http://jhmi.sums.ac.ir
Health Man & Info Sci 2023, 10(4), 194–207

HMIS
Health Management and Information Science

Original Article

Automated Fetal Head Circumference Measurement by Ultrasound using V-NET and Data Augmentation

Seyed Vahab Shojaedini[1*], Amir Sanyian[2], Mohammadreza Riahi[3], Mahsa Monajemi[4]

[1]Department of Biomedical Engineering, Iranian Research Organization for Science and Technology, Tehran, Iran
[2]Department of Computer Engineering, Faculty of Engineering, Islamic Azad University, Qazvin branch, Qazvin, Iran
[3]Department of Electrical Engineering, Shahid Beheshti University, Tehran, Iran
[4]Department of Biomedical Engineering, Faculty of Engineering, Islamic Azad University, Qazvin branch, Qazvin, Iran

Abstract

Introduction: One of the common methods for monitoring fetal growth is measuring its head circumference in ultrasound images taken from the mother's womb. In recent years, utilizing deep learning methods have been expanded in this application thanks to its potential in promoting the accuracy of estimating head circumference. However, the performance of deep neural networks is highly dependent on the volume of training data. On the other hand, the region of the fetal head is segmented with considerable errors, due to the presence of various types of noise in ultrasound images which leads to erroneous estimation of fetal head circumference. In this article, a new method is presented to improve fetal head circumference estimation in ultrasound images in which, by using unsupervised data augmentation, an attempt is made to increase the amount of training data of the deep network.

Methods: By utilizing an elliptical contour estimation method, an optimal contour was created to decrease the segmentation errors of the fetal head.

Results: Comparing the performance of the proposed scheme with the basic method as well as state-of-art schemes showed the improvement of fetal head circumference estimation with the help of the proposed algorithm in such way that not only the quality of fetal head circumference measurement with the Dice parameter has been improved by 0.6% and 3.24% respectively compared to the closest alternative and the basic method but also the variance of the obtained results in both types of these comparisons have improved dramatically.

Conclusion: These achievements demonstrate that the performance of the proposed method is also more focused and reliable in addition to being more accurate.

Keywords: Deep neural networks, Fetal head circumference, Ultrasound, Unsupervised data augmentation, Optimal contour estimation

Article History:
Received: 11 April 2023
Accepted: 19 July 2023

Please cite this paper as:
Shojaedini SV, Sanyian A, Riahi MR, Monajemi M. Automated Fetal Head Circumference Measurement by Ultrasound using V-NET and Data Augmentation. Health Man & Info Sci. 2023; 10(4): 194-207. doi: 10.30476/jhmi.2024.100180.1185.

***Correspondence to:**
Seyed Vahab Shojaedini, Department of Biomedical Engineering, Iranian Research Organization for Science & Technology, Tehran, Iran
Email: shojaeddini_va@yahoo.com

Introduction

Diagnostic ultrasound is a traditional imaging modality that utilizes either reflection or scattering of ultrasound waves to create images from soft organs. These images may provide valuable information for diagnosing and treating diseases. Thanks to the specifications of sound waves; this modality does not have an invasive and radioactive nature like CT or MRI. Therefore, it may be used in many clinical applications like in pregnancy clinics. The ultrasound images provide information about the fetus's Head Circumference (HC) and its hip joint. Obstetricians can estimate the gestational age, size, and weight from the fetus's HC and further the fetal movement in order to identify congenital diseases (1, 2).

In the first trimester of pregnancy, the doctors can measure the age of the fetus from the difference between the head and the seat of the fetus. Afterward, they estimate the fetal age from the head diameter and fetal femur length (3). The fetal situation is natural if the fetal age differs from the estimated age between one to ten days. However, if this difference is more than ten days, the obstetrician should check the health conditions of the fetus and the mother by using other examinations.

In conventional analysis, obstetricians measure the border around the fetal head manually.

Therefore, this method is highly subjective, its performance is limited by several parameters, and the most important challenge is different diagnoses by different specialists. In practice, this process requires specialized knowledge in parallel with the fact that it is tedious and time-consuming (4). As an alternative, automated image segmentation may be considered as an effective way to extract the fetal head contour (5). Unfortunately, there are some challenges that have caused automated segmentation of the fetal head to remain an open problem. Some of the above factors are the low contrast of the ultrasonic image under various conditions, its low signal-to-noise ratio, acoustic shadows, and speckle noise (5, 6).

During recent years, researchers have proposed a number of methods in order to solve the above limitations. These methods are divided into two categories: methods which are based on extracting handcrafted features and deep learning-based methods. Van den Heuvel et al. (7) propose a two-dimensional computer-aided diagnosis (CAD) to automatically measures the fetal head circumference from two dimensional images for the entire trimester of pregnancy. They extract Haar features to train a random forest classifier in order to locate the fetal skull. Finally, they estimate the fetal HC by using a combination of the Huff transform, dynamic programming, and oval fitting.

Zhang et al. (8) use a supervised learning-based approach for accurate segmentation. They use a multistep strategy for the measurement of the fetal head. In the first step, they reduce the spectral noise using a nonlinear propagation technique. Next, they utilize a multidimensional and multidirectional filter bank to extract specific features for the anatomical structure of the fetus. It is due to the assumption that brightness of the skull cross-section has an approximate Gauss-like curve. Finally, they construct a contour for the fetal head by using an oval curve fitting method. Gonzalez et al. (9) propose a fully automated approach for segmentation and measurement of the fetal head which mainly relies on tissue mapping and optimal oval recognition. They combine tissue mapping, morphological operations, and active contour with optimal oval detection. The final segmentation and measurement of the fetal skull are relatively acceptable.

Reudo et al. (10) propose a method based on distance calculation and blend-Altman measurements in order to segment the abdomen, head, and femur of the fetus in ultrasound images. Although they indicate that this approach can be used in clinical applications, it has poor performance for femur segmentation compared to the head. This weakness has two main reasons. The first is that the femur segmentation is more challenging compared to head segmentation. The second one is high dependency of this algorithm to the appearance of the femur.

Ni et al. (11) propose an automated method to detect fetal head based on the AdaBoost learning algorithm and then measure the head circumference by using the detected fetal head. To achieve significant improvement in both detection and speed, they train the AdaBoost classifier by using pseudo-rabid features. To reduce the sensitivity to spectral noise and intensity changes in ultrasound images, they identify the head boundary in the local area by utilizing a local phase-based method. Finally, they use the Hough transform in order to determine an ellipse on the contour of the head. In another research, Jatmiko et al. (12) combine the improved efficient Hough transform and boosting concept to detect the ellipse shape organs like the fetal head, abdomen, and femur.

Another family of methods for fetal head parameter estimation is the deep learning-based schemes in them; either the feature extraction or classification stages are within the network. Therefore, they are called end-to-end networks. The superior achievements of these networks in solving different problems make them popular among researchers. In the feature extraction stage, the network extracts low-level features and combines them in order to build high-level features. Next, a fully connected neural network is used for the classification stage.

Sundaresan et al. (13) propose an end-to-end convolutional neural network for detecting the fetal heart in the ultrasound video frames (echocardiography). Their purpose is detection of cardiac abnormalities and congenital heart disease. Also, Wu et al. (14) propose the cascading neural network method to fully automate the segment of the skull and abdomen of the fetus. Sinclair et al. (15) use the transfer learning technique based on VGG16 for estimation of the region of the fetal head. Unfortunately, this work has poor performance with the Dice evaluation criterion.

Sobhaninia et al. (16) propose a deep neural network based on the Link-Net for fetal skull segmentation in 2-D ultrasound images. They use a CNN for segmentation and oval estimation by minimizing a cost function. Experimental results on the fetal ultrasound dataset in different trimesters of pregnancy have been relatively consistent with the radiologist's diagnosis. Sobhaninia et al. (17) also propose another approach based on multidimensional CNN for measurement of the fetal head circumference. Although this network has fewer parameters and trains faster than conventional CNN, its results on ultrasound data sets of fetuses in different trimesters of pregnancy show some errors. Some deep learning-based methods use supervised augmentation in order to promote the quality of their training (16, 17). Also, some researchers use data augmentation that does not corrupt the image (e.g., flip) (15).

The aforementioned studies mainly indicate that deep learning methods have significant achievements for fetal HC segmentation, but the existence of sufficient data is one of the main requirements for them to work properly. Unfortunately, the accessible dataset for medical purposes in this area is limited. Therefore, data augmentation may be considered as a potential solution. Some augmentations like rotation and translation can corrupt the ultrasound images. Thus, some researchers prune the augmented images or use augmentation methods without destruction (15-17). Thus, here, this important question is raised whether the generation of fake data (We created new data by performing a series of operations on the original data, such as data augmentation, and in order not to be confused with the original data, we have called them "fake data" in our text; i.e., augmented ultrasound images), even at the cost of causing partial destruction in them, may lead to an improvement in the quality of classification and thus promote the overall results. In this study, it is shown that such a slight destruction can even have a constructive role in the generalization of the model. Accordingly, this article presents a novel method for training convolutional neural networks (CNNs) for fetal head circumference segmentation in ultrasound images using unsupervised data augmentation. The proposed approach leverages various data augmentation techniques, without the need for human supervision. The rationale behind this approach is that by carefully selecting the augmentation parameters, the generated images can still preserve the essential features of the fetal head, thereby improving the generalization ability of the deep learning model without the overhead of manual curation.

Methods

There is a trade-off between the number of training data and the model size to train a reliable convolutional neural network (CNN) model. A challenge with higher data complexity may require more complicated CNN models. These models have a large number of parameters and are prone to overfitting. Therefore, they may need massive training dataset, and collecting and preparing such a dataset is an overwhelming task. One common technique in the deep learning context is increasing the number of the training dataset by generating some fake data. Such a process may decrease the chance of overfitting. By using such fake data in training of the deep model, it may better generalize unseen data. There are some ways for generating such fake data. In computer vision, some of these techniques are based on geometric manipulation like flip, rotation, scale, crop, and translation, while others are not related to geometric manipulation like jpeg noise, Gaussian noise, changing color, and blur filters. The wise choice of these techniques is crucial. However, some methods are not suitable for some problems. For example, the jpeg noise or blur filters may negatively affect the training process for ultrasound images (e.g., our problem). Sometimes, such a fake data generation can change the image label. If these labels are not refined, they can introduce noise labels.

Some augmentation techniques like rotation and translation can remove some parts of an image. The deleted part can contain some parts of the fetus's skull. Therefore, there was human supervision for augmented images in recent research works (16, 17).In other words, a person reviews all augmented images and deletes images with a partial fetus's skull which is tedious. Block diagram of supervised image augmentation is shown in Figure 1a.

One may ask about the essence of supervision on these augmented images. To the best of our knowledge, there is no attempt to eliminate the human supervision of augmented training images. We can adjust the destructions of

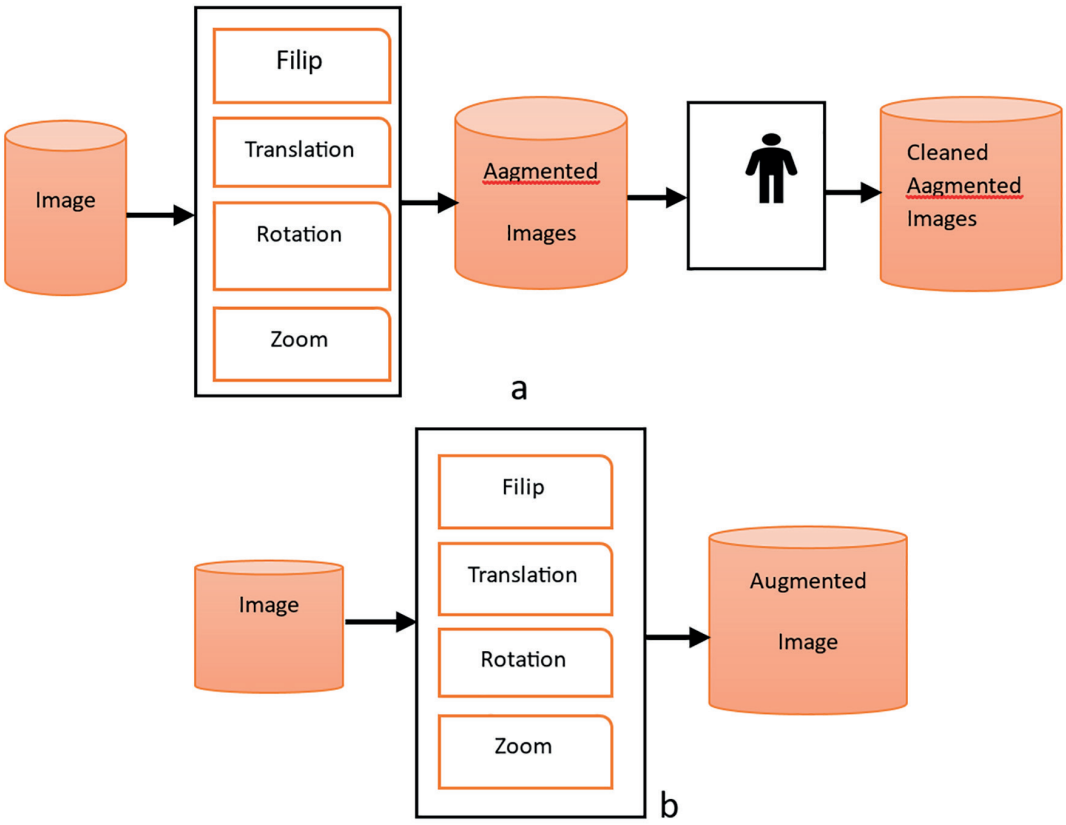

Figure 1: Supervised data-augmentation diagram (a) vs unsupervised data-augmentation diagram (b)

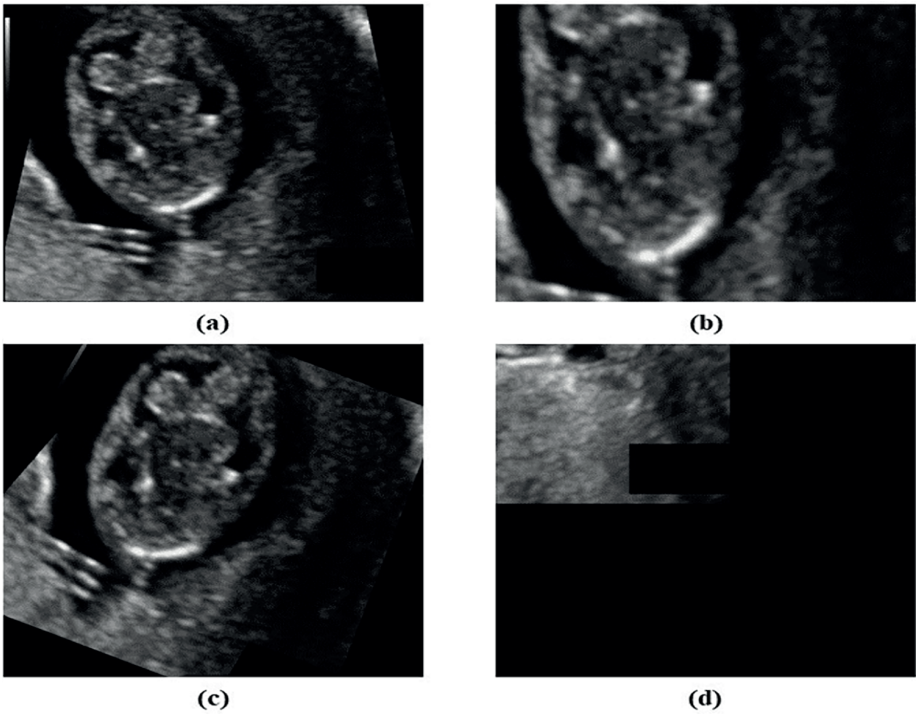

Figure 2: (a) the original image (b) corrupted image by zoom augmentation with 0.5 scale factor for height and width (c) corrupted image by rotation augmentation with 0.9 rotation factor (d) corrupted image by translation augmentation with 0.8 translation factor for height and width

an augmented image by choosing the augmentation parameters. For example, if the rotation degree sets to 45°, it may remove the major part of the fetus's skull. Therefore, we can ignore the human supervision for training images by selecting the augmentation parameters carefully. The pipeline of this unsupervised data augmentation is shown in Figure 1b. Examples of these disruptions are shown in Figure 2.

In this study, a version of fake data generation is introduced without any human monitoring for training a V-Net simultaneously. In other words, we do not save and prune the augmented training dataset and directly feed it into a V-Net. It is assumed that most of the fetus's skull is in the generated images. Therefore, some items including flip, translation, rotation, and zoom are utilized in the framework of the above idea. In Figures 3, 4 and 5, some generated images are shown in parallel with their labels. As a detailed explanation, since flip augmentation preserves the image content and its label, the images and their corresponding labels are flipped horizontally, vertically, or both. In translation technique, each pixel of an image can translate with the equation (1):

$$\begin{bmatrix}\hat{x}\\ \hat{y}\\ 1\end{bmatrix} = \begin{bmatrix}1 & 0 & x_0\\ 0 & 1 & y_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \qquad (1)$$

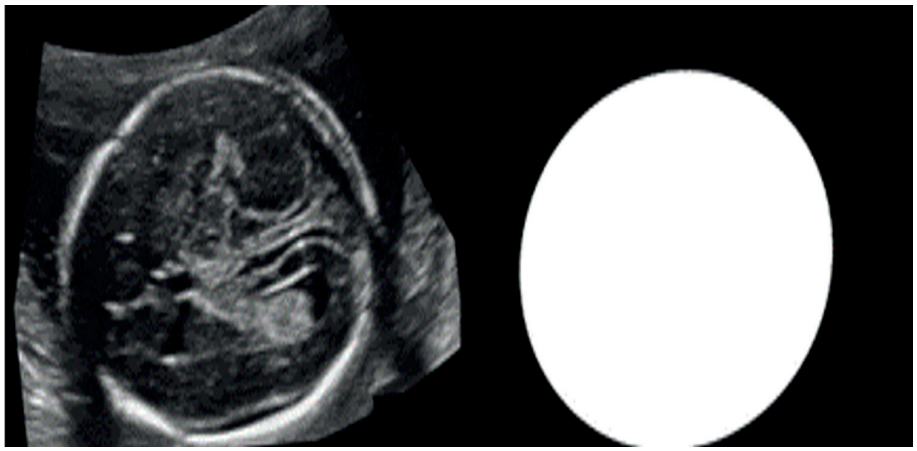

Figure 3: Corrupted image with random zoom

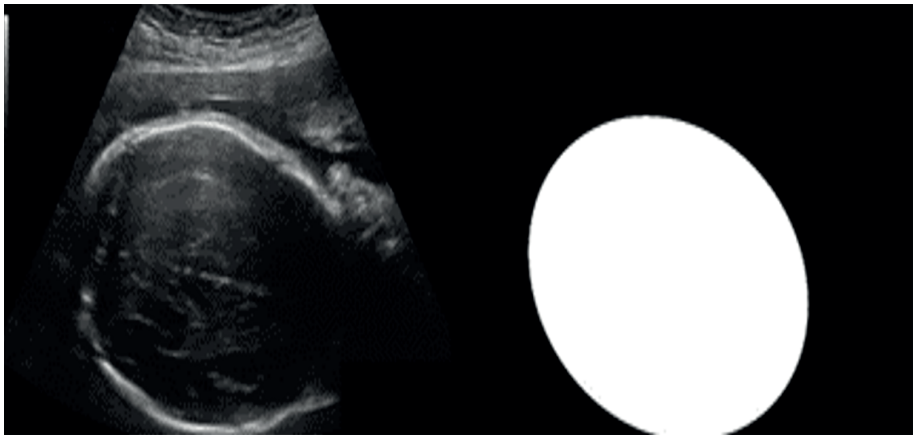

Figure 4: Corrupted image with random rotation

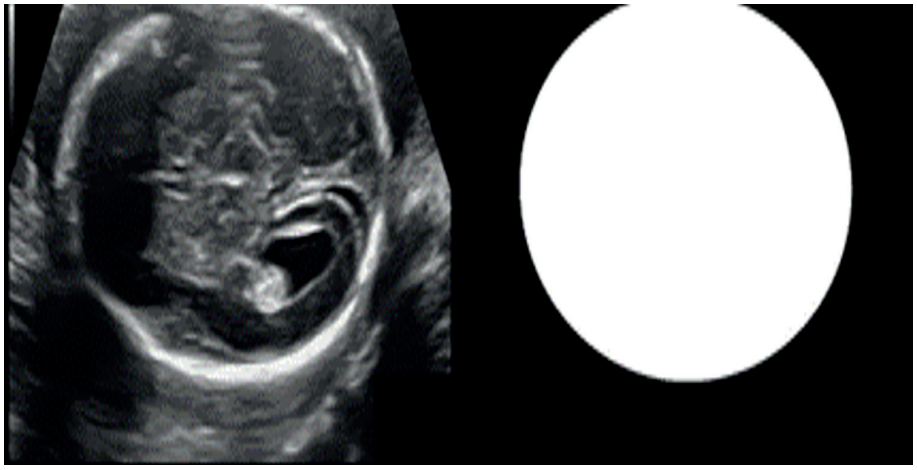

Figure 5: Corrupted image with random translation

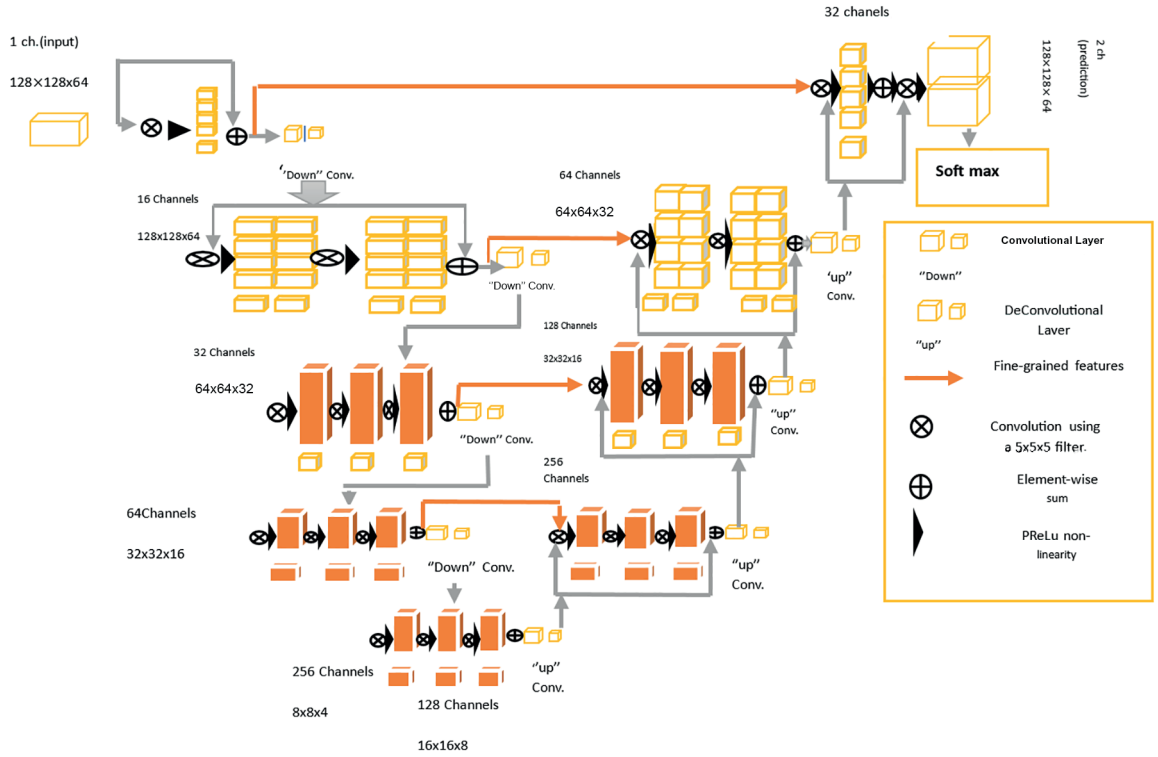

Figure 6: V-Net architecture for segmentation

where x_0 and y_0 are the translation values in x and y axes, respectively. The images are translated horizontally, vertically or both. For pixels outside the translated image, the same constant k with a value of 0.05 is applied. In rotation augmentation method, the transformation matrix of rotation is described in equation (2):

$$\begin{bmatrix}\hat{x}\\ \hat{y}\\ 1\end{bmatrix} = \begin{bmatrix}\cos(\theta) & \sin(\theta) & 0\\ -\sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \quad (2)$$

where the rotation degree is specified by θ. Each image is rotated in the range leading to 20° in either clockwise or counterclockwise directions. Finally, the scale transformation matrix of zoom augmentation technique is demonstrated in equation (3).

$$\begin{bmatrix}\hat{x}\\ \hat{y}\\ 1\end{bmatrix} = \begin{bmatrix}s_1 & 0 & 0\\ 0 & s_2 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \quad (3)$$

In which s_1 and s_2 are the scaling factor for x and y axes. The zoom-in or zoom-out process are performed randomly on images by utilizing a zooming factor up to 0.05 for the height and width while preserving the aspect ratio. The pixels outside the zoomed image are filled with the same constant with a value of 0. The Convolutional Neural Network for Volumetric Medical Image Segmentation (i.e., V-Net) was used as the core of deep learning in this article. This structure which is demonstrated in Figure 6 tries to perform fetal head segmentation in 2D ultrasound images. Based on the architecture, this network consists of the two branches. The first branch (i.e., left part) is a compression path**,** which is divided into different levels, and in each of them residual training is performed, consisting of one to three convolutional layers.

Along this branch, resolution is reduced during convolution and stride operation. This leads to halving the size of the resulting feature maps, similar to what happens in the pooling process in standard CNN deep networks. Correspondingly, the number of feature channels doubles at each level of the branch. The second branch (i.e., right part) decompresses the signal until its original size is reached. It expands the spatial support of the lower resolution feature maps in such way that two feature maps which are computed by convolutional layer are converted to probabilistic segmentations of the foreground and background regions by applying residual learning (same as the left branch) and soft-max voxel-wise.

After training a reliable model, a post-processing technique is used to estimate the extent of the fetal head and remove unwanted parts. Accordingly, first, the segmented image

is converted into a binary image by utilizing the thresholding method. Then, different contours are fitted to the binary image. Firstly, the largest contour among the above boundaries is selected as the fetal head pixels. Finally, the algebraic distance algorithm is applied in order to fit an ellipse on it. The basis of this algorithm is the minimization of an objective function which demonstrates the distance between extracted points and constructed curve. In this article, the conic fitting algorithm is utilized in order to solve such minimization problem as described below. The main components of this approach are as follows:

(a) A set of data points $p = \{a_i\}_{i=1}^{m}$ in which $a_i = (a_i, b_i)$;

(b) A family of curves $c(n)$ parameterized by the vector n;

(c) A distance metric $\delta(c(n), a)$ which measures the distance from a point a to the curve $c(n)$;

The main purpose of this problem is to find the value of parameter n for which the error function ϵ^2 (n) which is demonstrated by equation (4), reaches its global minimum. The curve $c(n)$ is then the curve that best fits the data. This best fitting value of n is called minimizer n and in this paper is shown as n_{min}.

$$\epsilon^2(\mathrm{n}) = \sum_{i=1}^{m} \delta(c(n), a_i) \qquad (4)$$

In the approach of this paper, the curve families considered are represented in the implicit form described in equation (5):

$$c(n) = \{a | F(n; a) = 0\} \qquad (5)$$

The family that we examine are general conic sections, with the following equation (6):

$$F(n; a) = A_{aa}a^2 + A_{ab}ab + A_{bb}b^2 + A_a a + A_b b + A_0 \qquad (6)$$

This equation can be rewritten by using the concept of dot product in such a way that data points and model parameters are separated.

$$F(n; a_i) = [a_i^2 a b_i b_i^2 a_i b_i].[A_{aa}A_{ab}A_{bb}A_a A_b A_0] = a_i.n \qquad (7)$$

Based on the concept of equation (8), the algebraic distance algorithm minimizes th objective function equation (8):

$$\epsilon^2(n) = \sum_{i=1}^{m} F(n; a)^2 = \|D_n\|^2 \qquad (8)$$

subject to the constraint that $\|n\|^2 = 1$.

In which the design matrix D is the $m \times 6$ matrix with row elements of a_i. The constrained objective function is defined as:

$$E = \|D_n\|^2 - \lambda(\|n\| - 1) = n^T D^T D_n - \lambda(n^T n - 1) \qquad (9)$$

It is minimized analytically to form an eigenvector problem (18) as equation (10):

$$\nabla_n E = 0 \Leftrightarrow 2D^T D_n - 2\lambda n = 0 \qquad (10)$$

In the above equation, λ is a Lagrange multiplier. The variable n_{min} which we mentioned as best fitting value before the relation (4) is also obtained as the eigenvector of $D^T Dn$ corresponding to the smallest eigenvalue based on the result of equation (10).

Results

In order to evaluate the effectiveness of the proposed method, it is necessary to simulate it in software framework followed by testing on real ultrasound images of the fetus. These images should be such that in them the border around the fetus's head is specified both in contour and numerically. In this way, the correctness of the contour and circumference resulting from the proposed method can be evaluated. To prepare the software, it is also necessary to implement the fetal head circumference estimation algorithm once by applying the method proposed in this article regarding the production of false images and once again without such innovation. Thus, python3 and Tensorflow paradigms were utilized as the testbed for implementation of the both methods on Intel ® Core i7-10700 computer with Ubuntu 20.04 operation system, 32 GB RAM, and an NVIDIA 2080 Ti. Furthermore, the OpenCV was used as a testbed for postprocessing algorithm including thresholding, contour extraction, and ellipse fitting. As explained above, the evaluation of the proposed method and any comparison requires a dataset in which the contours and numbers of the circumference of the head are precisely known. Thus, a dataset of HC18 challenge (9) was used in this research. This dataset contains 999 2D ultrasound images of the fetal head in the standard plane. Each ultrasonic image had 800 by 540 pixels with a pixel size ranging from 0.052 to 0.326 mm. This set was randomly divided into 700 images for training and 299 to test. In the training process, the 5-fold cross-validation strategy was adopted; thus, 20% of the training images were used for validation in each round of training. In the training set, the fetal head circumference (HC) has been

manually annotated by a trained sonographer. In the next step, based on the contour drawn by the sonographer, the annotated image is divided into two white and black regions by a thresholding-based software. The area inside the contour drawn by the sonographer is indicated with white color, while the area outside the contour is marked with black background. The resultant images are used as reference for training the network and measuring the results. A pair of the ultrasound image and its annotated label are shown in Figure 7. In addition, the rotation parameter varied from 0 to 180 degrees, and the zoom parameter was changed from 0.8 to 2, in experiments. Furthermore, both the horizontal and vertical translation values were also changed from zero to 10%. By combining the above scenarios, the augmentation process was performed.

Finally, in order to carry out the comparison process correctly, it is necessary that the conventional parameters used in the field of fetal head circumference measurement are also used as a basis for comparisons in this research. However, the most concrete parameter related to the comparison is fetal HC difference which is the difference between the estimated HC and its actual size. In equation 11, the HC_s is the estimated HC, and HC_a is the actual HC.

$$DF = HC_s - HC_a \qquad (11)$$

In addition to this quantitative parameter, a visual comparison of the border drawn for the remote environment in any method with the Benchmark border can also be inspiring. Thus, in this part of the article, we make comparisons on limited results and with the help of these two parameters to get a proper sense of how the proposed method works; then, in the next part, by adding other parameters, the evaluation will be done quantitatively and based on the entire set of images.

After the tests were conducted in two scenarios without and with using fake images, the obtained results clearly indicated the superiority of the method proposed in this article against its alternative which does not make use of fake images (i.e., basic method). Figures 8 and 9 show some visual results of these tests.

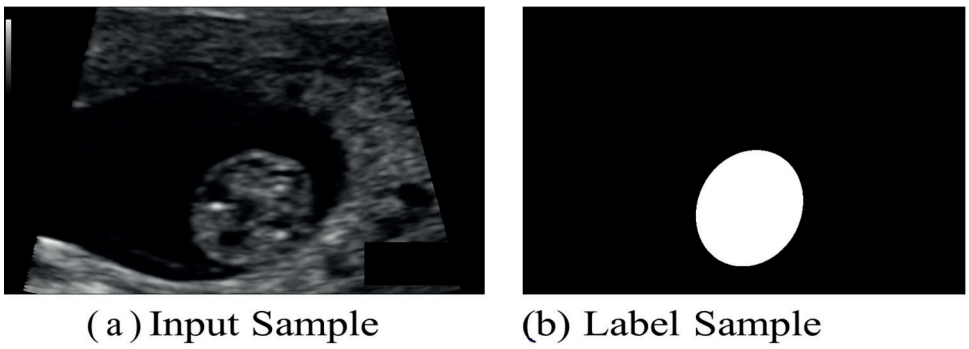

Figure 7: A sample image and its label

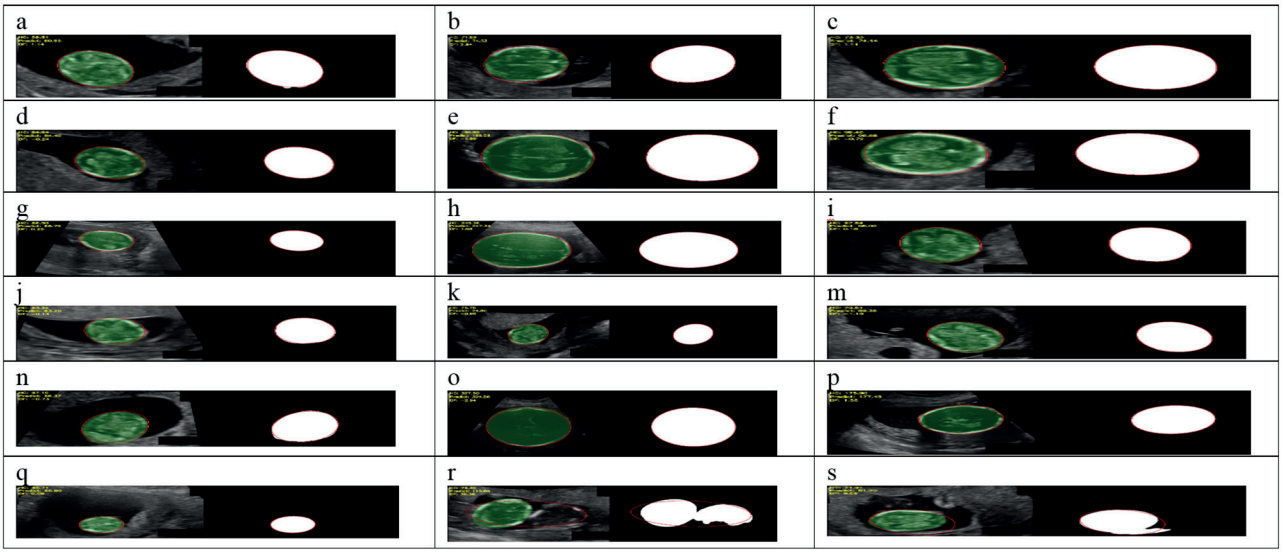

Figure 8: Some results obtained from the proposed method in parallel with their labeled references

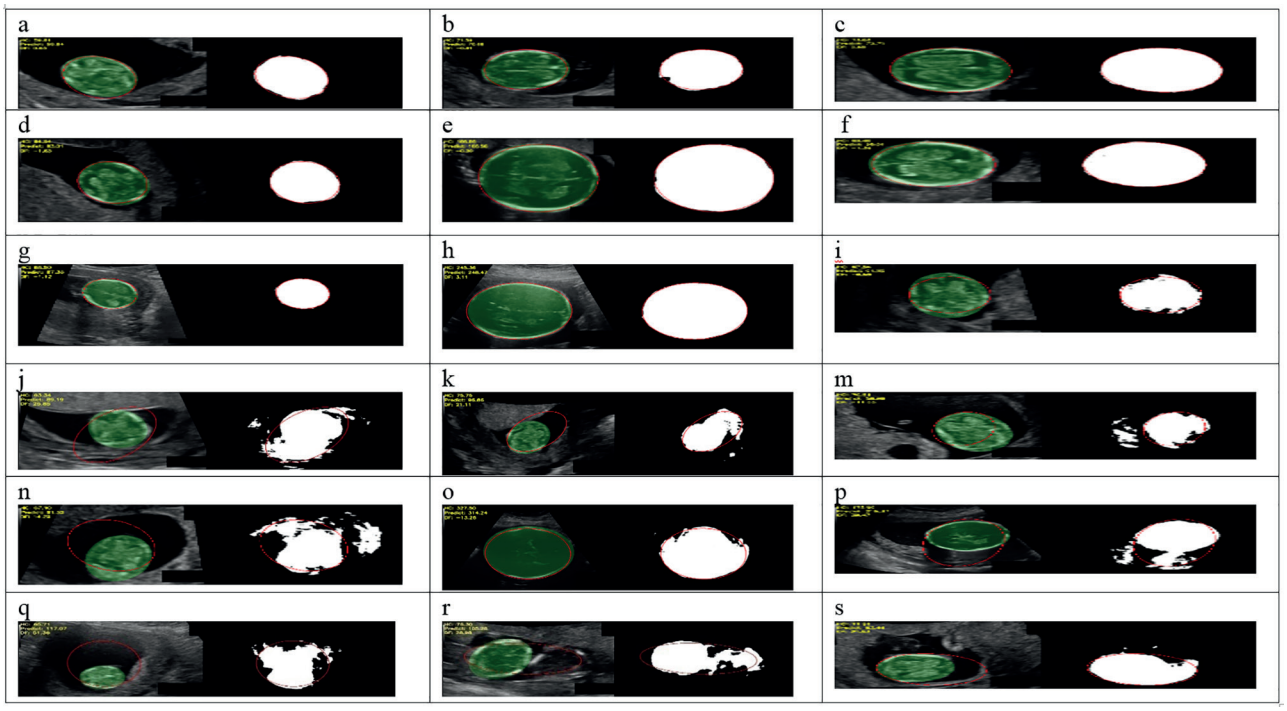

Figure 9: The results obtained from testing the basic method on those images that the results of the proposed method have been reported on them

In each of two above Figures, 8 samples of the results obtained from the proposed and basic methods are shown parallel to the relevant reference image, which shows how well the result of the proposed method matches the original contour (path of the red color). Also, the values of HC and DF parameters are included next to each image in order to clarify the quality of the results.

$$DSC = \frac{2 \times Area(GT) \cap Area(P)}{Area(GT) + Area(P)} \quad (12)$$

$$ADF = |HC(P) - HC(GT)| \quad (13)$$

$$Jaccard\ (X, Y) = \frac{X \cap Y}{X \cup Y} \quad (14)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

In general, these eighteen images, which actually represent the diversity in the results of the tests, can be analyzed in three main categories. The first group of images includes samples (a) to (h), the cases where the detection resulted from the basic method and the proposed method of this article were both acceptable.

In Table 1, the structure of the proposed method is illustrated. This architecture and training strategy aims to effectively segment the fetal head circumference in the ultrasound image dataset.

Table 1: Structure of deep learning method

Parameter	Description
Structure	4 convolutions, 4 max pooling, 4 encoders and decoders
Config	CNN (32,64,128,256)->MP(2,2)->Sig->Relu
Encoders, Decoders	4 encoder, 4 decoder
Number of trainable params	30,235,922
Optimizer	Adam
Batch size	32
Activation functions	ReLU
Loss function	Binary Cross entropy
Number of folds	5
Epochs per fold	20
Learning rate	1e-4

For example, in the sample shown in 8-c and 9-c, it can be observed that the measurement errors of the proposed and basic methods are 0.68 and 1.11, respectively, which both may be considered as negligible deviations and are not significantly different from each other.

The contour obtained in both these results shows acceptable match with the ground truth

contour (the contour drawn around the green area). In other samples of this category, we see a more or less similar situation. Although it should be noted that in this type of results, sometimes the results of the proposed method are slightly superior to the basic method (such as the cases of f, g and h), and sometimes the results of the basic method are somewhat better (such as the cases of a-e). However, as both types of results are acceptable, these small differences do not seem meaningful. The second category of samples are cases in which the quality of the results obtained from the proposed method is significantly better than those related to the basic method. In this type of results that obviously include images (i-q), the basic method has practically led to unacceptable boundaries around the tissue, while the boundary and the amount of error obtained in the estimation of the proposed method were acceptable. For instance, it is possible to check the images (8-i) and (9-i), in which the method of this article has led to a deviation of 0.18 which is acceptable and the basic method has led to a deviation of 6.59 (that is, about 37 times the error of the proposed method) which seems completely unacceptable. The comparison of the obtained texture boundaries in these two samples and the extent of their agreement with the ground truth (the path around the green reference area) also confirms the same fact.

As we will see in the continuation of the analytical results of the article, the accuracy of none of the methods is perfect; therefore, there are probably a few images in which the results of both the basic and proposed methods are unacceptable. We see examples of these images in samples (r-s). Therefore, in both of these examples, the measurement error of the environment surrounding the tissue in both methods was a two-digit number, and visually, the drawn contour is completely inconsistent with the ground truth boundary.

In this way, the visual results and the difference parameter between the estimation of the circumference of the fetal head and the benchmark value clearly indicate the superiority of the proposed method. Therefore, the obtained results briefly indicate that the performance of the proposed method in extracting the border around the head circumference of the fetus was either better than the proposed method, or at least in a similar quality. Thus, in the discussion section, the same topic is shown with more diverse parameters and based on quantitative analysis on the entire images.

Discussion

In the previous section, the results obtained from testing the proposed and alternative methods on the fetal ultrasound image dataset were compared. However, this comparison was reported visually and only on a limited number of images. In this section, the comparisons are reported numerically and based on the total results of the test process, which gives a more comprehensive view of the effectiveness of the proposed method. However, in order to make such a comparison, it is necessary to use the standard parameters that are used in the field of fetal detection in ultrasound images, as described below. These parameters include Dice Coefficient, Difference, Absolute Difference, Jaccard index, precision and accuracy.

The Dice Similarity Coefficient (DSC) is a comparison parameter that measures the spatial overlap between the ground truth segmentation (GT) and the predicted segmentation(P). It is calculated using the formula shown in Equation 12. The DSC value ranges from 0 to 1, with 1 indicating perfect overlap between the ground truth and prediction. It provides a quantitative assessment of how well the predicted segmentation matches the actual segmentation. A higher DSC value indicates better performance of the segmentation method.

To quantify the comparison between the results obtained by each examined method and the ground truth, the Absolute Difference Factor (ADF) was calculated. The ADF represents the absolute difference between the predicted fetal head circumference (HC) and the real/ground truth HC, expressed as a positive value in millimeters. The ADF is calculated using the formula shown in equation 13, where HC(GT) is the ground truth fetal head circumference and HC(P) is the predicted fetal head circumference. The ADF provides a measure of the deviation between the predicted and actual values, with lower ADF indicating better performance of the examined methods.

Since in several studies in this field, DF and ADF values were usually reported separately, in this research, despite the semantic connection of these two parameters, they were calculated

and reported separately. The Jaccard index is a statistic used to measure the similarity and diversity between two sample sets. It quantifies the similarity between finite sample sets, such as sets X and Y, as demonstrated in equation 14.

The Jaccard index ranges from 0 to 1, where 0 indicates that the two sets have no overlap (completely dissimilar), and 1 indicates that the two sets are identical (completely similar). The Jaccard index provides a way to numerically evaluate the degree of overlap or similarity between two finite sample sets.

Accuracy is a metric that calculates the percentage of predicted values that match the actual values for binary labels. Since the labels in the images of this article are binary, accuracy is expressed as the probability of the predictions being equal to 1, as shown in equation 15.

In this article, TP (True Positive) represents the count of pixels in a labeled image that belongs to an actual HC segment, and the baseline or proposed method correctly predicts them as an HC segment. TN (True Negative) represents the count of pixels in a labeled image that belongs to an actual non-HC segment, and the examined methods correctly predict them as a non-HC segment. FP (False Positive) refers to the number of pixels in a labeled image that belongs to a non-HC segment, but the examined methods determine them as an HC segment. FN (False Negative) indicates the pixels of HC segments that both methods failed to detect. The accuracy metric provides an overall measure of how well the prediction methods match the ground truth labels for the binary classification task.

In equation 16, precision is a metric that measures the proportion of true positive predictions out of all the positive predictions made by the model. It indicates how accurate the model is at identifying the positive class.

A high precision value means that the model is making very few false positive predictions, i.e., it is correctly identifying the positive instances most of the time. Precision is calculated as the ratio of true positives to the sum of true positives and false positives, and is a useful evaluation metric when the cost of false positives is high.

The tests of this section were also performed in two categories. In the first type of tests, the performance of the proposed method based on data augmentation was compared with the performance of the basic method in which data augmentation concept has not been utilized.

In this comparison, all the elements for two methods were the same, except that the innovation considered in this article occurred in one of them (the proposed method) and did not happen in another (i.e., basic method). In this way, the difference between the results of these two algorithms may be clearly considered as an indicator of the effectiveness of the proposed method of this research. The results of this scenario are shown in Table 2. The Dice value for the basic case is 94.37%, while it has been obtained equal to 97.61% for the proposed scheme; therefore, from the point of view of Dice parameter, 3.24% improvement can be considered for the method of this article.

A similar improvement is also observed in the parameters of ADF and DF, by extents of 4.17 and 4.3 millimeters, respectively. Another noteworthy point is the significant reduction of the tolerance of these two parameters in the proposed method compared to the basic method, which is considered an important achievement for the innovation of this article.

It may be observed that for the three parameters of Dice, DF and ADF, the variance of the results has decreased by 7.34%, 10.89mm and 11.2mm, which means that the proposed method has gained more reliability and reproducibility, in addition to being better than the basic method. Finally, the table shows that the proposed method led to improvements of orders 5%, 1%, and 3.4% in the Jaccard, accuracy, and precision parameters. Due to the shuffling, each training

Table 2: Comparison of DSC, DF and ADF parameters obtained for entire dataset belonging to the proposed and basic schemes

Method	The Basic Method	The Proposed Method
DSC (%)	94.37±9.29	97.61±1.95
DF (mm)	5.12±13.55	0.82±2.66
ADF (mm)	6.23±13.07	2.06±1.87
Jaccard Index (%)	90.3	95.3
Precision (%)	94.2	97.6
Accuracy (%)	97.5	98.7

Table 3: Comparison of DSC and ADF parameters obtained for entire dataset by applying the proposed method and some state of art alternatives

Method	DSC (%)	ADF (mm)
Proposed Method	97.61±1.95	2.06±1.87
Heuvel et. al. (7)	97.00±2.8	2.8±3.3
Sobhaninia et. al. (16)	96.84±2.89	2.12±1.87
Sobhaninia et. al. (17)	93.75	2.27

sample with specific augmentation happens only once with a high chance. Therefore, the variety of the training samples increases and the model can be better generalized. In the second type of tests, the performance of the proposed method is compared with three well-known methods in the field of intelligent fetal head extraction in ultrasound images. The results of this scenario shown in Table 3 indicate that the proposed method has a slight advantage even over its nearest alternative [i.e., (7)] and more considerable gain against other state-of-art schemes.

By comparing the results of the proposed method with the results obtained in (7), it may be observed that although the improvement in Dice parameter has not been so high (about 0.6%), the ADF parameter has been promoted significantly (i.e., approximately 0.74 mm) by using proposed method.

However, the reduction of tolerance results from 2.8% to 1.95% in measuring the dice parameter and from 3.3 mm to 1.87 mm in the ADF parameter is quite significant and indicates the improvement of the performance of the proposed method. In the case of the other two alternatives [i.e., (16, 17)], the superiority of the proposed method is much more noticeable, so that the advantage in the dice parameter has reached up to approximately 4% and in the ADF parameter up to 0.74 mm.

Conclusion

In this study, a new method was proposed to improve the automatic measurement of the fetal head circumference in ultrasound images. In this method, first, to address the challenge of the sensitivity of deep neural networks to the volume of training data, unsupervised data augmentation scheme was used, and then the VNET neural network was trained with these data. Usually, the region identified as the fetal head by the neural network includes two categories of missing and extra subregions, both of which cause errors in diameter or circumference measurements. In this research, by using the method of estimating the optimal ellipse contour, there was an attempt to minimize the amount of both of these error areas and thus estimate a more accurate circumference for the fetal head.

The results of testing the proposed method on real data and comparing it with existing alternatives clearly indicated the effectiveness of both ideas. In first scenario, the Dice parameter belonging to the head circumference estimated by the proposed scheme was about 3.2% higher than that obtained from basic method. In the same way, the proposed augmentation algorithm could improve the Jaccard, precision and accuracy parameters in a range between 1% to 5% compared to the method based on only the original data. In the second scenario, it was observed that compared to the methods published in the articles of this field, the proposed method was able to improve the dice parameter in the range of approximately [0.6 3.9] percent and the DF parameter in the range of [0.06 0.74] millimeters. Finally in both scenarios, the variances of the results of the proposed method were significantly lower than alternatives, which indicates that this method provides more reliability in addition to better performance.

Although the proposed method has been able to perform better than the basic and alternative methods in obtaining performance parameters, there are still limitations that may be solved in future researches. One of these cases is the increase in neural network training time due to the increase in data. Another limitation is that augmentation by the mentioned method does not necessarily lead to the best amount of diversity in the data. It seems that if other methods such as Generative adversarial network are used to generate fake data, it may be possible to obtain more diversity in the training data and thus improve the quality of the results.

Furthermore, utilizing more advanced neural network architectures, such as multi-scale networks or transformer-based networks, may improve the results because these methods will have a higher potential in modeling non-linear

data in the problem of fetal head circumference estimation.

Acknowledgement
All authors declare that the present article has not been published in whole or in part elsewhere and is not currently being considered for publication in another journal. Furthermore, authors have been personally and actively involved in substantive work leading to the manuscript, and will hold themselves jointly and individually responsible for its content.

Authors' Contribution
Seyed Vahab Shojaedini: conceptualization, data curation, supervision. Amir Saniyan: data curation, simulations and tests. Mohammad Reza Riahi: drafting the manuscript and primary writing. Mahsa Monajemi: review, and final editing.

Ethics Approval
All ethical principles were considered in this article.

Funding
There was no funding.

Consent for Publication
Not applicable.

Conflict of Interest
There are no conflicts of interest

References
1. Loughna P, Chitty L, Evans T, Chudleigh T. Fetal size and dating: charts recommended for clinical obstetric practice. *Ultrasound*. 2009;17(3):160-6.
2. Murthy BR. Imaging of Fetal Brain and Spine: An Atlas and Guide: Springer; 2019.
3. Poojari VG, Jose A, Pai MV. Sonographic Estimation of the Fetal Head Circumference: Accuracy and Factors Affecting the Error. *J Obstet Gynaecol India*. 2022;72(Suppl 1):134-8. doi: 10.1007/s13224-021-01574-y.
4. Lu W, Tan J, Floyd R. Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform. *Ultrasound Med Biol*. 2005;31(7):929-36. doi: 10.1016/j.ultrasmedbio.2005.04.002.
5. Jardim SM, Figueiredo MA. Segmentation of fetal ultrasound images. *Ultrasound Med Biol*. 2005;31(2):243-50. doi: 10.1016/j.ultrasmedbio.2004.11.003.
6. Wu L, Cheng JZ, Li S, Lei B, Wang T, Ni D. FUIQA: Fetal Ultrasound Image Quality Assessment With Deep Convolutional Networks. *IEEE Trans Cybern*. 2017;47(5):1336-49. doi: 10.1109/TCYB.2017.2671898.
7. van den Heuvel TLA, de Bruijn D, de Korte CL, Ginneken BV. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS One*. 2018;13(8):e0200412. doi: 10.1371/journal.pone.0200412.
8. Zhang L, Ye X, Lambrou T, Duan W, Allinson N, Dudley NJ. A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2D ultrasound images. *Phys Med Biol*. 2016;61(3):1095-115. doi: 10.1088/0031-9155/61/3/1095.
9. Perez-Gonzalez J, Muńoz JB, Porras MR, Arámbula-Cosío F, Medina-Bańuelos V, editors. 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 29, 30 & 31 October 2014; 2015.
10. Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, et al. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. *IEEE Trans Med Imaging*. 2014;33(4):797-813. doi: 10.1109/TMI.2013.2276943.
11. Ni D, Yang Y, Li S, Qin J, Ouyang S, Wang T, et al., editors. 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; 2013: p. 772-5.
12. Jatmiko W, Habibie I, Ma'sum MA, Rahmatullah R, Satwika IP. Automated telehealth system for fetal growth detection and approximation of ultrasound images. *International Journal on Smart Sensing and Intelligent Systems*. 2015;8(1):697-719.
13. Sundaresan V, Bridge CP, Ioannou C, Noble JA, editors. Automated characterization of the fetal heart in ultrasound images using fully

convolutional neural networks. 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017); 2017. p. 671-4.
14. Wu L, Xin Y, Li S, Wang T, Heng P-A, Ni D, editors. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017); 2017. p. 663-6.
15. Sinclair M, Baumgartner CF, Matthew J, Bai W, Martinez JC, Li Y, et al. Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018;2018:714-7. doi: 10.1109/EMBC.2018.8512278.
16. Sobhaninia Z, Rafiei S, Emami A, Karimi N, Najarian K, Samavi S, et al. Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:6545-8. doi: 10.1109/EMBC.2019.8856981.
17. Sobhaninia Z, Emami A, Karimi N, Samavi S, editors. Localization of fetal head in ultrasound images by multiscale view and deep neural networks. 2020 25th International Computer Conference, Computer Society of Iran (CSICC); 2020. p. 1-5.
18. Fitzgibbon AW, Fisher RB. A buyer's guide to conic fitting: Citeseer; 1996. p 513-5.