Fetal ultrasound image segmentation using YOLOv8

Chu Bao Minh

March 2024

1 Introduction

Ultrasound is a valuable tool for monitoring both the fetus and the mother during pregnancy due to its non-invasiveness and lower cost compared to modalities like CT or MRI. It is widely used for various clinical applications [1]. During pregnancy, ultrasound is commonly used to assess fetal development and calculate fetal biometric measurements such as biparietal diameter (BPD), head circumference (HC), and abdominal circumference (AC). Accurate measurements of the fetal head can be used to determine gestational age and evaluate fetal growth [2]. Experts examine ultrasound data to calculate dimensions of the fetal head, body, and movements, enabling them to identify abnormalities in fetal growth [3]. Precise quantification of anatomical structures, such as head circumference (HC), biparietal diameter, abdominal circumference, femur length, and humerus length, is crucial for evaluating fetal development [4]. However, manual measurement of ultrasound fetal biometric measurements is time-consuming and operator-dependent, relying on proper image acquisition angles and caliper positioning [5]. Therefore, automated image segmentation and measurement techniques are needed to streamline this process and assist in result analysis [6].

Ultrasound images often suffer from challenges such as varying intensities, speckle noise, acoustic shadows, discontinuous anatomical boundaries, and low contrast, which can complicate border recognition [7]. Head circumference (HC) is typically assessed in the standard plane, representing the perimeter of the fetal skull boundary in a cross-sectional view of the fetal head, as shown in Figure 1. Currently, physicians manually segment the fetal HC using ultrasound equipment. The segmented HC is then fitted with an ellipse to automate the longitudinal measurement of HC. However, the scarcity of experienced sonographers, particularly in low-resource settings, presents a significant challenge.

In order to address these issues, advanced automated image segmentation and measurement techniques are needed to improve efficiency and accuracy in ultrasound-based fetal diagnosis. These techniques can reduce the reliance on operator expertise, minimize human error, and enhance the accessibility of fetal assessments in diverse healthcare settings.

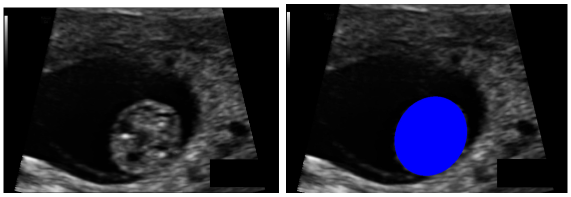


Figure 1: Sample images of 2D ultrasound fetal head area

Recent research has focused on utilizing deep learning algorithms, particularly convolutional neural networks (CNNs), for the estimation of fetal biometric characteristics in ultrasound images. These algorithms show promise in automating the determination of fetal biometric attributes, thereby enhancing efficiency and reducing diagnostic steps and time. In this context, a proposed automated fetus skull segmentation model using 2D ultrasound imaging is introduced. The model, based on a Dilated Multi-Scale LinkNet, demonstrates superior results compared to competing strategies, improving segmentation accuracy. By automating the skull segmentation process, this approach streamlines fetal ultrasound analysis, enabling accurate measurements and biometric attributes acquisition. This advancement has the potential to revolutionize prenatal care by improving efficiency and accuracy, benefiting both healthcare professionals and expectant parents.

2 Methodology

2.1 YOLOv8 Overview

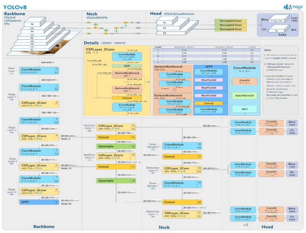


Figure 2: YOLOv8 architecture

We utilize YOLOv8 nano version for the segmentation task, as it is the new state-of-the-art computer vision model. This latest version has the same architecture as YOLOv5 with numerous improvements, such as a new neural network architecture that utilizes both Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). These features make it easier to annotate images for training the model. The FPN works by gradually reducing the spatial resolution of the input image while increasing the number of feature channels. This results in feature maps capable of detecting objects at different scales and resolutions. The PAN architecture, on the other hand, aggregates features from different levels of the network through skip connections. By doing so, the network can better capture features at multiple scales and resolutions, which is crucial for accurately detecting objects of different sizes and shapes[9].

YOLOv8 uses CSPDarknet53[8] as its backbone, a deep neural network that extracts features at multiple resolutions (scales) by progressively downsampling the input image. The feature maps produced at different resolutions contain information about objects at different scales in the image and different levels of detail and abstraction. YOLOv8 can incorporate different feature maps at different scales to learn about object shapes and textures, which helps it achieve high accuracy in most object detection tasks. YOLOv8 backbone consists of four sections, each with a single convolution followed by a c2f module[10]. The c2f module is a new introduction to CSPDarknet53. The module comprises splits where one end goes through a bottleneck module (Two 3x3 convolutions with residual connections). The bottleneck module output is further split N times where N corresponds to the YOLOv8 model size. These splits are all finally concatenated and passed through one final convolution layer. This final layer is the layer where we will get the activations. After obtaining bounding boxes, we then classify each pixel in each bounding box to perform segmentation.

2.2 Loss function and update rule

The generalized loss function and weight update procedure can be defined as follows:

$$\mathcal{L}(\theta) = \frac{\lambda_{box}}{N_{pos}}\mathcal{L}(\theta) + \frac{\lambda_{cls}}{N_{pos}}\mathcal{L}_{cls}(\theta) + \frac{\lambda_{dfl}}{N_{pos}}\mathcal{L}_{dfl}(\theta) + \phi{\theta_2}^2 \tag{1}$$

$$V^t = \beta V^(t-1) \tag{2}$$

$$\theta^t = \theta t - 1 - \eta V^t \tag{3}$$

Where 1 is the generalized loss function incorporating the individual loss weights and a regularization term with weight decay φ, 2 is the velocity term with momentum β, and 3 is the weight update rule and η is the learning rate. The specific YOLOv8 loss function can be defined as:

$$\mathcal{L} = \frac{\lambda_{box}}{N_{box}} \sum_{x,y} 1^*_{c_{x,y}} [1 - q_{x,y} + \frac{b_{x,y} - \hat{b}_{x,y}{}_2^2}{\rho} + \alpha_{x,y} \nu_{x,y}]$$

$$+ \frac{\lambda_{cls}}{N_{cls}} \sum_{x,y} \sum_{c \in classes} y_c \log \hat{y}_c + (1 - y_c) \log(1 - \hat{y}_c)$$

$$+ \frac{\lambda_{dfl}}{N_{dfl}} \sum_{x,y} 1^*_{c_{x,y}} \left[-(q_{(x,y)+1} - q_{x,y}) \log \hat{q}_{x,y} \right.$$

$$\left. + \, (q_{x,y} - q_{(x,y)-1}) \log \left(\hat{q}_{(x,y)+1} \right) \right]$$

where:

$$q_{x,y} = IoU_{x,y} = \frac{\hat{\beta}_{x,y} \cap \beta_{x,y}}{\hat{\beta}_{x,y} \cup \beta_{x,y}}$$

$$\nu_{x,y} = \frac{4}{\pi^2} (arctan(\frac{w_{x,y}}{h_{x,y}}) - arctan(\frac{\hat{w}_{x,y}}{\hat{h}_{x,y}}))^2$$

$$\alpha_{x,y} = \frac{\nu}{1 - q_{x,y}}$$

$$\hat{y}_c = \sigma(\cdot)$$

$$\hat{q}_{x,y} = softmax(\cdot)$$

and:
− Npos is the total number of cells containing an object.
− is an indicator function for the cells containing an object.
− $\beta_{x,y}$ is a tuple that represents the ground truth bounding box consisting of (x_coord,y_coord, width, height).
− $\hat{\beta}_{x,y}$ is the respective cell's predicted box.
− $b_{x,y}$ is a tuple that represents the central point of the ground truth bounding box.
− y_c is the ground truth label for class c (not grid cell c) for each individual grid cell (x,y) in the input, regardless if an object is present.
− $q_{(x,y)} + / - 1$ are the nearest predicted boxes IoUs (left and right) $\in c^*_{x,y}$
− $w_{x,y}$ and $h_{x,y}$ are the respective boxes width and height.
− ρ is the diagonal length of the smallest enclosing box covering the predicted and ground truth boxes.

Each cell then determines its best candidate for predicting the bounding box of the object. This loss function includes the CIoU (complete IoU) loss proposed by Zheng et al. [11] as the box loss, the standard binary cross entropy for multi-label classification as the classification loss (allowing each cell to predict more than 1 class), and the distribution focal loss proposed by Li et al.[12] as the 3rd term. Finally, we add BCE loss to do the classification on each pixel.

2.3 Model diagnosis

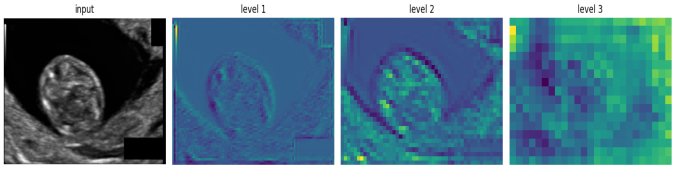


Figure 3: Feature activation maps for CT scan. From left to right, we have 3 features extracted from the model's CSPDarknet53 backbone

Figure 3 shows the original CT scan image and the activation of the four c2f stages in the network, with each stage being more profound in the network from the second image right. The Activation Map corresponding to the shallowest c2f module shows the broadest activation. This module shows the abstraction of organs and determine what is these organs look like. The second activation map corresponds to the second c2f module in our backbone. It shows strong activation in the general shape of the organs. It appears that this layer is attempting to infer what type of each organ looks like in the image by highlighting these features. Finally, the model's final c2f module activates extremely fine-grained details and outlines in the respective images.

2.4 Segmentation result

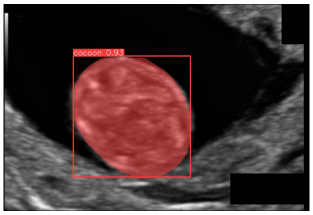
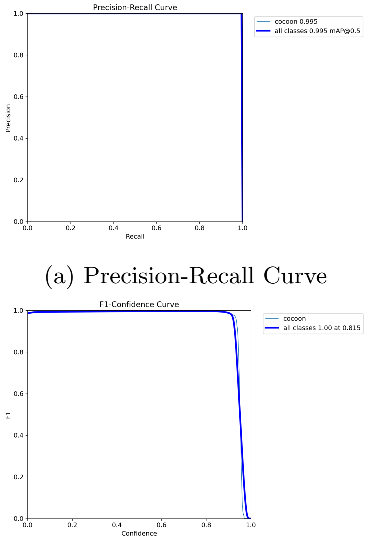


Figure 4: Output of segmentation

After training for 50 epochs, the model shows an accurate segmentation in Figure 4 with a high confidence score (over 80%) and good metrics below:



(a) Precision-Recall Curve

(b) F1 Score Curve

Figure 5: Performance Curves

Figure 5a shows that the average AUC of the precision-recall curve is very high (99.5% at IoU threshold=0.5), which means the model excels at localizing the fetal. Figure 5b shows a very high F1 score: 1 at IoU threshold 0.815.

There are some miss segmentation case in the figure below

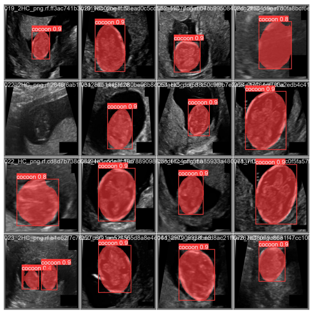


Figure 6: Some predictions

This failure happens when the fetal is too blended with the background, hardly visible even to human eye (Figure 6 second row) or there is another organ looks similar to the fetal (Figure 6 last row). However, we can increase the IoU threshold to eliminate the false segmentation.

In conclusion, even though the nano version of YOLOV8 contains only 3 million parameters and uses less computational cost (only 8.1 GFLOPs), it still handles the segmentation very well.

3 Conclusion

The reliability of a fetal segmentation method using YOLOv8 is demonstrated in this study. The method exhibits excellent metrics, indicating its accuracy and effectiveness in the task. Notably, the inference time is remarkably short, taking approximately 20 milliseconds. This swift processing time significantly enhances productivity in fetal head analysis, allowing for more efficient and streamlined procedures.

4 Future work

The presented results employing YOLOv8 for fetal head segmentation show promising outcomes. Future work should focus on several key areas to propel the field of medical image analysis forward and enhance clinical applicability.

Fine-tuning the existing models or considering more advanced architectures, possibly through transfer learning with pre-trained models on larger medical image datasets, presents an opportunity to achieve heightened accuracy and generalization. Systematic hyperparameter tuning, encompassing adjustments to learning rates, batch sizes, and other relevant parameters, may lead to the identification of optimal configurations. Furthermore, incorporating model interpretability techniques will aid in understanding and visualizing decision-making processes, particularly important in critical medical applications. Optimizing models for reduced inference time ensures practical usability in clinical settings. Establishing continuous monitoring and updating mechanisms based on new data and evolving medical standards is vital for sustained accuracy and reliability. In summary, future research in medical image analysis should embrace advanced architectures, data augmentation, interpretability, and collaboration with healthcare professionals to create robust, efficient, and clinically applicable models, ultimately improving patient care in diverse medical scenarios.

References

[1] S. Rueda et al., Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge, IEEE Trans. Med. Imaging, vol. 33, no. 4, pp. 797–813, 2014.

[2] Loughna P, Chitty L, Evans T, Chudleigh T (2009) Fetal size and dating: charts recommended for clinical obstetric practice. Ultrasound 17(3):160–166.

[3] V. Sundaresan, C. P. Bridge, C. Ioannou, and J. A. Noble, Automated characterization of the fetal heart in ultrasound images usingfully convolutional neural networks, in Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI), Apr. 2017, pp. 671-674.

[4] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree, IEEE Trans. Med. Imag., vol. 27, no. 9, pp. 1342-1355, Sep. 2008.

[5] R. C. Sanders and A. E. James, The Principles and Practice of Ultra-sonography in Obstetrics and Gynecology. New York, NY, USA: Appleton, 1985.

[6] S. M. G. V. B. Jardim and M. A. T. Figueiredo, Segmentation of fetal ultrasound images, Ultrasound Med. Biol., vol. 31, no. 2, pp. 243-250, Feb. 2005.

[7] Wisnu Jatmiko, Ikhsanul Habibie, M. Anwar Ma'sum, Robeth Rahmatullah, I Putu Satwika, Automated Telehealth System for Fetal Growth Detection and Approximation of Ultrasound Images, International Journal on Smart Sensing and Intelligent Systems, vol. 8, no. 1, 2015.

[8] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *ArXiv*, abs/1804.02767. Retrieved from https://api.semanticscholar.org/CorpusID:4714433

[9] Terven, J. R., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction.* doi:10.3390/make5010007

[10] Solawetz, J. (2023, Dec). What is Yolov8? the ultimate guide. *Roboflow Blog.* Retrieved from https://blog.roboflow.com/whats-new-in-yolov8/

[11] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 12993-13000). Retrieved from https://doi.org/10.1609/aaai.v34i07.6999

[12] Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., & Yang, J. (2020). Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *ArXiv*, abs/2006.04388. Retrieved from https://api.semanticscholar.org/CorpusID:219531292