Deep Learning–Based Regression for Fetal Head Circumference Estimation from Ultrasound Images

Nguyen Thi Lan Huong
Student ID: 23BI14191
University of Science and Technology of Hanoi (USTH)
Lecturer: Tran Giang Son
Email: huongntl.23bi14191@usth.edu.vn

***Abstract*—Fetal head circumference (HC) measurement is a fundamental task in prenatal ultrasound assessment for monitoring fetal growth. This paper presents a deep learning–based approach for automatic HC estimation from 2D ultrasound images using a convolutional neural network. A regression model based on a lightweight ResNet architecture is employed to directly predict HC values from grayscale ultrasound images. The proposed method is evaluated on the HC18 dataset and trained using the mean absolute error (MAE) loss function. Experimental results demonstrate that the model achieves stable and competitive regression performance, indicating the effectiveness of convolutional neural networks for automated fetal biometric estimation from ultrasound data.**

***Index Terms*—Fetal head circumference, Ultrasound imaging, Deep learning, Convolutional neural network, Regression, HC18 dataset**

I. Introduction

Fetal head circumference (HC) measurement is a fundamental task in prenatal care for assessing fetal growth and detecting potential abnormalities. Ultrasound imaging is a non-invasive and widely used modality for fetal biometric analysis. However, manual HC measurement from ultrasound images is time-consuming and subject to inter-observer variability, motivating the development of automated measurement methods.

Traditional approaches for HC estimation often rely on manual delineation or handcrafted image features, which require expert knowledge and may lack robustness to noise and image variability. In contrast, deep learning methods enable automatic feature extraction directly from raw ultrasound images, providing a data-driven solution for fetal biometric estimation.

In this work, we investigate a convolutional neural network–based regression approach for automatic fetal head circumference estimation from 2D ultrasound images. A lightweight CNN model is trained and evaluated on the HC18 dataset using the mean absolute error (MAE) metric. Experimental results demonstrate the feasibility of deep learning–based regression for reliable and efficient HC estimation from ultrasound data.

II. Dataset Description

The experiments were conducted on the HC18 dataset, which consists of 2D fetal ultrasound images acquired during routine prenatal examinations. Each image is annotated with a ground truth head circumference (HC) measurement, provided in millimeters. The annotations are stored in a CSV file containing the image filename, pixel size (in millimeters per pixel), and the corresponding HC value.

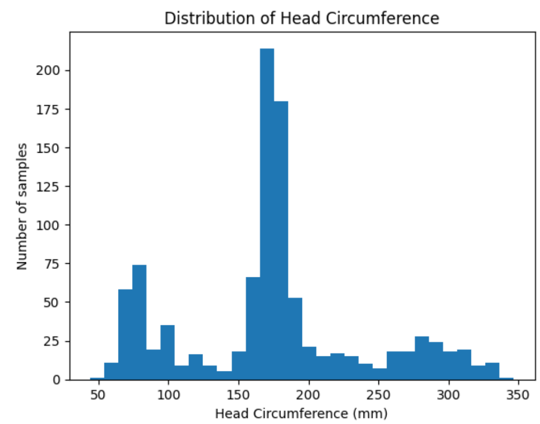

Fig. 1. Distribution of head circumference values in the HC18 dataset.

The dataset contains a total of 999 ultrasound images in. The head circumference values range from 44.3 mm to 346.4 mm, with a mean value of 174.38 mm, indicating a wide coverage of fetal growth stages. As illustrated in Fig. 1, the HC distribution is not uniform and exhibits multiple concentration regions, which increases the difficulty of the regression task and requires the model to generalize across both small and large fetal head sizes.

In addition to HC annotations, the dataset provides pixel size information for each image. The pixel size varies from 0.0494 to 0.3933 mm per pixel, with an average value of 0.1398 mm per pixel, as shown in Fig. 2. This variability reflects differences in ultrasound acquisition settings and imaging devices and contributes to the overall heterogeneity of the dataset.

Representative fetal ultrasound images from the HC18 dataset are shown in Fig. 3. The images are grayscale and exhibit typical ultrasound characteristics such as speckle noise, low contrast, and blurred anatomical boundaries. These factors make accurate localization of the fetal head contour challenging and highlight the need for robust feature extraction methods.

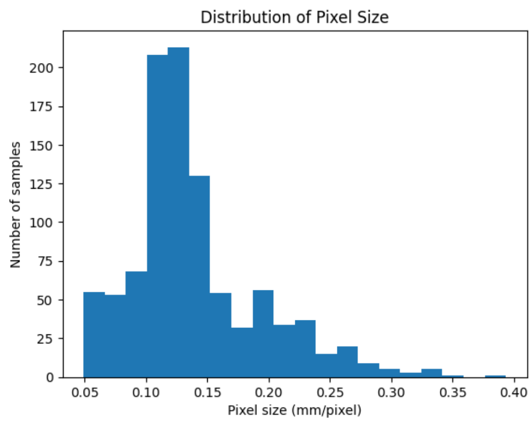

Fig. 2. Distribution of pixel size (mm per pixel) in the HC18 dataset.

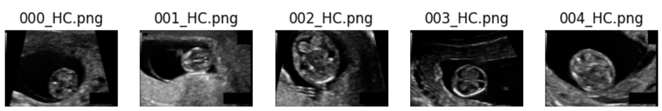

Fig. 3. Representative fetal ultrasound images from the HC18 dataset illustrating variations in fetal head appearance and imaging conditions.

For model development and evaluation, the dataset was randomly split into a training set and a validation set, with 80% of the samples used for training and 20% used for validation. The official HC18 test set does not provide ground truth HC annotations; therefore, an internal train–validation split was adopted to assess the generalization performance of the proposed regression model.

A. Test Set Description

The HC18 test set consists of 335 fetal ultrasound images provided without ground truth head circumference annotations. The pixel size values range from 0.0526 to 0.3713 mm per pixel, with a mean value of 0.1415 mm per pixel. The images have similar spatial resolutions, with heights ranging from 540 to 544 pixels and widths close to 800 pixels. The test set was used exclusively for inference after model training and validation, and no quantitative evaluation metrics were computed due to the absence of ground truth labels.

III. Methodology

This section describes the proposed deep learning–based regression framework for automatic fetal head circumference estimation from ultrasound images. The overall pipeline includes image preprocessing, model architecture, training strategy, and evaluation metrics.

A. Image Preprocessing

All ultrasound images in the HC18 dataset exhibit varying spatial resolutions. Since convolutional neural networks require fixed-size inputs, each image was converted to grayscale and resized to a resolution of 224×224 pixels. Pixel intensity values were normalized to the range $[0, 1]$ to stabilize training and accelerate convergence. No additional handcrafted feature extraction was applied, allowing the model to learn discriminative features directly from raw image data.

B. Model Architecture

A convolutional neural network based on the ResNet18 architecture was employed as the regression backbone. ResNet18 was selected due to its moderate depth and residual learning mechanism, which helps mitigate the vanishing gradient problem and enables stable training on noisy ultrasound images. The first convolutional layer was modified to accept single-channel grayscale images. The final fully connected layer was replaced with a single linear neuron to output a continuous head circumference value.

C. Training Strategy

The model was trained using a supervised regression setup. The mean absolute error (MAE) was adopted as the loss function, as it directly reflects the absolute difference between predicted and ground truth head circumference values measured in millimeters. Model optimization was performed using the Adam optimizer. The dataset was split into training and validation subsets using an 80:20 ratio to assess generalization performance. Multiple training configurations were evaluated by varying key hyperparameters such as learning rate, batch size, and number of training epochs.

D. Evaluation Metric

Model performance was evaluated using the mean absolute error (MAE), defined as

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| ,$$

where y_i and $\hat{y}_i$ denote the ground truth and predicted head circumference values, respectively, and N is the number of samples. MAE is expressed in millimeters and provides an intuitive measure of regression accuracy.

IV. Experimental Results

A. Hyperparameter Analysis

To investigate the impact of different hyperparameter configurations on regression performance, multiple experiments were conducted using the same ResNet18-based regression model. The learning rate, batch size, and number of training epochs were varied, while all other components of the training pipeline were kept unchanged. Model performance was evaluated using the best validation mean absolute error (MAE) achieved during training.

TABLE I
Hyperparameter experiments and best validation MAE.

Learning Rate	Batch Size	Epochs	Best Val MAE (mm)
1×10^{-3}	16	20	14.24
1×10^{-4}	16	30	34.04
1×10^{-4}	8	30	14.84

Table I summarizes the experimental results. The configuration with a learning rate of 1×10^{-3}, batch size of 16, and 20 training epochs achieved the lowest validation MAE of 14.24 mm. This indicates that a relatively larger learning rate enables faster convergence and more effective optimization for this task. In contrast, reducing the learning rate to 1×10^{-4} while increasing the number of training epochs to 30 resulted in a significantly higher validation MAE of 34.04 mm, suggesting underfitting within the given training budget.

When the batch size was reduced from 16 to 8 while maintaining the lower learning rate, the validation MAE improved to 14.84 mm. This improvement suggests that a smaller batch size can partially compensate for a lower learning rate by introducing additional gradient noise, leading to better generalization. However, this configuration still did not outperform the model trained with a learning rate of 1×10^{-3}. Overall, the results demonstrate that the learning rate plays a dominant role in determining model performance for fetal head circumference regression.

B. Comparison with HC18 Leaderboard

The HC18 dataset is associated with the HC18 Grand Challenge, where submitted methods are evaluated on a hidden test set and ranked on a public leaderboard. Most top-performing approaches on the leaderboard rely on multi-stage pipelines that include fetal head segmentation followed by ellipse fitting to estimate head circumference. These methods typically report substantially lower errors than direct regression approaches.

In contrast, the proposed method formulates head circumference estimation as a single-stage regression problem, directly predicting HC values from raw ultrasound images without explicit segmentation or geometric modeling. As a result, the achieved validation MAE of approximately 14 mm is higher than the errors reported by top leaderboard methods. However, it is important to note that the evaluation protocols are not directly comparable. The leaderboard results are obtained using specialized pipelines and challenge-specific metrics, whereas this work focuses on a simplified regression framework evaluated using MAE on an internal validation split.

Despite this performance gap, the regression-based approach offers advantages in terms of simplicity and computational efficiency. The results demonstrate that a lightweight convolutional neural network can learn meaningful representations from ultrasound images and provide reasonable HC estimates without requiring additional annotation such as segmentation masks. This makes the proposed approach suitable as a baseline model and highlights potential directions for future improvement, such as integrating segmentation-based supervision or multi-task learning strategies.

V. Discussion

The experimental results demonstrate that hyperparameter selection has a significant impact on the performance of deep learning–based regression for fetal head circumference estimation. Among the evaluated configurations, the learning rate was found to be the most influential factor. A learning rate of 1×10^{-3} enabled faster convergence and resulted in the lowest validation MAE, whereas a smaller learning rate led to suboptimal performance within the same training budget. This suggests that overly conservative optimization settings may hinder effective learning in this task.

The effect of batch size was also observed, particularly when training with a smaller learning rate. Reducing the batch size from 16 to 8 improved validation performance, indicating that increased gradient variability can enhance generalization. However, this improvement was limited and did not surpass the performance achieved with a higher learning rate, further highlighting the dominant role of learning rate selection.

When compared with the HC18 challenge leaderboard, the proposed regression-based approach exhibits higher error values. This performance gap can be attributed to fundamental differences in methodology. Most leaderboard methods employ multi-stage pipelines involving fetal head segmentation and ellipse fitting, which explicitly model head geometry and benefit from additional supervision. In contrast, the proposed method directly predicts head circumference values from raw ultrasound images without explicit anatomical modeling. As a result, the two approaches are not directly comparable in terms of evaluation protocol and complexity.

Despite its relatively higher error, the proposed method offers several advantages. It provides a simple, end-to-end learning framework that does not require pixel-level annotations or complex post-processing steps. This simplicity reduces annotation cost and computational overhead, making the approach suitable as a baseline model and for rapid prototyping. The results indicate that convolutional neural networks are capable of learning meaningful representations from noisy ultrasound images, even under a simplified regression formulation.

Overall, the findings suggest that while direct regression is less accurate than segmentation-based methods for head circumference estimation, it remains a viable and efficient baseline. Future improvements may be achieved by incorporating multi-task learning, integrating anatomical priors, or combining regression with segmentation-based supervision.

VI. Conclusion

This paper presented a deep learning–based regression approach for automatic fetal head circumference estimation from 2D ultrasound images. A lightweight ResNet18-based convolutional neural network was employed to directly predict head circumference values from grayscale ultrasound images. The proposed method was evaluated on the HC18 dataset using the mean absolute error (MAE) metric. Experimental results demonstrated that appropriate hyperparameter selection, particularly the learning rate, plays a crucial role in achieving stable and reliable regression performance. Although the proposed approach does not reach the accuracy of segmentation-based methods reported on the HC18 leaderboard, it provides a simple and efficient baseline for fetal biometric estimation without requiring additional annotations.

VII. Future Work

Several directions can be explored to further improve the performance of the proposed approach. First, incorporating segmentation-based supervision or multi-task learning could help the model better capture anatomical structures and reduce prediction error. Second, leveraging pixel size information more explicitly during training may improve robustness across different imaging resolutions. Additionally, advanced data augmentation techniques and larger training datasets could enhance model generalization. Finally, extending the framework to jointly estimate multiple fetal biometric parameters may further increase its clinical applicability.

VIII. References

[1] HC18 Grand Challenge, "HC18 Grand Challenge: Automatic Measurement of Fetal Head Circumference from Ultrasound Images," Grand Challenge in Medical Image Analysis, 2018.

[2] T. L. A. van den Heuvel, D. de Bruijn, C. L. de Korte, and B. van Ginneken, "Automated measurement of fetal head circumference using 2D ultrasound images," HC18 dataset, Zenodo, 2018. doi:10.5281/zenodo.1327317.

[3] M. Fiorentino *et al.*, "Automatic measurement of fetal biometric parameters in ultrasound images," *Computer Methods and Programs in Biomedicine*, 2020.