is being more than ten days. The expert should have an appropriate approximation from the border around

*Corresponding Author: Tel: +98-2156276311; Email: shojadini@irost.org

the fetal head to obtain cited information. Therefore, estimation of the fetal head from various ranges of obstetricians has been an indispensable part of the traditional analysis (i.e., analysis by the physician) and made this process challenging. In a same fetal head, each obstetrician obtains different results compared to another obstetrician when manual measurement method is used. Furthermore, the need for specialized knowledge, time-consuming nature, and tedious trend are other limitations of manual measurement of fetal head [4]. As an alternative, automated image segmentation may be considered not only as a productive manner to extract the fetal head contour but also it has high potential to play a key role in preparing fetal head details such as meticulous measurement [5]. Unfortunately, some issues such as the low signal-to-noise ratio, acoustic shadows, the low contrast of the ultrasonic image under different conditions, and speckle noise have been inseparable parts of automated segmentation and the major cause to remain this as an open problem [5, 6]. To address the mentioned challenges, researchers have developed various approaches categorized into handcrafted feature-based and deep learning-based techniques. One category attempts to identify the fetal head boundary through tissue classification and optimal ellipse detection [7]. Other approaches [8] employ a combination of Hough transformation, dynamic programming, and ellipse fitting to automatically compute fetal head circumference from 2D ultrasound images throughout all pregnancy trimesters. Additional techniques [9] leverage multidimensional and multidirectional filter banks to capture features representing the anatomical characteristics of the fetal head structure.

Ultimately, a contour around the fetal head using an oval curve fitting method were the main achievement of the above method. Some techniques [10] recommended a distance calculation-based and blend-Altman measurements in order to segment the head, abdomen, and femur of fetal in ultrasound images. The obtained results demonstrated feeble performance of femur segmentation compared to head segmentation due to method's algorithm dependency to the appearance of the femur. AdaBoost learning algorithm has been suggested as a basis for automated method to recognize fetal heads in some other researches [11]. Training the AdaBoost classifier on pseudo-rabid features has played a key role in meaningful achievement in both detection and speed. Furthermore, recognizing the head boundary in the local area with the local phase-based method has been the primary cause of dwindling the sensitivity to spectral noise and intensity fluctuations in ultrasound images. As a reinforcing process, the Huff transform sometimes has been utilized as a post-processing to determine an ellipse on the contour of the head.

The deep learning-based method is characterized by solving various problems. This method, which is also known as end-to-end networks, thanks to its powerful feature extraction and classification stage can extract low-level features and combine them to build high-level features agreeably. In some researches deep learning paradigm have been utilized in the form of using the transfer learning for fetal HC estimation [12].

A deep neural network based on the Link-Net for fetal skull segmentation in 2-D ultrasound images was applied in some researches [13]. This technique tries to segment fetal ultrasonic image as well as estimating oval by minimizing a cost function. The results showed that this scheme has generated considerable diagnosis in terms of consistent with the radiologist's results. Another widely examined approach in deep learning paradigm has been based on multidimensional CNN (Convolutional Neural Network) for the fetal HC measurement. The principal characteristics of this method is based on using fewer parameters, training in faster manner, and receiving acceptable results compared to the costly networks [14].The cascading neural network method to entirely automate the segment of the skull and abdomen of the fetus has been proposed in some other researches [15].

To improve the quality, supervised augmentation and data augmentation were utilized by some researchers to prevent image corruption during training procedures [16, 17, 18]. The methods based on fine tuning are also other techniques in the field of deep learning that have been used to estimate the border of the fetal head in ultrasound images. A suitable example is fine-tuning U-Net network with a lightweight MobileNet as the encoder to perform segmentation on a set of fetal head ultrasound images [19]. Such a combination scheme has been attracting considerable interest due to gaining acceptable results with the smaller trainable parameter sizes. Another proposed scheme in deep learning paradigm is 3D V-Net structure which led to improved segmentation of the fetal skull and therefore the more accurate head circumference measurement [20].

Although early studies have demonstrated the deep learning methods had a meaningful impact on promoting fetal HC segmentation, however, the performance of this class of methods strongly depends on the amount of labeled data. Regarding the estimation of fetal head circumference from ultrasound images, there are also a limited number of valid and labeled images. Therefore, the limited access to valid labeled data sets causes that researchers have insufficient selection to train a deep neural network which leads to overfitting in training procedure of deep neural network.In this article, the framework of deep attention learning is proposed to address the overfitting challenge raised from insufficient training data in estimation of fetal head circumference in ultrasound images. This improvement is performed in the form of aggregation of two techniques known as

channel attention map and spatial attention map. The first technique exploits the inter-channel relationship of features and thus it evokes the concept of "what" in the input ultrasound image. The second concept (i.e., spatial attention map) focuses on “where” concept in ultrasound image by utilizing the inter-spatial relationships of its features. Aggregation of the above attention maps may construct an adaptive feature refinement procedure, which feed a data to the network that the logical connections between its samples are strengthened as well as its redundancy has been reduced significantly. As result training deep neural network by using such refined data may promote its performance in recognizing border of head in fetal ultrasound images, by means of suppressing the overfitting during training procedure.

Attention mechanisms enable deep learning models to selectively focus on diagnostically relevant image regions while suppressing irrelevant background information, mimicking the human visual perception process. This selective emphasis improves feature discrimination, enhances model interpretability by highlighting decision-critical areas, and significantly reduces overfitting—particularly valuable in medical imaging where annotated training data is scarce and image quality is inherently variable. The remaining sections of this manuscript are structured as follows: Section 2 presents the deep attention-based approach we propose. In Section 3, the results that have been obtained from applying the proposed method are illustrated. In Section 4, the fetal head circumferences which have been estimated by using proposed method are compared to the results obtained from the basic structure by using their effective indexes. In the last section of the article, conclusion is presented.

Materials and Methods

One of the most suitable types of deep neural networks is UNet, which has demonstrated its power in medical image segmentation as well as several types of segmentation in special organs of the body. The potential of this method is owing to its up and down linear and non-linear convolutional filters which have high aptitude to locate the attributes of input images. This means that this technique may scrutinize a variety of features of input images [21]. The U-shaped deep neural network consists of four encoders and four decoder blocks each block involves the spatial dimensions and the feature channels. The encoder sections fulfil feature extracting and learning a conciseness portrayal of the input image [22]. To decrease the computational costs as well as reducing the number of trainable parameters, the output of a sigmoid activation function attaches to a 2×2 max-pooling layer, where the spatial dimension including the width and height of the feature maps dwindled by half. The encoders connect to the decoders with the bridge layer, which involves 2 convolutions and a ReLU (Rectified Linear Unit) activation function. Eventually, the segmentation mask is produced by the decoders layer, when the results transfer from previous layers. The UNet architecture and other details may be observed in Figure 1.

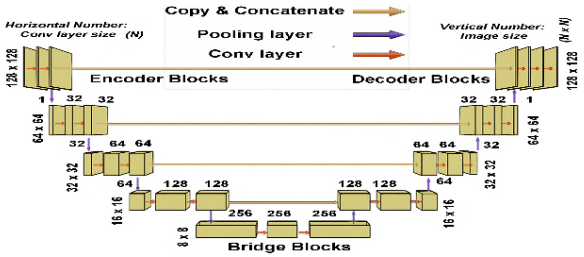

Figure 1. Basic method architecture (UNet).

Although the UNet is an appropriate choice to segment the medical images, some issues such as challenges in the training process and generating final optimized results prompt researchers to suppress this network with their proposed methods. The Channel Spatial Attention Convolution (CSAC) structure, which inspires the attention mechanism rules in the human brain, is the proposed idea of this article to address the above challenges. This component incorporates into a UNet architecture and ensures the restoration of statistical elements and clinical parameters of fetal head segmentation. The Attention strategy serves a crucial function in human perception. The key feature of human visual processing is that individuals do not try to analyze an entire scene simultaneously. Rather, humans employ a series of selective observations and concentrate specifically on significant regions to achieve superior comprehension of visual information. The significance of the Attention mechanism has been investigated in some studies [23]. In deep neural networks the Attention module is not an independent network (for example such as the UNet network), but it integrates to a deep neural network in order to promote the performance. In fact, the UNet mechanism extracts general input features, whereas UNet+CSAC emphasizes (i.e., refines) the meaningful and important parts of input features. Obtaining the above two types of attention is performed in practice by two modules. The Channel Attention Module and Spatial Attention Module are two sequential sub-modules that shape the basis of a CSAC and guarantee to refine the input features [24, 25]. The basic concept of a CSAC module containing its sub-modules may be observed in Figure 2.

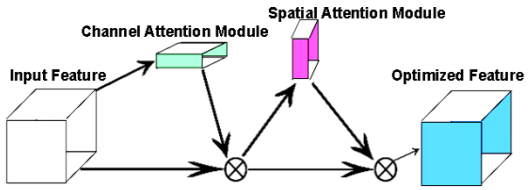

Figure 2. Structure of a CSAC module.

The intermediate feature map $M \in \mathbb{R}^{C\times H\times W}$ enters to a Channel Attention sub-module then the result convert to 1D Channel Attention map $A_{channel} \in \mathbb{R}^{C\times 1\times 1}$ and passes to a 2D spatial attention map $A_{spatial} \in \mathbb{R}^{1\times H\times W}$. Eventually this process may be summarized as Equation (1):

$$M' = A_{channel}(M) \otimes M, M'' = A_{spatial}(M') \otimes M' \quad (1)$$

where $\otimes$ indicates element-wise multiplication. The multiplication results transmit through the process appropriately. Eventually, M'' is the optimized and refined output features. Focusing on 'what are significant parts in an input image' is the foundation of channel attention module. Before arriving input images into the channel attention module, image squeezing and max-pooling were used to increase the performance of channel attention. To Construct channel attention map $A_{channel} \in \mathbb{R}^{C\times 1\times 1}$, the intermediate feature map $M \in \mathbb{R}^{C\times H\times W}$ forwards to a shared network. One hidden layer and a multi-layer perceptron is the essential parts of this shared network. Finally, the output feature vectors of shared network are obtained by using element-wise summation merge together. Briefly, the calculation of channel attention may be observed as Equation (2:

$$A_{channel}(M) = \sigma\big(MLP \times MaxPool(M)\big) = W_i(W_r \times \big(M_{maxpool}^{channel}\big)) \quad (2)$$

Where σ indicates sigmoid function and $W_i, W_r \in \mathbb{R}^{C}$ are the MLP's weights. Furthermore W_i, W_r are the weights of input features and ReLU activation function respectively. Process steps of channel attention sub-module may be demonstrated in Figure 3.

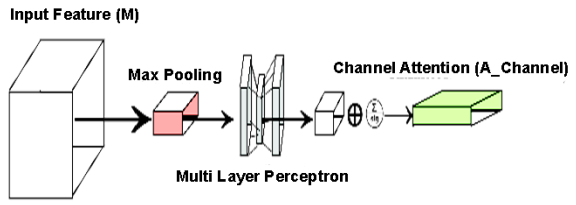

Figure 3. Structure of a Channel Attention Module.

On the contrary, the spatial attention concentrates on 'where are meaningful parts which resulted from the channel attention module'. In Similar manner with channel attention module, the spatial attention utilizes max-pooling in order to increase the quality of features extraction. Next, to generate a spatial attention map $A_{spatial}(M) \in \mathbb{R}^{H\times W}$, the module must have a convolution layer in order to suppress the obtained features and generate a 2D map $M_{maxpool}^{spatial} \in \mathbb{R}^{1\times H\times W}$. The calculation of spatial attention may be demonstrated as Equation (3):

$$A_{spatial}(M) = \sigma\big(f^{3\times 3} \times MaxPool(M)\big) = \sigma\Big(f^{3\times 3} \times \big(M_{maxpool}^{spatial}\big)\Big) \quad (3)$$

In which σ indicates sigmoid function and $f^{3\times 3}$ demonstrates the filter size of convolution layer. Process steps of spatial attention sub-module may be demonstrated in Figure 4.

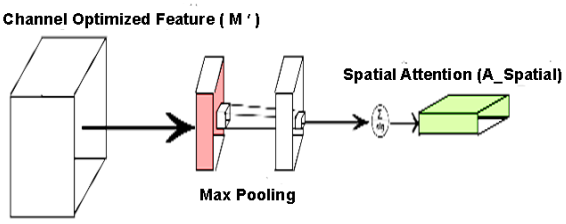

Figure 4. Structure of a Saptial Attention Module.

To implement the proposed idea, the network should apply the CSAC modules contains channel attention and spatial attention sub-modules into the convolution layer blocks of the UNet (i.e., basic structure). For this scenario both channel attention and spatial attention placed in a sequential manner. Implementation of a CSAC module into a convolution layer block of the UNet may be observed in Figure 5.

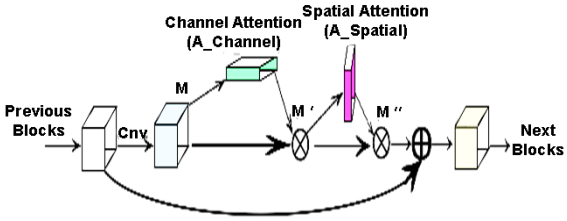

Figure 5. A Convolution Block of CSAC module.

As described in following sentences, after apply CSAC into the convolution blocks of the UNet network, some decisions have been made. Compared to a pure UNet network, the size of filters has transformed to 128, 256, 512, and 1024 respectively, while the filter size of the UNet network (basic method) is 32, 64, 128, and 256. Using smaller filter sizes has reduced the computational complexity rate, while larger filters given the chance to the network to seek more precious patterns during the learning process. In the first step of UNet+CSAC architecture, an ultrasound image with $128 \times 128 \times 1$ size lays on the first encoder block which has multi convolution filters in 32, and 64 sizes respectively. Other encoder blocks have the approach like the first encoder block including its filter shape, connectivity, and ReLU activation function. The generated data then is transmitted to the next encoder block by a 2×2 max-pooling, which reduces the size of a processed image. In general, the main features of the input images are extracted by encoder blocks. As the left side of Figure 6 represents, (i.e., based on encoder blocks), the size of convolutional filters is increased while the size of input images is reduced by max-pooling layers. Some examples of input image

transformations during the encoding process may be observed in Figure 6.

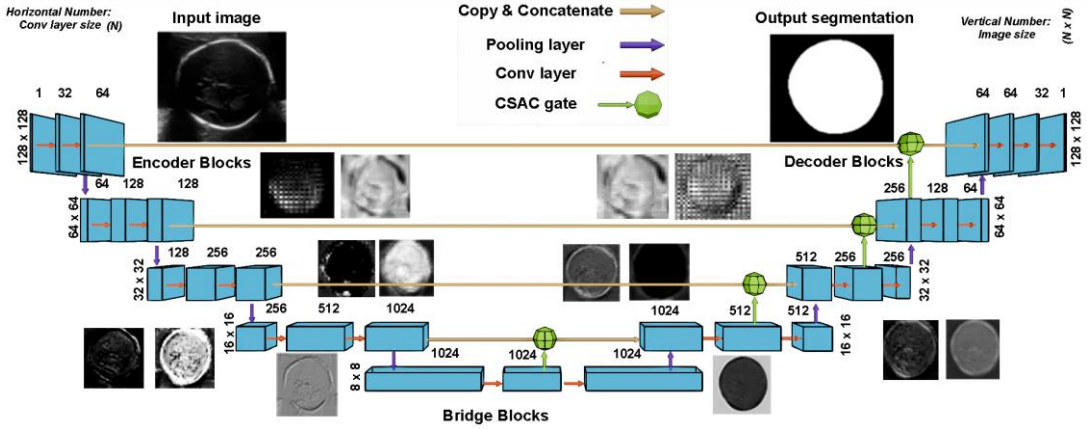

Figure 6. Proposed method architecture (UNet+CSAC).

To create a connection between the encoder and the decoder blocks, this architecture used a bridge layer, which has similar behavior to other layers. Also, the first gate of CSAC may be seen in the bridge layer. The right side of the Figure 6 is relevant to decoder blocks, which profit from the encoder blocks with copy and concatenation lines. To achieve the appropriate HC segmentation, the aim of decoder blocks is to extract the details of an input image with a focus on targeted segmentation (i.e., HC area). The mechanism of decoder blocks as well as encoder blocks involves pooling layers, convolutional filters, and activation functions with three CSAC gates, which are the strength of the proposed method. With a concentration on the meaningful portion of an input image, which is produced by each level output, it guarantees to get suppressed results in comparison with a UNet network. The final architecture of the UNet+CSAC is demonstrated in Figure 6.

Results

In order to evaluate the performance of the proposed method it was applied on real images belonging to dataset of HC18 challenge [32]. This dataset contains 999 ultrasound images of size 800 by 540 in each of them the contour and value of the circumference of fetal head are annotated. Figure 7 shows two examples of images used in this article and their corresponding annotations. The aforementioned dataset was divided into 60%, 20% and 20% portions for training, validation and testing of the neural network, respectively. Furthermore the 5-fold cross validation strategy was adopted for investigating the quality of training the network and monitoring overfitting phenomena. The testbed was prepared on python3 and TensorFlow paradigms on an Intel ® Core i7-10700 computer with Ubuntu 22.04 operation system, 32 GB RAM, and an NVIDIA 2080 Ti. This program runs on Jupyter Notebook, an open-source software for interactive computing.

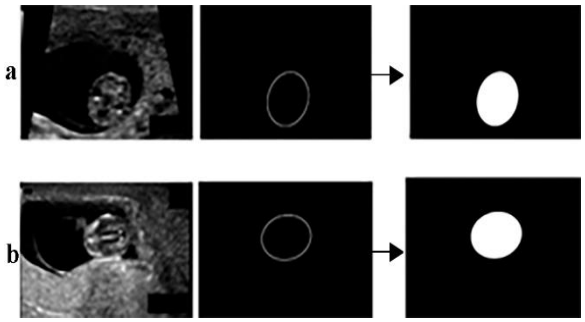

Figure 7. Some examples of image preprocessing before training the network with fill free space of labeled HC in order to handle the unbalanced distribution of HC space.

To evaluate the performance of proposed method, two scenarios were adopted to conduct the tests. In the first scenario a UNet neural network, which has been used in various researches to detect and measure HC automatically, was applied on dataset. In the second scenario the above UNet structure was optimized by adding the attention-based structure proposed in this article (i.e., CSAC) and the previous tests were repeated. Hence, any dissimilarity in the results of these two methods can be rationally assigned to the use or non-use of the proposed method. Table 1 explains the eventual architecture of these two applied models as well as the setting of their hyperparameters which led to the results of this article. In order to quantify the comparison between the results obtained by each of the examined methods with ground truth, Absolute Difference measure (i.e., ADF) was calculated as the absolute difference between predicted HC and real HC as a positive amount in millimetres.

$$ADF: |HC(P) - HC(GT)| \quad (4)$$

In above equation, HC(GT) and HC(P) demonstrate Ground-Truth and predicted fetal head circumferences in millimetres. Other comparison criteria are introduced and used in the discussion section.

Table 1. Structure of basic and proposed method.

Network	Basic Method	Proposed Method
Structure	4 convolutions, 4 max pooling, 4 enocoders and decoders	4 convolutions, 4 max pooling, 4 enocoders and decoders
Encoders, Decoders	4 encoder, 4 decoder	4 encoder, 4 decoder
Number of trainable params	30,235,922	30,522,784
Optimizer	Adam	Adam
Batch size	32	32
Activation functions	ReLU	ReLU
Loss function	Binary Cross entropy	Binary Cross entropy
Number of folds	5	5
Epochs per fold, L rate	20, 1e-4	20, 1e-4

Figure 8 represents some of results which have been obtained from the abovementioned scenarios. There are three images in each section. The left image demonstrates the original test subject, with the actual circumference of the fetus's head has been written numerically in millimeters, below it. The middle image shows the contour estimated for the fetal head circumference by the proposed method. Below this image, the head circumference resulting from the above estimate is written numerically in millimeter units, and its error value compared to the ground truth head circumference is written in parentheses based on ADF criterion. Finally, the right image represents the estimated contour belonging to fetal head circumference by using basic method (i.e., UNet). Similar to previous case, the head circumference resulting from the above estimate is written numerically in millimeter units bellow it, and its error value compared to the ground truth head circumference is written in parentheses based on ADF criterion. It should be mentioned that in both center and right-hand images, the blue region indicates the actual HC region whereas the yellow contour represents the predicted boundary. Most results from both experimental scenarios correspond to instances where the baseline and proposed approaches successfully estimated the boundary and thus the fetal head circumference measurement with satisfactory precision. More specifically, rows (A) through (I) in Figure 9 display several examples of this condition. Through visual examination of these samples, one can observe that the boundary generated by either baseline or optimized networks (yellow trajectory) aligns well with the ground truth region (blue). Simultaneously, examining the circumference values and their corresponding ADFs shown beneath these images confirms that both approaches demonstrate errors below 5 millimeters relative to the actual fetal head circumferences. The subsequent category of results pertains to images where the baseline approach has severely miscalculated the boundary and circumference size around the head, whereas the proposed approach has successfully estimated these measurements with satisfactory precision. More specifically, rows (J)-(O) in Figure (8) present several examples of this condition. In contrast to the earlier scenario, visual examination of outcomes in this image group reveals substantial discrepancy between the blue region (ground truth) and the estimated boundary (yellow trajectory) when the baseline (i.e., UNet) method was utilized. The samples demonstrate that in such instances, the baseline approach typically fails to capture a considerable portion of the fetal head region. In accordance with these results, the amount of the circumferences measured in these cases using the basic method have the ADF error of more than 200 millimeters. Unfavorable results obtained from the basic method for this category of images (J-O) have been obtained in conditions where it can be seen that for all of them, the proposed method of this article has been able to obtain a contour that is completely consistent with the ground truth in parallel with acceptable ADF error in calculated circumference value (e.g., less than 5 millimeters). What is remarkable in images (J)-(O), the boundary separating the fetal head from the background appears more indistinct compared to images (A)-(I). This phenomenon appears to contribute to incorrect outcomes when applying the baseline method. However, findings demonstrate that our proposed approach has successfully compensated for this reduction in edge definition. Beyond the two aforementioned scenarios, there existed a small subset of images where neither the baseline nor proposed methods could perform accurate analysis. Several instances of this subset are presented in rows (P)-(R) of Figure 8. What differentiates this subset from the preceding two is the extremely poor contrast between the fetal head region and surrounding background, making it challenging to identify the separation between these regions even for specialists. This contrasts with the clearly visible boundaries in images (A)-(I) and the partially visible boundaries in images (J)-(O). Consequently, in the two earlier subsets, either both methods or minimally the proposed approach operated correctly, whereas in this latest subset, both have been unsuccessful. Following a general assessment of both proposed and alternative approaches' performance obtained through visual examination of selected results in the preceding section, we now evaluate both methods across all images to perform more comprehensive analysis. Additionally, beyond ADF which was presented and employed in the prior section, four additional evaluation metrics are utilized including Jaccard, Binary Accuracy, Dice Similarity Coefficient (i.e., DSC), and Hausdorff (i.e., HD). The Jaccard serves as a metric for quantifying the similarity and difference of sample

collections. It quantifies similarity between finite sample collections (for instance A, B) as shown in Equation (5).

$$\text{Jaccard (A, B)}: \frac{A \cap B}{A \cup B} \quad (5)$$

The Binary Accuracy computes the proportion of predicted outcomes that correspond with ground truth outcomes for binary classifications. Given that the classification is binary like our labeled images, it comprises the probability value of predictions equaling 1 and can be formulated as Equation (6).

$$\text{Binary Accuracy}: \frac{(\text{TP} + \text{TN})}{(\text{TP}+\text{FN}+\text{FP}+\text{TN})} \quad (6)$$

In our work, TP denotes the number of pixels in a labeled image belonging to an actual HC region that the baseline or proposed approach correctly identifies as HC region. Similarly, TN signifies the number of pixels in a labeled image belonging to an actual non-HC region that the evaluated approaches correctly classify as non-HC region. Conversely, FP indicates the quantity of pixels in a labeled image belonging to non-HC region which the evaluated approaches categorize as HC region. In contrast, FN represents the pixels of HC regions that both approaches failed to identify. According to these definitions, the comparison metric DSC quantifies the spatial agreement between two segmentation regions (i.e., Ground Truth and Prediction) as Equation (7).

$$\text{DSC}: \frac{2 \times \text{Area(GT)} \cap \text{Area(P)}}{\text{Area(GT)} + \text{Area(P)}} \quad (7)$$

Finally, Hausdorff parameter which is illustrated in Equation (8), calculates how far two subsets of a Ground-Truth and Predicted fetal head are from each other.

$$\begin{aligned} H_{P,GT} \max(h_{p,gt}, h_{gt,p}), h_{p,gt} \\ = \max_{p \in P}(\min_{gt \in G}(d_{p,gt})), h_{gt,p} \\ = \max_{gt \in GT}(\min_{p \in P}(d_{gt,p})) \end{aligned} \quad (8)$$

where P: $\{p_1, p_2, p_3, \ldots, p_n\}$ represents the pixels in the predicted fetal head, and GT: $\{g_1, g_2, g_3, \ldots, g_n\}$ represents the pixels belong to Ground-Truth. The $d_{p,gt}$ is the Euclidean distance between p and gt and vice versa. This implies that $h_{p,gt}$ initially identifies the closest point in GT for each point in P, followed by selecting the maximum of these distances, representing the most discordant point of P. Then, the Hausdorff distance $H_{P,GT}$ equals the maximum of $h_{p,gt}$ and $h_{gt,p}$ To ensure dependable outcomes, the training and testing protocols were executed 200 times for both baseline and proposed approaches. In each iteration, all five metrics were computed for the two evaluated approaches and displayed as box and whisker plots presented in Figure 8. Through this visualization approach, one can assess the potential advantages and disadvantages of our proposed approach relative to the baseline approach from multiple perspectives based on statistically robust repetitions. These visualizations demonstrate that incorporating the innovation presented in our work into the baseline UNet architecture has yielded substantial enhancements across all five evaluation metrics. One can note that following numerous reproducibility experiments on the proposed approach, the mean values for the five metrics of Jaccard, Binary accuracy, DSC, ADF, and HD equaled 97.4%, 98.5%, 98.3%, 0.67 mm, and 0.54 mm respectively. The corresponding values for the baseline approach averaged 94.3%, 96.2%, 95.1%, 2.5 mm, & 1.6 mm. These outcomes reveal that employing the proposed approach has produced a 3.1% improvement in HC calculation accuracy using Jaccard, alongside a 3.2% increase in DSC compared to the alternative approach. Furthermore, from the perspectives of ADF and HD (i.e., distance metrics) one can observe that the mean ADF and HD values for UNet+CSAC were Figure 8. ADF, Hausdorff, DSC, Binary Accuracy, and Jaccard comparison between UNet and UNet+CSAC. 1.83 mm and 1.06 mm lower than the baseline UNet, which highlights the substantial benefit of utilizing the proposed approach.

To statistically verify the reported accuracy improvement (error reduced from 2.5 mm to 0.67 mm), we performed formal significance testing in addition to box plot analysis. After checking normality, we applied a paired t-test (or Wilcoxon signed-rank test when required). The results indicate that the improvement is statistically significant ($p < 0.05$). We also report mean and standard deviation values to support the comparison.

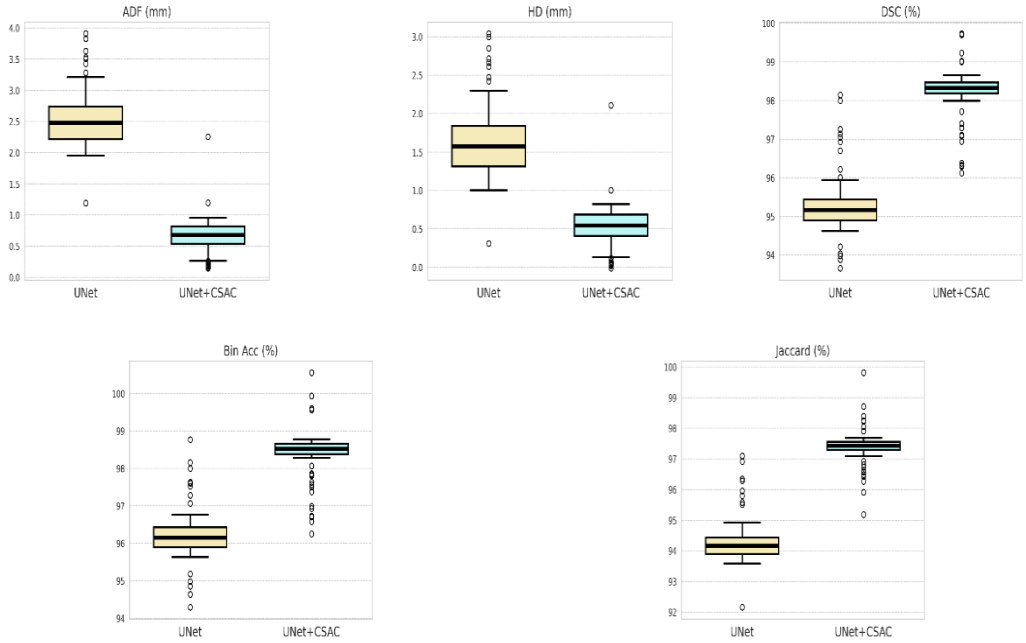

Figure 8. ADF, Hausdorff, DSC, Binary Accuracy, and Jaccard comparison between UNet and UNet+CSAC.

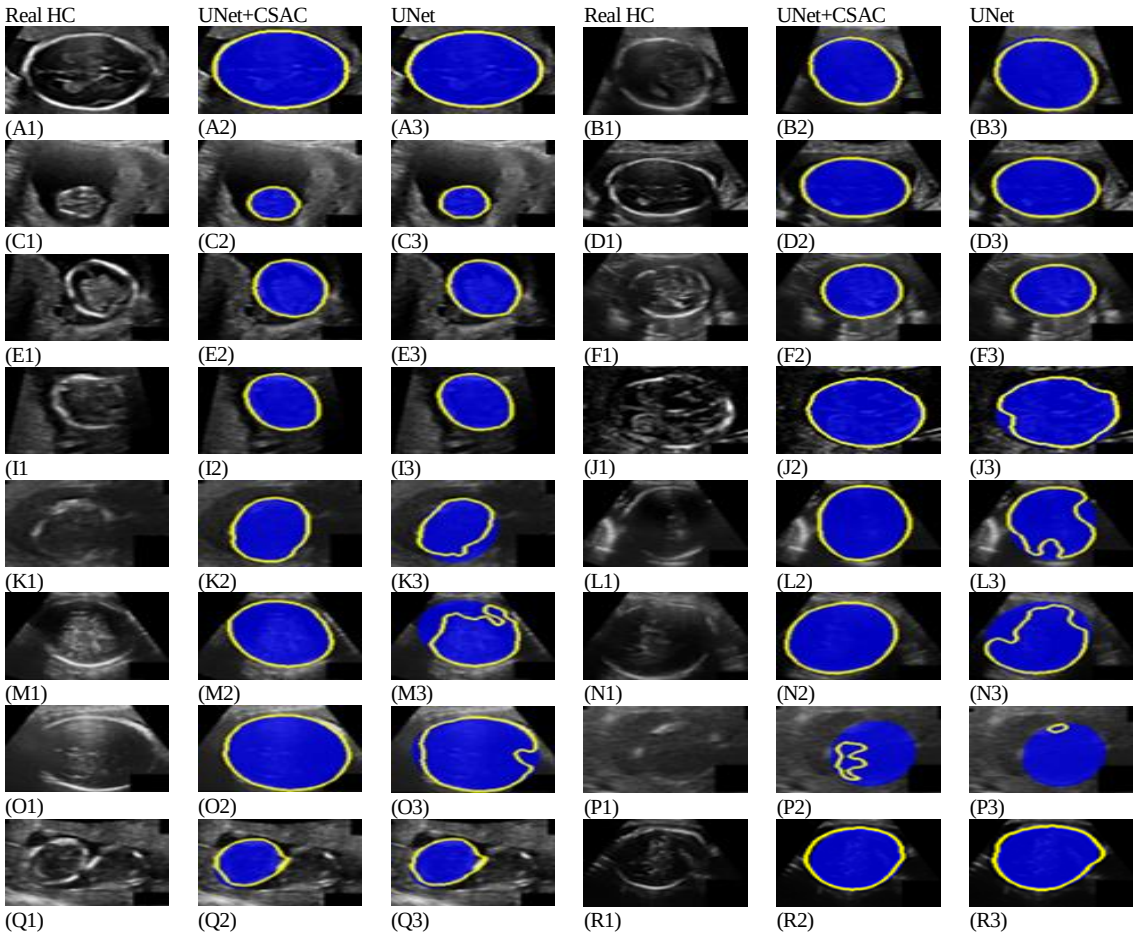

Figure 9 Visual HC results: Sets (A1-A3) per HC image. (A)-(I): Both UNet+CSAC and UNet perform well. (J)-(O): UNet+CSAC outperforms UNet. (P)-(R): Both methods fail.

Discussion

For a better understanding of the reason of the significant improvements in the results due to the use of the proposed method, in Figure 10, examples of the loss curve during training and validation process for both UNet+CSAC and UNet methods have been demonstrated. The five train-validation curves of the top row correspond to the basic method and the curves of the bottom row of this figure are also from UNet+CSAC experiments. The visual inspection of these graphs indicates that the training-validation curves in the basic method have more mismatches than the graphs obtained from proposed method of this article. This fact, especially have become more prominent in the graphs c and d, in the form of sudden jumps, and in graph b, in the form of two curves moving away from each other in the final epochs.

This contrasts with graphs (c)-(d), which illustrate the UNet+CSAC training process where none of these patterns appear. Nevertheless, a more precise understanding of this enhancement emerged when the aforementioned discrepancy was quantified as the mean separation between training and validation curves across the final 10 epochs, and displayed in Figure 11 as a box and whisker diagram for all 200 executions of each method. These boxes reveal that the magnitude of the mean separation between training and validation curves achieved for baseline UNet ranged within [0.05-1.6] with median of 0.25. This contrasts with the proposed approach where, for identical test quantities, the equivalent value only varied within [0.02-0.5] with median of 0.02. Thus, one can observe that the degree of alignment between training and validation curves in UNet+CSAC has not only improved up to twofold compared to baseline UNet, but its consistency has also been enhanced substantially.

Based on what was mentioned in the above paragraphs, the use of the proposed method of this article to strengthen UNet by using the attention-based method known as CSAC may improve the obtained results. This promotion may be justified due to effectiveness of CSAC in reducing the occurrence of the overfitting phenomenon and therefore provide better learning of neural network. Here the question may be raised, why standard methods of this field such as dropout have not been used to reduce the problem of overfitting, and a more complex method has been used, such as CSAC. Table 2 shows that use of the drop-out method has not been able to significantly improve the obtained measurements. Meanwhile, our proposed method in this article creates a considerable improvement in the results, which indicates the significant effectiveness of the proposed method compared to the classical methods for solving overfitting problem.

Another well-known classic method to deal with overfitting is the use of augmentation. Based on this, in another test with the same purpose as what was intended in the dropout test, the performance of data augmentation was compared with the basic method and the proposed method of this article. Comparing the numbers of the first two rows of the Table 3 with the third row clearly indicates that the improvement provided by augmentation is significantly weaker than the improvement provided by our CSAC method. For instance, regarding the AD parameter, which is one of the most important measurements, it can be seen that the use of augmentation has only resulted in a slight improvement compared to UNet scheme (AD reduction from 2.53 mm to 2.48 mm). Meanwhile, using the method proposed in this article (UNet+CSAC) has been able to bring this error to 0.67 mm, which is a significant improvement.

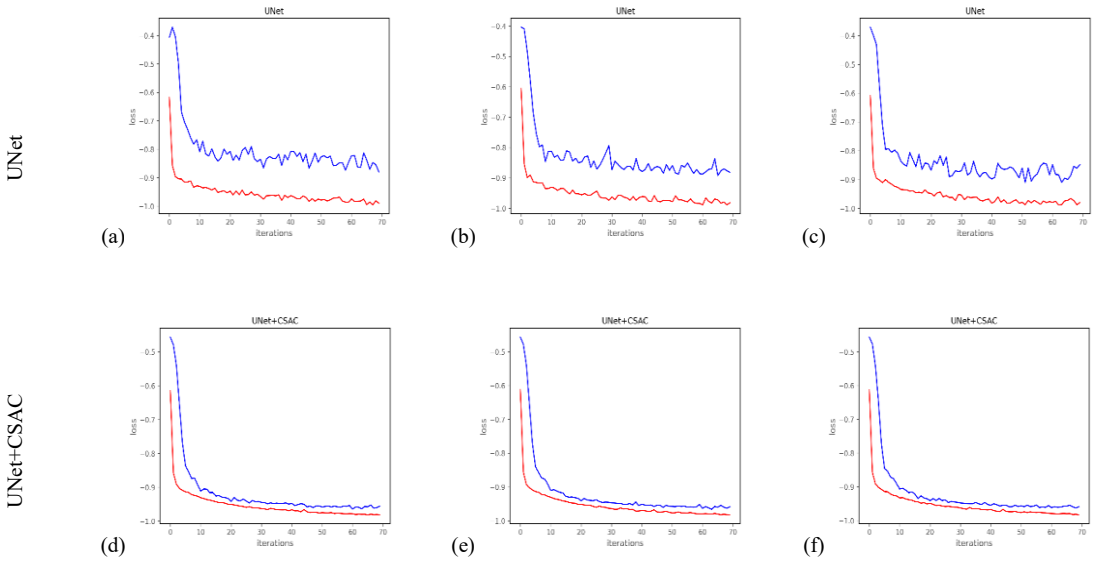

Figure 10. Comparison of some training-validation curves acquired from (a-c). basic UNet method and (d-f). proposed UNet+CSAC method.

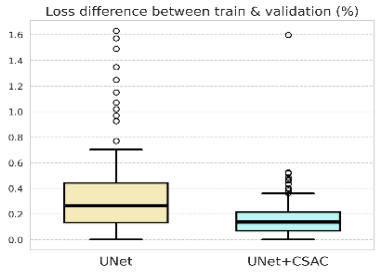

Figure 11. The box-whisker plot which demonstrate the gaps obtained between training and validation curves belonging to UNet and UNet+CSAC methods.

Table 2. Investigating the Effectiveness of classic dropout method.

Method	AD (mm)	HD (mm)	DSC (%)
UNet	2.53±0.91	1.61±0.92	95.22±0.75
UNet+Dropout	2.48±0.83	۱٫۵۴±۱٫۰۱	97.61±0.94
UNet+CSAC	0.67±0.35	0.54±0.31	98.31±0.52

Table 3. Investigating the Effectiveness of classic Augmentation method.

Method	AD (mm)	HD (mm)	DSC (%)
UNet	2.53±0.91	1.61±0.92	95.22±0.75
UNet+Augmentation	2.48±0.59	1.47±0.98	97.58±1.22
UNet+CSAC	0.67±0.35	0.54±0.31	98.31±0.52

Table 4. Comparison between UNet & VNet 2D.

Method	AD (mm)	HD (mm)	DSC (%)
UNet	2.53±0.91	1.61±0.92	95.22±0.75
VNet 2D	2.89 ± 2.20	1.94 ± 1.17	95.22 ± 2.16

Table 5. Comparison between the proposed and some state-of-art methods

Method	AD (mm)	HD (mm)	DSC (%)
Random Forest [26]	2.83±3.16	1.83±1.60	97.10±2.73
GVF-Net [27]	2.42±1.93	2.18±2.40	95.53±3.98
Mask R-CNN [28]	2.33±2.21	1.39±0.82	97.73±1.32
Mini Link-Net [29]	2.12±1.87	1.72±1.39	96.84±2.89
UNet+CSAC	0.67±0.35	0.54±0.31	98.31±0.52

In another test, the validity of the VNet method was tested as a structure of the same family as the UNet on the data of this article. The test conditions were exactly the same as the UNet test conditions and the results of the above two methods are compared in the table 4. The comparisons indicate that UNet model performs better than the 2D version of VNet with all three criteria AD, HD and Dice. It is worth noting that although the duration of program execution was not among the main criteria, but even in terms of program execution speed, UNet performs up to twice faster than VNet. These results, in addition to the favorable performance of UNet in a multitude of medical image segmentation applications that have already been reported in several articles, justify the philosophy of using UNet as the basic model in the experiments of this article.

In order to clearly compare between this article's proposed method and state-of-arts methods, the important articles that used the same data were checked. In this regard, we specifically focused on studies whose evaluation criteria were the same as our criteria. These results show that this article's proposed method has a suitable position among alternatives both in terms of the absolute number of results and in terms of the standard deviation and dispersion of the results. The comparison between the results of this article's proposed method and several prominent works in the field of fetal head circumference detection, all of which have used HC18 data, has been done in the Table 5. In a study Heuvel et al [26] had segmented the HC images using the random forest method. In addition to the fact that this article's proposed method was superior to their method in all components, on the other hand, the standard deviation for this article's proposed method measurement coefficients were significantly less than its alternative. This may be considered as an advantage. In another study Rong et al [27] had used the GVF method and its combination with the UNet (GVF-Net). Considering that

finding the edge is an important thing in segmentation, their edge-based method may be considered appropriate. It may be observed in table that this article's proposed method has a significant advantage over the above method in terms all components. Also, in this article's proposed method, the average standard deviation for this article's measurement coefficients were significantly less than its alternative. In the research of Al-Bander et al [28] the Mask R-CNN was used in order to identify the fetal head area. However, the mentioned method seems rational, but in comparison with this article's proposed method, weaker results have been obtained for all measurement coefficients and deviations. The findings of Sobhaninia et al [29] research showed that a network with small number of CNN blocks can be more effective than a classical neural network with many layers. However, the comparison of the results of article's proposed method scheme and their study indicated the superiority of article's proposed method in all components in parallel with their standard deviations. A similar study used an attention mechanism (DAG) [30] similar to our method, which will be explained further below. The results of these comparisons are shown in the Table 5.

Finally, it should be investigated why the CSAC method was used among the different Attention-based methods in this article. To justify this issue, a well-known method in the field of Attention called DAG (Deeply Attention Gated), which has a history of being used in fetal head circumference measurement [30], was tested. To make the comparisons fair, this module was also added to UNet and its performance was achieved in conditions completely similar to the method proposed in this article. The results obtained from our proposed method (UNet+CSAC) and this alternative (UNet+DAG) have been mentioned in Table 6. These results clearly indicate that under the same conditions, the attention-based method proposed in this article (CSAC) has been able to achieve better performance in all parameters than the its attention-based alternative.

Table 6. Comparison between UNet+CSAC & UNet+DAG.

Method	AD (mm)	HD (mm)	DSC (%)
VNet+ DAG	1.77 ± 1.70	1.27 ± 0.80	97.93 ± 1.25
UNet+DAG	1.69±1.27	1.19 ± 0.74	97.79±1.16
UNet+CSAC	0.67±0.35	0.54±0.31	98.31±0.52

Conclusion

In this article, a new method was proposed to improve the performance of deep neural networks in the automatic estimation of fetal head circumference using ultrasound images. The proposed method incorporates both channel and spatial attention mechanisms to emphasize diagnostically relevant components of the input data during the learning process. This approach reduces the influence of redundant information, which is typically abundant in low-contrast ultrasound images, on the network training procedure.

Evaluation of this technique on ultrasound images from the HC18 dataset demonstrated improvements in both training and testing phases of a UNet architecture. Results from 100 executions showed that the attention-enhanced UNet (UNet+CSAC) reduced overfitting probability by approximately 50% compared to the standard UNet. Additionally, the proposed method achieved a 2% improvement in fetal head region detection accuracy and approximately 1 millimeter improvement in HC estimation compared to the baseline UNet.

These findings suggest that the proposed attention-based approach represents a viable option for automated fetal head analysis in ultrasound imaging, warranting further validation in diverse clinical settings.

Acknowledgment

The authors would like to thank Zenodo community to provide an open access HC dataset.

References

1. Poojari, V. G., Jose, A., & Pai, M. V. (2021). Sonographic estimation of the fetal head circumference: Accuracy and factors affecting the error. The Journal of Obstetrics and Gynecology of India, 1–5.
2. Abramowicz, J. S., & Longman, R. E. (2023). First-trimester ultrasound: A comprehensive guide. Springer Nature.
3. Gabbe, S. G., Niebyl, J. R., Simpson, J. L., Landon, M. B., Galan, H. L., Jauniaux, E. R., Driscoll, D. A., Berghella, V., & Grobman, W. A. (2016). Obstetrics: Normal and problem pregnancies e-book. Elsevier Health Sciences.
4. Zeng, W., Luo, J., Cheng, J., & Lu, Y. (2022). Efficient fetal ultrasound image segmentation for automatic head circumference measurement using a lightweight deep convolutional neural network. Medical Physics, 49(8), 5081–5092.
5. Wu, L., Cheng, J.-Z., Li, S., Lei, B., Wang, T., & Ni, D. (2017). FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks. IEEE Transactions on Cybernetics, 47(5), 1336–1349.
6. Rasheed, K., Junejo, F., Malik, A., & Saqib, M. (2021). Automated fetal head classification and segmentation using ultrasound video. IEEE Access, 9, 160249–160267.
7. Perez-Gonzalez, J., Muñoz, J. B., Porras, M. R., Arámbula-Cosío, F., & Medina-Bañuelos, V. (2015). Automatic fetal head measurements from ultrasound images using optimal ellipse detection and texture maps. VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina.
8. van den Heuvel, T. L., de Bruijn, D., de Korte, C. L., & Ginneken, B. v. (2018). Automated measurement of fetal head circumference using 2D ultrasound images. PLoS One, 13(8), e0200412.
9. Zhang, L., Ye, X., Lambrou, T., Duan, W., Allinson, N., & Dudley, N. J. (2016). A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2D

ultrasound images. Physics in Medicine & Biology, 61(3), 1095.
10. Rueda, S., Fathima, S., Knight, C. L., Yaqub, M., Papageorghiou, A. T., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., & Tohka, J. (2013). Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge. IEEE Transactions on Medical Imaging, 33(4), 797–813.
11. Ni, D., Yang, Y., Li, S., Qin, J., Ouyang, S., Wang, T., & Heng, P. A. (2013). Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters. 2013 IEEE 10th International Symposium on Biomedical Imaging.
12. Alzubaidi, M., Agus, M., Shah, U., Makhlouf, M., Alyafei, K., & Househ, M. (2022). Ensemble transfer learning for fetal head analysis: From segmentation to gestational age and weight prediction. Diagnostics, 12(9), 2229. https://www.mdpi.com/2075-4418/12/9/2229
13. Dai, W., & Zhai, J. (2023). Fetal head circumference detection based on Dlink-Net model. Chinese Intelligent Systems Conference.
14. Wang, X., Wang, W., & Cai, X. (2022). Automatic measurement of fetal head circumference using a novel GCN-assisted deep convolutional network. Computers in Biology and Medicine, 145, 105515.
15. Wu, L., Xin, Y., Li, S., Wang, T., Heng, P.-A., & Ni, D. (2017). Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).
16. van den Heuvel, T. L., Petros, H., Santini, S., de Korte, C. L., & van Ginneken, B. (2019). Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries. Ultrasound in Medicine & Biology, 45(3), 773–785.
17. Zhang, J., Petitjean, C., Lopez, P., & Ainouz, S. (2020). Direct estimation of fetal head circumference from ultrasound images based on regression CNN. Medical Imaging with Deep Learning.
18. Skeika, E. L., Da Luz, M. R., Fernandes, B. J. T., Siqueira, H. V., & De Andrade, M. L. S. C. (2020). Convolutional neural network to detect and measure fetal skull circumference in ultrasound imaging. IEEE Access, 8, 191519–191529.
19. Zeng, W., Luo, J., Cheng, J., & Lu, Y. (2022). Efficient fetal ultrasound image segmentation for automatic head circumference measurement using a lightweight deep convolutional neural network. Medical Physics, 49(8), 5081–5092.
20. Horgan, R., Nehme, L., & Abuhamad, A. (2023). Artificial intelligence in obstetric ultrasound: A scoping review. Prenatal Diagnosis, 43(9), 1176–1219.
21. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., & Wu, J. (2020). UNet 3+: A full-scale connected UNet for medical image segmentation. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
22. Xu, Y., Hou, S., Wang, X., Li, D., & Lu, L. (2023). A medical image segmentation method based on improved UNet 3+ network. Diagnostics, 13(3), 576.
23. Larochelle, H., & Hinton, G. E. (2010). Learning to combine foveal glimpses with a third-order Boltzmann machine. Advances in Neural Information Processing Systems, 23.
24. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., & Tang, X. (2017). Residual attention network for image classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
25. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
26. van den Heuvel, T. L., de Bruijn, D., de Korte, C. L., & van Ginneken, B. (2018). Automated measurement of fetal head circumference using 2D ultrasound images. PLoS One, 13(8), e0200412.
27. Rong, Y., Xiang, D., Zhu, W., Shi, F., Gao, E., Fan, Z., & Chen, X. (2019). Deriving external forces via convolutional neural networks for biomedical image segmentation. Biomedical Optics Express, 10(8), 3800–3814.
28. Al-Bander, B., Alzahrani, T., Alzahrani, S., Williams, B. M., & Zheng, Y. (2019). Improving fetal head contour detection by object localisation with deep learning. Annual Conference on Medical Image Understanding and Analysis (pp. 142–150). Springer.
29. Sobhaninia, Z., Emami, A., Karimi, N., & Samavi, S. (2019). Localization of fetal head in ultrasound images by multiscale view and deep neural networks. arXiv preprint.
30. Zeng, Y., Tsui, P. H., Wu, W., Zhou, Z., & Wu, S. (2021). Fetal ultrasound image segmentation for automatic head circumference biometry using deeply supervised attention-gated V-Net. Journal of Digital Imaging, 34, 134–148.
31. Online: http://doi.org/10.5281/zenodo.1322001