Adam Lang
4/22/2024
Springboard Data Science Fellow

Capstone 2 Project Proposal Fetal Head Detection and Measurement on Ultrasound Images

Two-dimensional (2D) ultrasound imaging is the gold standard for prenatal screening and fetal monitoring as it is low cost, non-invasive, real-time, and has no radiation hazards. The fetal head circumference (HC) can monitor fetal growth, estimate gestational age, and predict delivery methods for pregnant women. Therefore, fetal HC is one of the most important biological characteristics in prenatal ultrasounds. In current clinical practice, fetal HC measurements are performed manually by Sonographers which is not convenient, time-consuming, and heavily dependent on human skill and experience. However, an accurate measure of the fetal HC is not possible without precise detection of the fetal skull boundary which is often difficult in 2D ultrasound due to numerous imaging factors; all of which can lead to incomplete and incorrect detection of the fetal skull boundary and false measurements (Yang et al, 2022). It is also well known that the manual biometric measurements of the fetal head vary significantly by inter- and intra-observer variability and thus the concept of having a more accurate standard measuring system such as a computer vision algorithm could reduce the amount of time and variability because it would not be subject to intra-observer variability (van den Heuvel, 2018).

Furthermore, in medically underserved areas of the U.S. and across the world, Medical Sonographers are not always available which is why companies like Google Health are leading the way in developing algorithms for ultrasound detection and measurement of the fetal head (BBC News, 2024). In addition, Ultimately, refining image detection algorithms for fetal ultrasound can help everyone involved in prenatal care whether they live in a big city or a rural underserved region. It is well known that 99% of maternal deaths worldwide occur in developing countries and enhancing the prenatal care before, during and after childbirth can and will save lives of countless women and newborns (van den Heuvel, 2018).

Criteria for success and Scope of solution space

The goal of this project is to utilize a dataset of 1,334 2D ultrasound images (999 training, 335 test) to perform standard deep learning computer vision tasks in object recognition: Image classification (fetal head vs. other), Image Classification Class Label for Gestation Age (GA): 1st trimester, 2nd trimester, 3rd trimester (although these measures are also closely related to the object segmentation task), Object detection (fetal head present), Object localization (skull boundary), and Object segmentation (fetal head measurements). The most basic level of success would be to have a positive prediction on each task mentioned.

The more specific criteria for success will follow the standard machine learning metrics for computer vision tasks from previous studies that utilized this same dataset:

1. Dice Similarity Coefficient (DSC)
2. Jaccard Similarity score
3. Mean Pixel Accuracy (mPA)
4. Mean intersection over union (mIoU)

The metrics above have previously been used for the “Object Segmentation” tasks to measure actual vs predicted. The DSC is the more standard measure and previous studies have had results between 92 and 97% (Chenarlogh et al. 2022). In addition, other common metrics used for image classification tasks that should be used to assess success include (Hoque, 2024).

1. Accuracy
2. Precision
3. Recall
4. Precision-Recall Curve (optimization of the precision-recall tradeoff)
5. F1 score
6. ROC curve
7. AUC score
8. Confusion Matrix

Constraints

The first major constraint that I anticipate encountering is the known fact that there is an imbalance in the dataset as published by the authors in the original paper. While the data is balanced appropriately with 75% training data and 25% test data, the imbalance is in the class labels for the trimester of pregnancy. The class labels are as follows:

Trimester	Train Set	% of train set	Test Set	% of test set
1st	165	16.5%	55	16.5%
2nd	693	69%	233	69%
3rd	141	14%	47	14%
Totals	999	75%	335	25%

Why is this imbalance an issue or constraint? **It was shown in the original study that produced this dataset, *“that it is important to separate the results for each trimester, because the uncertainty of the estimated gestational age (GA) increases with GA due to the natural variation in fetal size increases with GA* (van den Heuval, 2018)**.” It is also well known that the GA can be estimated more accurately in the first trimester, but the fetal skull is not clearly visible in the first trimester, which makes automated detection of the fetal head circumference (HC) and segmentation measurement a more challenging task (van den Heuval, 2018). The fetal skull is also very soft in the first trimester, and does not always appear brighter than the inside of the fetal head. Therefore, it is often very difficult to detect the edge of the fetal head, especially when it lies close to the wall of the uterus (van der Heuval, 2018). This means that the pixel size increases from 1st to 2nd to 3rd trimesters. The pixel size differences will be a constraint to simple detection of the fetal head, let alone trying to measure the segmentation of the head circumference. Thus to deal with this we should consider the different algorithms available to handle these constraints: U-NET, YOLO and CNN’s (all types).

A few more constraints to consider include:

- The data for this study was acquired in only one hospital using two different ultrasound devices from the same vendor which could introduce immediate bias to the training data.
- Most images were acquired after 12 and 20 weeks of pregnancy, since these are standard time intervals for routine ultrasound screening for pregnant women in the Netherlands where the study was conducted. This time frame varies by country and region, one international study reported that the first ultrasound is usually obtained between 18 and 22 weeks and the 2nd ultrasound between 32 and 36 weeks (Kim et al. 2018). The American College of Gynecology suggests having at least 1 ultrasound between 18 and 22 weeks but this varies based on fetal development, the mother, and other variables (ACOG, 2024). Obviously this means that these results may only represent a sample of the true population of fetal ultrasounds and may not be able to generalize on larger test sets.
- Lastly, the size of each 2D ultrasound image is 800 by 540 pixels with a pixel size ranging from 0.052 to 0.326 mm. This large variation in pixel size is a result of adjustments in the ultrasound settings by the sonographer (depth settings and amount of zoom are routinely varied during the examination) to account for the different sizes of the fetuses. This means that the training data for this project is highly dependent upon the sonographer who obtained the measurements. Again, this should be considered as we would try to generalize the results on a larger test set.

Stakeholders

This is a hypothetical case study that can be applied in the real world. I will assume the stakeholders would be:

- Google Health and other companies using AI to improve medical care.
- Medical providers (e.g. Physicians, Nurses, Radiologists, Sonographers) in the U.S. and other countries.
- Hospitals and Clinics in the U.S. and other countries as well as their respective clinical informatics teams.
- GE Healthcare and other companies that manufacture fetal ultrasound devices.
- Data Scientists and Machine Learning Engineers

Data Sources

The dataset for this project originated from a medical imaging challenge that began in the year 2018 called *"Automated measurement of fetal head circumference using 2d ultrasound images".* The results were originally published in a journal article by van den Heuvel et al.

Metadata information is accessed on the IEEE website: **IEEE Dataport 2D ultrasound**

Dataset download via Zenodo: **Download ultrasound dataset**

References

- ACOG, 2024. “Ultrasound Exams.” Accessed on 4/24/2024. Retrieved from: https://www.acog.org/womens-health/faqs/ultrasound-exams
- Ashkani Chenarlogh V, Ghelich Oghli M, Shabanzadeh A, et al. Fast and Accurate U-Net Model for Fetal Ultrasound Image Segmentation. *Ultrasonic Imaging*. 2022;44(1):25-38. doi:10.1177/01617346211069882
- BBC News, 2024. How AI could ‘save lives’ of pregnant women in Kenya. Link: https://www.youtube.com/watch?v=IZA-5cg3_WI
- G. Wang, X. Luo, R. Gu, S. Yang, Y. Qu, S. Zhai, Q. Zhao, K. Li, S. Zhang. (2023). PyMIC: A deep learning toolkit for annotation-efficient medical image segmentation. Computer Methods and Programs in Biomedicine (CMPB). February 2023, 107398. Github: https://github.com/HiLab-git/PyMIC
- Hoque, 2024. “How to Evaluate An Image Classification Model.” Accessed on 4/24/2024. Retrieved from: https://docs.clarifai.com/tutorials/how-to-evaluate-an-image-classification-model
- Kim ET, Singh K, Moran A, Armbruster D, Kozuki N. Obstetric ultrasound use in low and middle income countries: a narrative review. Reprod Health. 2018 Jul 20;15(1):129. doi: 10.1186/s12978-018-0571-y.
- Machine Learning Mastery – Object Recognition with Deep Learning: https://machinelearningmastery.com/object-recognition-with-deep-learning/
- Thomas L. A. van den Heuvel, Dagmar de Bruijn, Chris L. de Korte and Bram van Ginneken. Automated measurement of fetal head circumference using 2D ultrasound images. *PloS one*, 13.8 (2018): e0200412.
- Thomas L. A. van den Heuvel, Dagmar de Bruijn, Chris L. de Korte and Bram van Ginneken. Automated measurement of fetal head circumference using 2D ultrasound images [Data set]. Zenodo. http://doi.org/10.5281/zenodo.1322001
- V. Nagabotu and A. Namburu, Precise segmentation of fetal head in ultrasound images using improved U-Net model, *ETRI Journal* (2023), 1–12, DOI 10.4218/etrij.2023-0057.
- Yang et al, 2022. “A new approach to automatic measure fetal head circumference in ultrasound images using convolutional neural networks.” Computers in Biology and Medicine. Link: https://www.sciencedirect.com/science/article/abs/pii/S0010482522005674?via%3Dihub