Fetal Head Circumference Prediction Using Machine Learning

Nguyen Minh Dat
Machine Learning in Medicine
USTH University
Email: datnm.23bi14089@usth.edu.vn

***Abstract*—This study proposes an automated approach for fetal head circumference (HC) estimation using 2D ultrasound images and a U-Net–based segmentation model. The model is trained to segment the fetal head region, after which the head contour is extracted to compute HC. Experimental results show stable training behavior and reasonable segmentation performance; however, inaccuracies in contour prediction still affect HC estimation accuracy. These findings suggest that further improvements are needed to enhance segmentation quality and overall prediction reliability.**

I. Introduction

During pregnancy, monitoring fetal development plays a crucial role in ensuring fetal health. Head Circumference (HC) is a key biometric measurement used to assess whether the fetal head size is within the normal range for a given gestational age, and it is considered an essential parameter in prenatal ultrasound examinations.

With the rapid development of modern computational techniques, advanced scientific models have been increasingly adopted to assist clinicians in achieving faster and more accurate diagnoses. In this context, this study focuses on applying machine learning models to predict fetal head circumference, aiming to support medical decision-making and improve diagnostic efficiency.

II. Data Description

The dataset used in this study is entitled *Automated Measurement of Fetal Head Circumference Using 2D Ultrasound Images*. It consists of two subsets: a training set and a testing set. The training set contains 999 fetal ultrasound images along with 999 corresponding annotation images, while the testing set includes 355 ultrasound images. All ultrasound images are provided in grayscale format with a fixed spatial resolution. The annotation images provide ground-truth information required for supervised learning.

In addition to the image data, the dataset provides several associated features that are used for head circumference (HC) measurement and evaluation. These features are summarized in Table I.

TABLE I
Description of dataset features

Feature	Description
Filename	Name of the ultrasound image file corresponding to each sample.
Pixel size (mm)	Physical size represented by each pixel in millimeters, used to convert pixel-based measurements to real-world units.
Head Circumference (mm)	Ground-truth fetal head circumference measured in millimeters, provided by expert annotation.

III. Methodology

A. Data Preparation

Since the image directory contains both ultrasound images and corresponding annotation images, they were first separated into two individual folders before the training process.

Image Preprocessing: All ultrasound images and annotation masks were resized to a fixed resolution of 256×256 pixels to ensure consistent input dimensions for the model. The ultrasound images were normalized to the range $[0, 1]$ to stabilize the training process.

Mask Preprocessing: The annotation images were converted into binary masks, where pixel values of 0 represent the background, and 1 represent the fetal head region.

B. Annotation Representation

In this study, two types of annotation representations were used. The first type is contour-based annotations, which provide only the boundary of the fetal head. The second type is a filled binary mask, where pixels inside the head contour are labeled as foreground, and pixels outside are labeled as background.

C. Model Selection

The U-Net architecture was selected due to its strong performance in medical image segmentation tasks, particularly in scenarios that require accurate boundary and shape detection. In the context of fetal head circumference measurement, precise extraction of the head region from ultrasound images is a critical step.

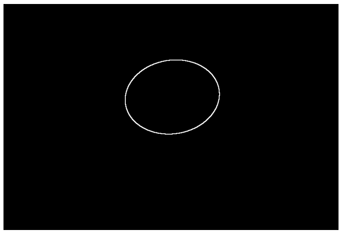

Fig. 1. contour-based annotations

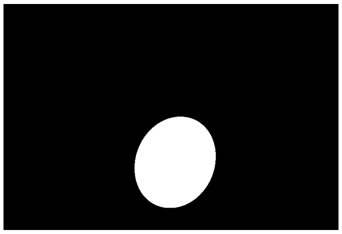

Fig. 2. filled binary mask

U-Net effectively captures global contextual information while preserving fine-grained spatial details through skip connections between the encoder and decoder, making it well-suited for segmenting the fetal head region at the pixel level.

IV. Results

A. Traning 1: contour-based annotations

For the first training experiment, the original annotations provided by the dataset were used. After 50 training epochs, the model loss gradually decreased from 1.29 at epoch 1 to 0.26 at epoch 50.

The trained model was then used to predict the head circumference (HC), which was calculated using the following formula:

$$HC_{\text{mm}} = \text{Perimeter}_{\text{pixel}} \times \text{PixelSize}_{\text{mm/pixel}}$$

TABLE II
Predicted head circumference on the test set

Filename	**Pixel size (mm/pixel)**	**HC$_{pred}$ (mm)**
000_HC.png	0.2359	265.60
001_HC.png	0.0680	8.74
002_HC.png	0.1652	82.31
003_HC.png	0.0967	134.44
004_HC.png	0.2022	292.58
005_HC.png	0.1167	155.66

Based on the visual results and quantitative evaluation, the predicted masks are not fully closed, which leads to incomplete head region extraction. As a result, the estimated head circumference values are not yet optimal or fully accurate.

B. Traning 2: filled binary mask

Overall, the model shows effective learning behavior over time, with a steady decrease in training loss and consistent improvements in training Dice and IoU scores. The validation results also demonstrate a clear improvement trend toward later epochs, particularly after around epoch 14. However, some instability is observed at early stages, with sudden increases in validation loss at epochs 3 and 8, indicating that the training process was not fully stable at the beginning.

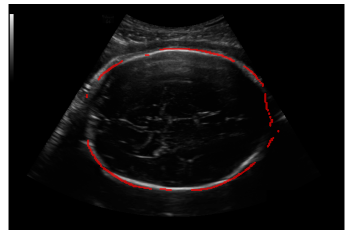

Fig. 3. prediction 1

TABLE III
Training and validation performance over epochs

Epoch	Train Loss	Train Dice	Train IoU	Val Loss	Val Dice	Val IoU
1	11.01	0.405	0.280	0.979	0.582	0.414
2	0.834	0.673	0.511	0.839	0.686	0.528
3	0.765	0.701	0.544	11.75	0.418	0.266
4	0.730	0.719	0.566	0.965	0.521	0.355
5	0.682	0.740	0.591	0.761	0.687	0.528
6	0.651	0.752	0.607	0.668	0.744	0.599
7	0.615	0.768	0.628	0.771	0.685	0.526
8	0.599	0.774	0.636	1.262	0.369	0.229
9	0.589	0.779	0.642	0.600	0.776	0.641
10	0.561	0.789	0.657	0.730	0.681	0.521
11	0.544	0.795	0.664	0.624	0.774	0.638
12	0.518	0.806	0.679	0.914	0.632	0.466
13	0.509	0.811	0.685	0.633	0.777	0.642
14	0.496	0.815	0.693	0.573	0.792	0.663
15	0.479	0.822	0.701	0.551	0.780	0.643
16	0.476	0.823	0.702	0.510	0.808	0.684
17	0.459	0.829	0.713	0.476	0.820	0.700
18	0.456	0.830	0.713	0.497	0.812	0.688
19	0.431	0.841	0.729	0.570	0.798	0.671
20	0.441	0.836	0.723	0.460	**0.825**	**0.707**

Despite this, the final performance is strong, achieving a validation Dice score of approximately 0.82 and a validation IoU of around 0.71 at epoch 20, which indicates good segmentation quality. The gap between training and validation metrics remains acceptable, suggesting no severe overfitting. Additionally, the model continues to improve until the final epochs without early plateauing, reflecting a good generalization capability.

In this training run, the model is able to produce closed contours. However, it still over-segments certain regions and does not fully match the ground truth annotations from the training set.In this training run, the model is able to produce

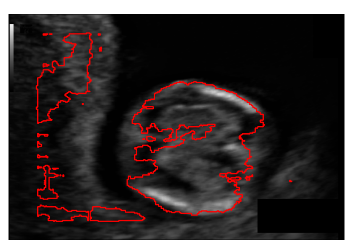

Fig. 4. prediction 2

TABLE IV

PREDICTED HEAD CIRCUMFERENCE (HC) MEASUREMENTS

Filename	Pixel size (mm)	HC_{pred} (mm)
000_HC.png	0.2359	587.33
001_HC.png	0.0680	173.98
002_HC.png	0.1652	379.95
003_HC.png	0.0967	286.40
004_HC.png	0.2022	618.43
005_HC.png	0.1167	295.51

closed contours. However, it still over-segments certain regions and does not fully match the ground truth annotations from the training set.

V. CONCLUSION

Overall, the training process in each experiment was relatively smooth. However, based on the evaluation on the test set, the results are not yet satisfactory. In the first training run, the predicted contours generally followed the expected shape but contained broken segments. In the second training run, the predicted contours were less accurate, which led to unreasonable head circumference estimation results.Overall, the training process in each experiment was relatively smooth. However, based on the evaluation on the test set, the results are not yet satisfactory. In the first training run, the predicted contours generally followed the expected shape but contained broken segments. In the second training run, the predicted contours were less accurate, which led to unreasonable head circumference estimation results.