

Automated measurement of fetal head circumference using 2D ultrasound images

Nguyen Hoang Tung

Student ID: 22BA13318

University of Science and Technology of Ha Noi (USTH)

Email: tungtimo0808@gmail.com

I. EXPLORATORY DATA ANALYSIS

A. Dataset Overview

1) *Training Set*: The training set is provided in the file `training_set_pixel_size_and_HC.csv`. It contains 999 ultrasound images. Each image is associated with the following attributes:

- **filename**: the name of the ultrasound image file.
- **pixel size (mm)**: the physical size of one pixel in millimeters, which depends on the ultrasound machine.
- **head circumference (mm)**: the ground-truth head circumference value.

The head circumference values are obtained from manual annotation. An ellipse is fitted to the fetal head boundary.

2) *Test Set*: The test set is provided in the file `test_set_pixel_size.csv`. It contains 335 ultrasound images. Each sample includes:

- **filename**: the name of the ultrasound image file.
- **pixel size (mm)**: the physical size of one pixel in millimeters.

The test set does not contain the head circumference values. These values are predicted by the trained model.

B. Class Distribution

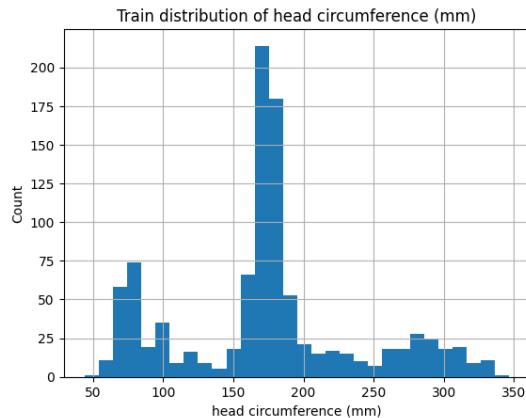


Fig. 1. Distribution of head circumference in the training set

This figure shows that the head circumference values in the training set cover a wide range, from about 50 mm to more than 340 mm. This indicates that the dataset includes

fetuses at many different development stages. The histogram suggests that the data may come from different gestational age groups. This wide distribution makes the regression task more challenging.

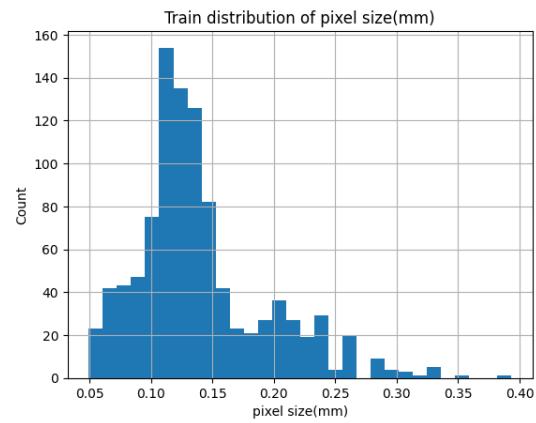


Fig. 2. Distribution of pixel size in the training set

We can observe that most pixel size values in the training set are also between 0.08 mm and 0.16 mm. This distribution is very similar to the one in the test set. This means that the training and test data were collected under similar imaging conditions. However, there are still a few samples with larger pixel sizes, which can be considered as outliers.

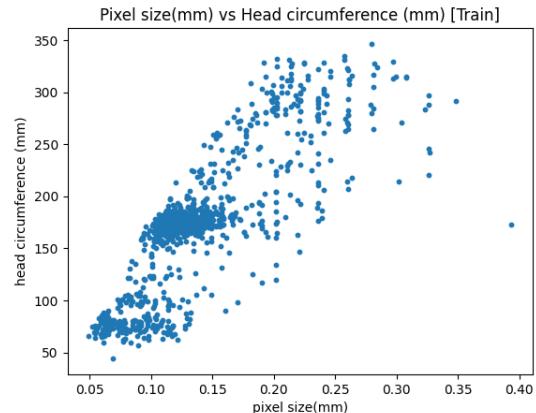


Fig. 3. Relationship between pixel size and head circumference in the training set

This scatter plot shows a clear positive relationship between pixel size and head circumference. When the pixel size increases, the head circumference also tends to increase. However, the points are widely spread and do not follow a single straight line. This means that the relationship is not perfectly linear. Therefore, a simple linear model may not be enough to solve this problem.

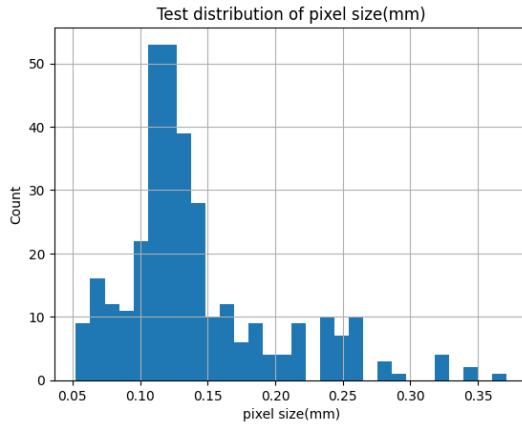


Fig. 4. Distribution of pixel size in the test set

From this figure, we can see that most pixel size values in the test set are between 0.08 mm and 0.16 mm. This means that most test images were captured with similar zoom levels. A small number of samples have much larger pixel sizes, which indicates that a few images were acquired with different settings. Overall, the distribution is not uniform and contains some outliers.

C. Conclusion and Insights from EDA

From the exploratory data analysis:

- First, the head circumference values in the training set cover a wide range, from very small to very large values. This indicates that the dataset includes fetuses at many different development stages. Because of the regression task is not simple and requires a model that can handle large variations.
- Second, the analysis of pixel size shows that most images use similar scales, but there are still some outliers with much larger pixel sizes. These samples may make the learning process more difficult and should be handled carefully.
- Finally, the scatter plot between pixel size and head circumference shows a clear positive trend. It proves that linear model may not be sufficient. A more powerful model, such as a deep neural network. It is needed to learn the complex mapping from the input image and pixel size to the head circumference value.

Overall, the dataset is diverse and realistic. It is suitable for training and evaluating a regression model for fetal head circumference estimation.

D. Model Implementation

- 1) *Data Loading:* Two CSV files are used in this work:

- The training file contains image filenames and ground-truth head circumference (HC) values.
- The test file contains image filenames only.

Two image folders are defined: `training_set/` and `test_set/`. The filename column and the HC column are automatically detected from the CSV headers to make the implementation more robust to minor changes in column names.

- 2) *Image Preprocessing:* Each ultrasound image is processed as follows:

- The image is loaded in grayscale.
- The image is resized to 224×224 .
- Pixel values are normalized to the range $[0, 1]$.
- The data are reshaped to $(N, 224, 224, 1)$.

If a filename does not match an existing file, multiple filename formats and extensions are checked. Samples with missing images or missing HC values are removed from the dataset.

- 3) *Train–Validation Split:* The training set is split into training and validation subsets using an 80/20 ratio. A fixed random seed is used to ensure reproducible results.

- 4) *Network Architecture:* A ResNet50 model pre-trained on ImageNet is used as the backbone network. The classification head is removed by setting `include_top=False`. The input size of the backbone is $(224, 224, 3)$.

Since the input images are grayscale, they are converted to RGB using `grayscale_to_rgb` before being passed to the network. The ResNet preprocessing function is then applied.

The regression head consists of:

- A Global Average Pooling layer.
- A fully connected layer with 256 units and ReLU activation.
- A Dropout layer with rate 0.3.
- A final fully connected layer with one output for regression.

During training, the backbone network is frozen and only the regression head is updated.

- 5) *Training Setup:* The model is trained with the following settings:

- Optimizer: Adam with learning rate 10^{-3} .
- Loss function: Mean Absolute Error (MAE).
- Batch size: 8.
- Maximum number of epochs: 30.

Two callbacks are used:

- Early Stopping on validation MAE with patience 5.
- ReduceLROnPlateau on validation MAE with factor 0.5 and patience 2.

The model weights with the best validation MAE are restored at the end of training.

6) *Evaluation:* After training, the model is evaluated on the validation set. The following regression metrics are reported:

- Mean Absolute Error (MAE),
- Root Mean Squared Error (RMSE),
- Coefficient of determination (R^2).

These metrics provide a clear view of the prediction error and the overall goodness of fit.

7) *Test Prediction and Submission:* For the test set, the same preprocessing steps are applied. The trained model is used to generate predictions, which are saved to a CSV file containing two columns:

- filename,
- HC.

This file is used as the final submission.

II. EXPERIMENTAL RESULTS

In this section, we present the experimental results of the proposed model for fetal head circumference estimation. The performance is evaluated on the validation set using standard regression metrics.

A. Evaluation Metrics

To evaluate the regression performance, we use three common metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)

B. Training Process

The model is trained using the training set and evaluated on the validation set. During training, we monitor the training loss and the validation loss. The Early Stopping and ReduceLROnPlateau strategies are used to avoid overfitting and to stabilize the training process.

C. Quantitative Results

Table I shows the regression performance of the model on the validation set.

TABLE I
PERFORMANCE ON THE TEST SET

Metric	MAE (mm)	RMSE (mm)	R^2
ResNet-based model	15.34	20.64	0.89

The results show that the model can predict the head circumference with a reasonable error. The MAE and RMSE values indicate that the average prediction error is within an acceptable range. The R^2 score shows that the model can explain a large part of the variance in the data.

D. Discussion

Overall, the experimental results show that the proposed ResNet-based regression model is able to learn meaningful features from ultrasound images. It can estimate the fetal head circumference with good accuracy. However, there is still room for improvement, especially for difficult samples and extreme cases.

III. COMPARISON WITH HC18 LEADERBOARD

Results of the test set are compared with representative methods from the HC18 Grand Challenge leaderboard. Methods on this leaderboard typically use segmentation followed by ellipse fitting to derive head circumference.

A. Comparison with Published Results

Representative methods from the HC18 leaderboard demonstrate lower measurement errors through segmentation-based approaches. For example, state-of-the-art models reported in the literature using the HC18 dataset typically achieve mean absolute differences below 2 mm with high segmentation accuracy.

While direct comparison across datasets and evaluation protocols is not fully standardized, the large difference in error magnitude indicates that segmentation-based pipelines currently outperform simple regression approaches in terms of absolute accuracy.

B. Discussion

The comparison highlights a clear trade-off in fetal head circumference estimation. Segmentation-based methods achieve higher accuracy but require detailed annotations and complex pipelines, while the proposed regression-based approach provides a simpler and annotation-efficient alternative at the cost of higher error.

IV. CONCLUSION

In this work, we studied how to estimation head circumference (HC) from ultrasound images using a deep learning approach. Instead of a traditional segmentation-based pipeline, we proposed a regression-based method that directly predicts the HC value from the input image and pixel size using a ResNet-based network.

Experimental results show that the model learns meaningful visual features and achieves stable performance on the test set. Although its error is higher than state-of-the-art segmentation-based methods on the HC18 benchmark, the proposed approach has a much simpler pipeline and does not require pixel-level annotations.

This shows a clear trade-off between accuracy and complexity: segmentation-based methods are more accurate but require detailed annotations, while the proposed regression-based method is lightweight and annotation-efficient.

In future work, performance can be improved by combining regression with shape information, using stronger backbones, or applying multi-task learning.

REFERENCES

- [1] M. A. Rahman, S. Khan, N. Werghi, and M. A. Khan, “Automatic measurement of fetal head circumference from ultrasound images using deep learning,” *HC18 Grand Challenge: Head Circumference Estimation*, 2018. [Online]. Available: <https://hc18.grand-challenge.org/>
- [2] A. F. Baumgartner, K. Kamnitsas, J. Matthew, et al., “SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2204–2215, 2017.