Fetal head segmentation

Nguyen Xuan Minh Vu

March 17, 2024

1 Introduction

Nowadays, before giving birth to a baby, the mother must do many prenatal check-ups for health monitoring. This ensures that the baby will be born healthy and detect any problems soon. One of the most common procedures for prenatal check-ups is ultrasound. The ultrasound works by emitting very high-frequency signals (or sounds) to build and reconstruct the image of structure inside the body.

Ultrasound is used to investigate many body parts of the fetus, especially the fetal head. The fetal head is egg-shaped, being broader posteriorly and symmetric without irregularity of contour. Ossification of the skull vault is complete by 12 weeks gestation with the sutures remaining visible throughout pregnancy. By doing the ultrasound, the doctor can recognize any abnormality in the fetus' head.

The segmentation task is an important procedure in ultrasound as it can accurately capture the fetal head position inside the mother's body using a 2D ultrasound image. Usually, it is done manually by the doctor; however, this may cost time when there are too many images. It is also complicated to utilize modern machines as the data must be uploaded handly instead of segmenting automatically. Therefore, during the scope of this report, I introduce a segmentation model using convolutional neural networks (CNN) to detect the shape and position of the fetus's head, which serves the works of analyzing and investigating abnormalities later.

This project will use a variant of the Unet model and the HC18 dataset.

2 Background

2.1 Head circumference (HC)

During pregnancy, ultrasound imaging measures fetal biometrics. One of these measurements is the fetal head circumference (HC). The HC can be used to estimate the gestational age and monitor the growth of the fetus. The HC is measured in a specific cross-section of the fetal head, which is called the standard plane.

2.2 Convolutional Neural Network (CNN)

CNN excels at processing data like images, speech, and audio signals. This is due to their unique architecture composed of three main layers: The convolutional layer is to extract information. The pooling layer reduces the dimensionality of the data and captures spatial variance. Lastly, the fully connected layer combines the extracted features from earlier layers and makes the final classification or prediction.

3 Dataset

In this project, I will use the HC18 dataset, provided by Grand Challenge. This dataset includes 999 images for the training set and 335 images for the test set. Those images are captured by ultrasound, and present the real fetus's head inside the mother's body. With each image in the training dataset, there is an annotation dataset, which bounds the shape of the fetus's head. This is used as the label of the corresponding ultrasound image. On the other hand, the test set is independently used to evaluate the model so there is no annotation.

4 Method

4.1 Data preprocessing

There are 999 data samples in the training set, most of which are 540x800. However, some images have different shapes so I decided to cut them off. Therefore, the number of data for training is 975. They are resized to 256x256.
For the annotation images, the area inside the ellipse boundary is filled to white to get segmented images.

4.2 Network architecture

The network architecture is described in 1. The model is in the "U" shape. The left side will undertake the task of extracting features. It does the 3-by-3 convolutions and uses batch normalization and the ReLU activation function after each convolution layer. For down-sampling, the 2-by-2 MaxPool with stride = 2 will increase the channel size twice. The output of the left side is then fed to the right side to do the predicting process and produce the result of a segmented image. It uses 2-by-2 up-convolutions results concatenate with the corresponding layers on the left side and continue do the 3-by-3 convolution. In the last layer, the 1-by-1 convolution, BatchNorm, and Sigmoid are performed to achieve the segmented prediction with the same height and width as the original resized image.
The special of this architecture is that there is no fully connected layer. Therefore, the input size can be varied and the learning process can be faster.

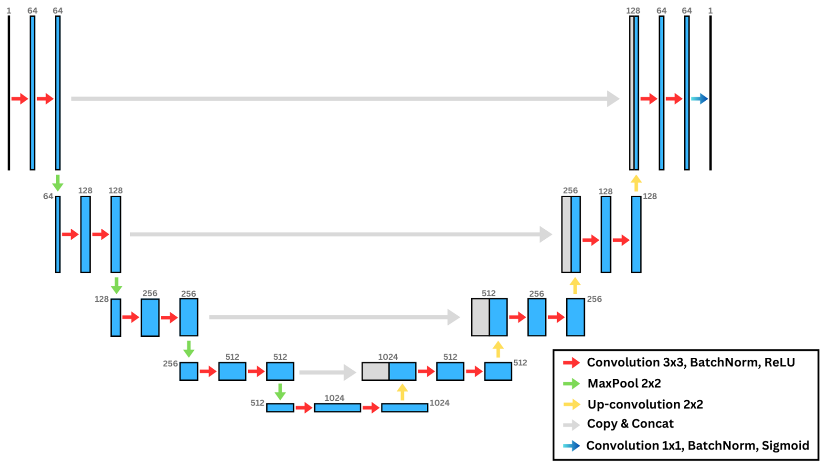

Convolution 3x3, BatchNorm, ReLU
MaxPool 2x2
Up-convolution 2x2
Copy & Concat
Convolution 1x1, BatchNorm, Sigmoid

Figure 1: Overall Unet custom model

There are some adjustments in the model compared with the original one to suit the current problems. Firstly, the Unet model uses copying and cropping to concatenate two blocks as the size is lost in the convolution steps. I only do the concatenation but at each convolution layer, I add the padding so that the size will not change in that process and I can get the last output with the same size as the input. Secondly, I use the BatchNorm method between the convolution and the ReLU, which makes the model faster and more stable as the data is continuously re-scaled and re-centered. Lastly, I use the Sigmoid activation function to the last layer to get a matrix with values between 0 and 1 to do the criterion.

4.3 Training

For training, I use the Adam optimizer algorithm with a learning rate = 0.001, and weight decay = 0.00000001. For the loss function, I use the Binary Cross Entropy Loss. The batch size for training is set as 16 while validating is 4 in this work. The model runs through 100 epochs.

4.4 Data splitting

The training dataset is divided into a smaller training set and a validation set. From 975 images, I used 750 samples for the training set and 225 samples for the validation set.

4.5 Data postprocessing

The returned prediction is in the shape of 256x256 and the shape of the segmented area is not a perfect ellipse. To get a prediction the same as the annotation, I first set the threshold = 0.2 to get a binary image. From that, I can get the contours of the predicted fetus' head area. Using the contours, the most fitted ellipse shape is estimated, and the ellipse features: the central point, the major axis, the minor axis, and the angle are extracted. Lastly, I only keep the boundary of the ellipse and resize that image back to 540x800. The illustration of the process is shown in 2. The comparison between the original annotation image and the segmented image is illustrated in 3.

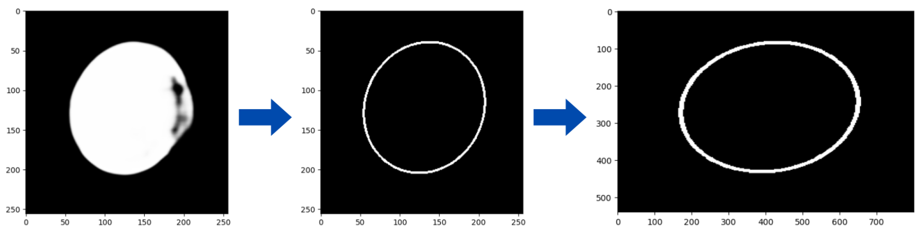
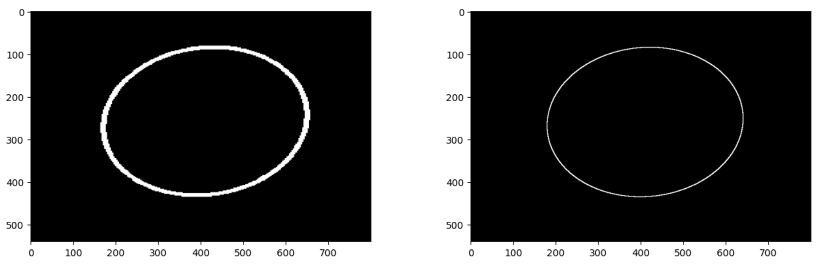


Figure 2: Postprocessing process illustration



(a) Annotation image

(b) Output

Figure 3: Compare the initial input and the output

5 Evaluation

5.1 Loss Progression

The losses after 100 epochs are illustrated in 4. For the first 30 epochs, both loss in training and validating are significantly decreased, nearly reaching 0.1.

After that, the validating loss is stable between 0.11 and 0.8 while the training loss still decreases.

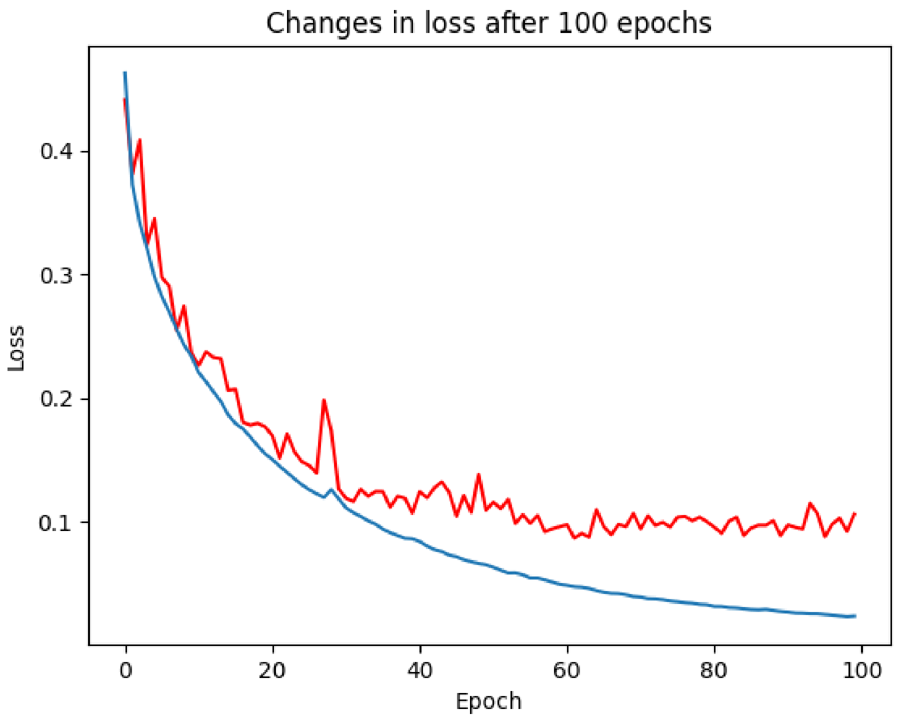


Figure 4: Losses in 100 epochs (black: training loss, red: validating loss)

5.2 Metric

The Mean Absolute Error (MAE) metric is used to evaluate this model. After 100 epochs, the MAE is computed as 0.0417 for validating data and 0.0176 for training data.

5.3 Result

The results displayed in this section are the prediction on another dataset with no annotation. Two random HC images are shown in 5.

By putting the original images in the model, we can achieve the ellipse shape of the fetus's head as 6.

The model has shown its superior result as the ellipse accurately bounds the fetus's head area. For a better visualization, I will combine the ellipse and the image in one image. They are shown in 7.

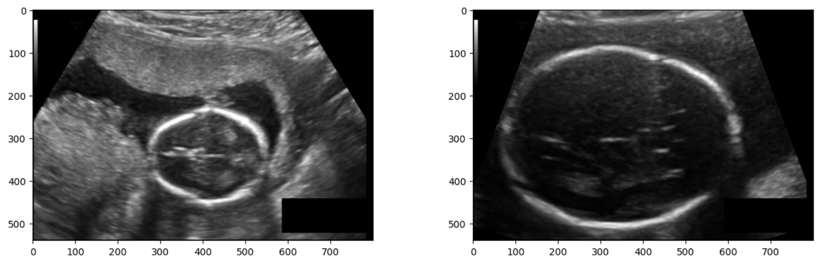

(a) Image 1 (b) Image 2

Figure 5: 2 random images from test dataset

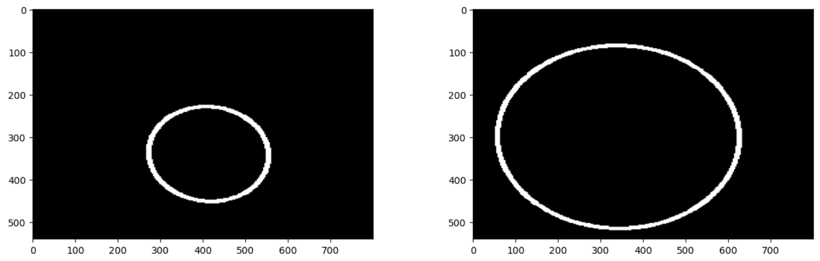

(a) Ellipse 1 (b) Ellipse 2

Figure 6: Ellipse shapes bound the fetus's head

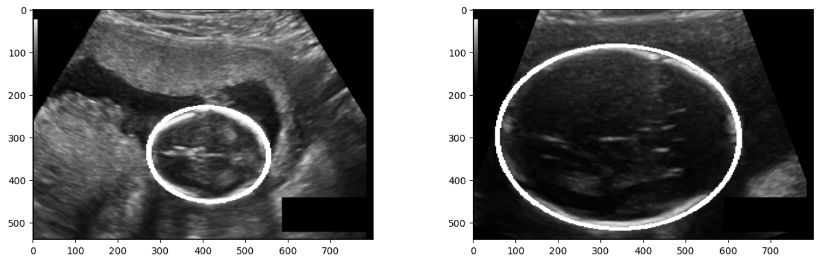

(a) Merge 1 (b) Merge 2

Figure 7: Ellipse shapes on the real images

6 Conclusion

In general, the model performs significantly well on most of the data in both training and validating datasets. There are some cases that the model cannot fully detect the segment area as in the real image, there is not a closed boundary. However, the post-processing process can handle them and return a perfect ellipse shape that covers the segmented area.
I will continue this work in the future to improve the model, do with more

datasets, and implement this model in other fields of medicine.