Practice 2: A Study of Deep Learning for Measurement of The Fetal Head Circumference

Hoang Khanh Dong - 22BA13072

January 2026

Abstract—**Accurate measurement of fetal head circumference (HC) from ultrasound images is essential for monitoring fetal growth and estimating gestational age during pregnancy. Manual measurement by sonographers is time-consuming and subject to inter-observer variability. This study presents a deep learning-based approach for automated HC measurement using 2D ultrasound images from the HC18 Grand Challenge dataset**

I. INTRODUCTION

My approach contains two main steps. First, semantic segmentation is used to delineate the fetal head boundary, followed by an ellipse fitting algorithm to compute the circumference. The methodology demonstrates the potential of deep learning techniques to assist clinicians in obtaining accurate and consistent fetal biometric measurements.

II. DATASET

The dataset is from the HC18 Grand Challenge dataset, which contains 1334 2D ultrasound images (999 training, 335 testing) with pixel sizes from 0.052 to 0.326mm. The images are of variable size and quality, with different levels of noise and artifacts.

In the training set, there is 975 over 999 (97.6%) images with the size of 800x540 pixels, and the rest of the images are of different sizes.

In the test set, there is 327 over 335 (97.6%) images with the size of 800x540 pixels, and the rest of the images are also of different sizes.

Fig. 1. Example of a 2D ultrasound image from the HC18 dataset.

III. METHODOLOGY

A. Data Preprocessing

The HC18 dataset provides ellipse annotations as outline images rather than filled segmentation masks. To prepare the ground truth for training, I implemented a label extraction pipeline.

First, the annotation image is loaded in grayscale and binarized using a threshold of 127. Next, contour detection is applied using the external contour retrieval mode to extract the ellipse boundary. Finally, the detected contour is filled to create a solid binary mask where pixels inside the fetal head region are set to 1 and background pixels to 0.

This filled mask serves as the ground truth segmentation label for training the neural network.
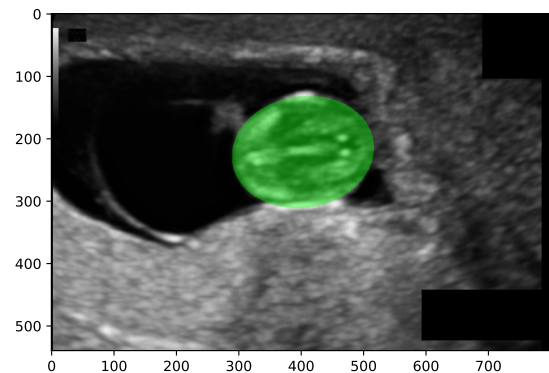

Fig. 2. Overlay of the ultrasound image and the ground truth segmentation mask.

B. Semantic Segmentation

1) Model Architecture: My model takes the idea from the CU-Net architecture (Cascade U-Net) which is a convolutional neural network that is widely used for semantic segmentation tasks. The model is composed of two encoder-decoder paths, which is slightly modified from the original architecture. A visualization of the model architecture is shown in Figure 3.
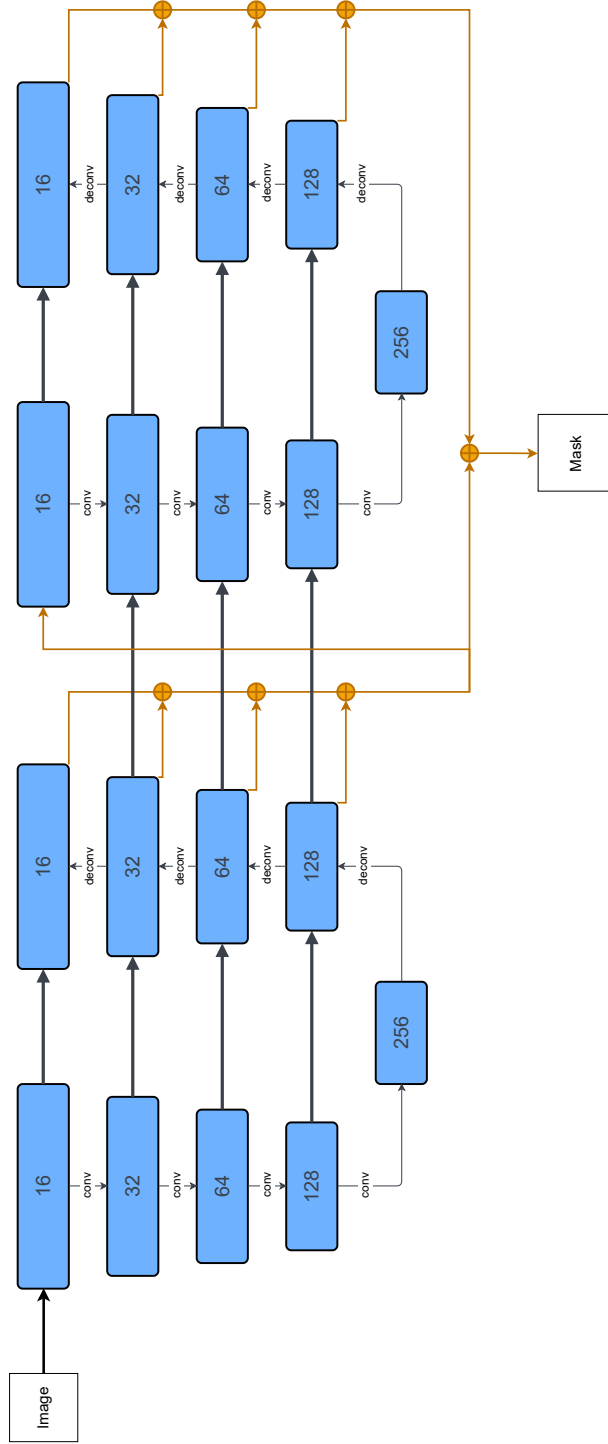
DoubleConv blocks, where each block contains two 3×3 convolutions with Batch Normalization and ReLU activation, plus a residual shortcut connection for improved gradient flow. The number of channels progressively increases from 16 to 32, then to 64, 128, and finally 256 at the bottleneck. Each downsampling stage applies Max Pooling with a kernel size of 2.

The decoder path mirrors the encoder with four upsampling stages using transposed convolutions, and skip connections that concatenate encoder features with decoder features at each level. Crucially, the first U-Net employs deep supervision, where the outputs from all decoder levels are projected to a single channel using 1×1 convolutions, upsampled to the original resolution using bilinear interpolation, and summed together.

The second U-Net refines the initial segmentation by receiving the output from the first U-Net. It additionally incorporates skip connections from the decoder features of the first U-Net, allowing the second network to leverage multi-scale features from the first stage. The second U-Net follows the same structure with deep supervision. The final prediction is obtained by fusing the outputs from both U-Net branches, enabling the model to capture both coarse and fine-grained features for accurate boundary delineation.

Fig. 3. Model Architecture

2) Network Components: The model begins with an input layer accepting grayscale ultrasound images of shape (batch_size, 1, 256, 256). The first U-Net performs initial feature extraction through an encoder-decoder structure. The encoder path consists of four downsampling stages using

C. Training Strategy

The model was trained using the Adam optimizer with a learning rate of $1×10^{-4}$ and Binary Cross-Entropy with Logits Loss as the objective function. I employed a batch size of 4 with 100 epochs. The dataset was split into 90% for training (999 images) and 10% for validation (99 images) using a fixed random seed for reproducibility.

A key part of the strategy was the use of a ReduceLROn-Plateau learning rate scheduler, which monitors the validation Dice score and reduces the learning rate by a factor of 0.5 if no improvement is observed for 10 consecutive epochs. This allows the model to escape local minima and continue learning. The best model was automatically saved based on the highest validation Dice score during training.

For evaluation metrics, I used Dice coefficient and Intersection over Union (IoU), which are standard metrics for segmentation tasks. Images were resized to 256×256 pixels and normalized to the range [0, 1] before feeding to the network.

Figure 4 illustrates the training and validation loss over 100 epochs. The model demonstrates rapid convergence within the first 20 epochs. While the training loss continues to decrease, the validation loss starts to increase from epoch 20, indicating that the model can be mild overfitting in the later stages.
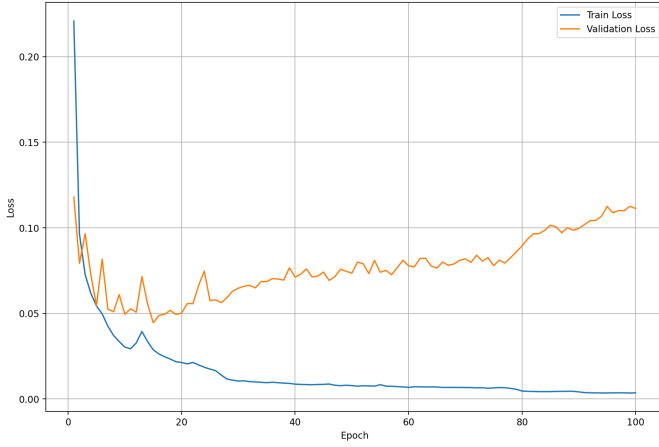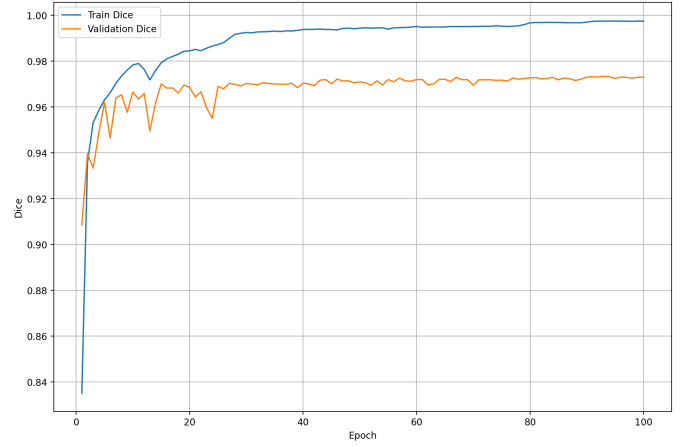
Fig. 4. Training and validation loss curves.

Fig. 5. Dice coefficient over training epochs.

D. Measurement of Fetal Head Circumference

After obtaining the segmentation mask from the CU-Net model, the head circumference is computed through a series of image processing and geometric fitting steps.

First, the predicted probability mask is binarized using a threshold of 0.5. Contour detection is then applied to extract the boundary of the segmented region. The largest contour is selected, assuming it corresponds to the fetal head. An ellipse is fitted to this contour using the least-squares method, which returns the center coordinates (x, y), the major axis diameter MA, the minor axis diameter ma, and the rotation angle.

The semi-major axis a and semi-minor axis b are computed as:

$$a = \frac{MA}{2}, \quad b = \frac{ma}{2} \tag{1}$$

The perimeter of an ellipse does not have a closed-form solution, so I use Ramanujan's second approximation:

$$h = \frac{(a-b)^2}{(a+b)^2} \tag{2}$$

$$P \approx \pi(a+b)\left(1 + \frac{3h}{10 + \sqrt{4 - 3h}}\right) \tag{3}$$

where P is the perimeter in pixels. Finally, the head circumference in millimeters is obtained by multiplying the perimeter by the pixel size provided in the dataset:

$$HC_{mm} = P \times \text{pixel\_size} \tag{4}$$

IV. RESULT

A. Semantic Segmentation

The CU-Net model achieved strong segmentation performance on the validation set. After 100 epochs of training, the best model reached a Dice coefficient of **0.9733** and an Intersection over Union (IoU) of **0.9480** at epoch 93.

Figure 5 and Figure 6 show the Dice and IoU curves during training. Both metrics rapidly improve within the first 20 epochs and stabilize at high values, consistent with the loss curves discussed earlier.
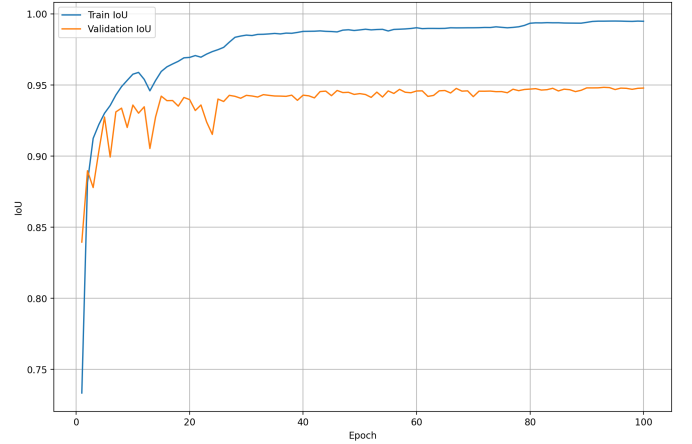

Fig. 6. IoU score over training epochs.

B. Measurement of Fetal Head Circumference

The head circumference was computed from the predicted segmentation masks using ellipse fitting and Ramanujan's perimeter approximation. On the validation set of 99 images, 100% of predictions were valid (all masks produced a contour suitable for ellipse fitting).

The regression analysis yielded excellent results, summarized in Table I. The model achieved a Mean Absolute Percentage Error of only 1.79%, indicating near-perfect linear correlation between predicted and ground truth measurements.

TABLE I
HEAD CIRCUMFERENCE MEASUREMENT RESULTS ON VALIDATION SET

Metric	Value
Total Samples	99
Valid Predictions	99 (100%)
Mean Absolute Error (MAE)	2.33 mm
Root Mean Squared Error (RMSE)	3.98 mm
Mean Absolute Percentage Error (MAPE)	1.79%
Root Mean Squared Percentage Error (RMSPE)	4.84%
R-squared (R^2)	0.9962

Figure 7 shows the scatter plot of predicted versus ground truth head circumference values. The points closely follow the ideal diagonal line, confirming the high accuracy of the measurement pipeline.
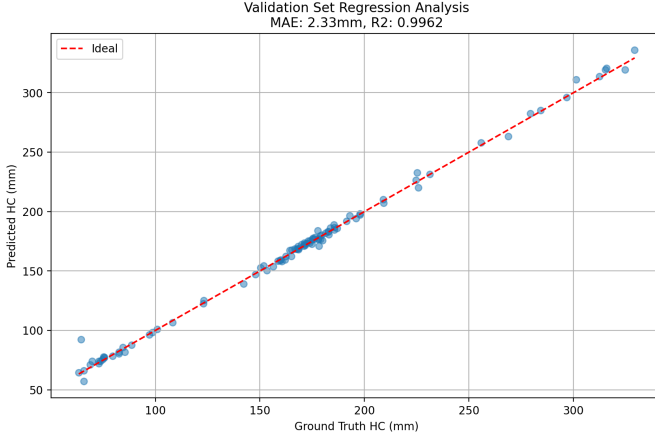

Fig. 7. Regression analysis of predicted vs. ground truth head circumference.

V. Discussion

The CU-Net model achieved a Dice of 0.9733, IoU of 0.9480, and HC measurement MAE of 2.33 mm ($R^2 = 0.9962$) on the validation set. Since the HC18 challenge no longer accepts submissions, these results cannot be directly compared to the official leaderboard.

For reference, the top methods on the HC18 leaderboard [1] are shown in Table II. The top solutions use nnU-Net and achieve MAE around 0.74 mm—better than my 2.33 mm. This gap may be attributed to differences in evaluation sets, their use of 1000 training epochs (vs. my 100), and nnU-Net's automated optimization pipeline.

TABLE II
TOP 3 RESULTS FROM HC18 GRAND CHALLENGE LEADERBOARD

Metric	1st (nnU-Net)	2nd (nnU-Net)	3rd (nnU-Net)
Created	8 June 2023	6 June 2023	10 June 2023
MAE (mm)	**0.74 ± 0.73**	**0.74 ± 0.73**	**0.81 ± 1.26**
MD (mm)	0.14 ± 1.03	0.14 ± 1.03	0.21 ± 1.49
DICE (%)	99.43 ± 0.63	99.43 ± 0.63	99.41 ± 0.73
HD (mm)	0.39 ± 0.40	0.39 ± 0.40	0.41 ± 0.56

VI. Conclusion

This study presented a CU-Net-based approach for automated fetal HC measurement, achieving strong validation performance (Dice: 0.9733, MAE: 2.33 mm). While not directly comparable to the HC18 leaderboard. Future work could explore nnU-Net, extended training, and ensemble methods to close the gap with state-of-the-art solutions.

---

[1] https://hc18.grand-challenge.org/evaluation/challenge/leaderboard/