Janeshvar Sivakumar

Segmentation-Based Approach to Landmark Detection

As a segmentation task would require, and similar to what I did earlier in Part A, I began by performing additional exploratory data analysis to draw clear conclusions. The foreground-to-background ratio was extremely low, and as expected with ultrasound data, the dataset was very noisy. The fetal head appeared in many orientations, and most importantly, the skull was visible primarily as a thin, bright boundary.

My first instinct was to treat this as a conventional skull segmentation task, where the model predicts the full skull region. I implemented a standard U-Net with class imbalance handling and trained it on the provided masks. The results were very poor. Dice scores remained near zero, predictions collapsed to background, and even single-image overfitting failed. Initially, I attempted to address this through changes in loss functions, class weighting, and preprocessing, but none of these attempts led to meaningful improvements.

At this point, I paused and returned to the images themselves. I realized that I was asking the model to predict a filled region when the image signal only clearly supported the boundary. In other words, the learning target did not match the information present in the image.

Based on these observations, I reformulated the task as skull boundary segmentation rather than region segmentation. I converted the provided skull masks into thin boundary masks by extracting their contours.

I experimented with common preprocessing techniques such as CLAHE and median filtering. While these methods are often useful in natural image tasks, visual inspection showed that they degraded the thin skull boundary in ultrasound images. Along with quantitative metrics such as SNR, I made preprocessing decisions primarily based on visual evidence. In the end, I settled on minimal preprocessing consisting only of grayscale normalization. This reinforced the idea that in medical imaging, more preprocessing is not always better.

Data augmentation was introduced only after establishing a stable baseline. Exploratory analysis showed significant variation in fetal head orientation, while skull shape remained relatively consistent. Based on this, I applied mild, anatomically valid augmentations. I explicitly avoided elastic deformation, aggressive scaling, or intensity manipulation, as these could distort skull geometry and violate anatomical realism.

For the baseline model, I chose a standard U-Net, as it is widely used in medical imaging and was also included as a reference architecture in the challenge. The baseline configuration was intentionally simple:

- Input resolution: 256 × 256, downsampled
- Target: 1-pixel skull boundary
- Loss: weighted binary cross-entropy combined with soft Dice
- No data augmentation
- No postprocessing

From the predicted boundary, I extracted the dominant contour and fit an ellipse. The major and minor axes of this ellipse naturally corresponded to OFD and BPD, respectively.

I evaluated the baseline using simple, interpretable checks. These included whether an ellipse could be successfully fitted, whether BPD was smaller than OFD, and whether the axes were approximately perpendicular. These checks reflect basic anatomical constraints and are more meaningful in a clinical context than segmentation metrics alone. While the baseline performed reasonably well, failures occurred when boundary predictions were fragmented, which motivated further investigation.

At this stage, I froze the baseline and introduced controlled modifications one at a time.

Experiment 1: Higher resolution
Increasing the input resolution improved local boundary smoothness in some cases, but did not consistently improve downstream geometric stability. This suggested that resolution alone is not always the primary bottleneck in ultrasound imaging.

Experiment 2: Thicker boundaries
Increasing the boundary thickness resulted in a higher Dice score, but also introduced distortion due to neighborhood noise, which negatively affected geometric extraction.

Experiment 3: Mild augmentation improves robustness
Given the variability in head orientation, I introduced mild, anatomically valid augmentations such as small rotations and translations. This improved robustness to orientation changes and reduced some geometric failures without distorting skull shape.

Experiment 4: Postprocessing improves geometric stability
Finally, I explored simple inference-time postprocessing. Morphological closing and largest connected component selection significantly improved ellipse fitting success across all models. This reinforced an important lesson that not all robustness needs to be learned.

Results:

	Mean Dice	Boundary F1	Precision	Recall	Ellipse Success Rate	Mean BPD	Mean OFD	OFD/BPD Ratio (mean)	OFD/BPD Ratio (std)
baseline	0.232730	0.240687	0.158547	0.441896	1.0	139.115385	173.393497	1.274890	0.174057
hyp1_highres	0.119120	0.135033	0.079585	0.273043	1.0	267.903585	339.085526	1.388575	0.630050
hyp3_augmented	0.183679	0.183680	0.102206	0.920872	1.0	145.443567	179.413280	1.260400	0.166879

Future work
Given the time constraints and the fact that I had my final year thesis review during this period, there were several components I could not explore in depth. These include more systematic ablation studies of preprocessing choices, model architectures, and augmentation strategies. Beyond methodological improvements, I would also like to explore the use of temporal information from adjacent frames to further stabilize geometry.

Key takeaways from the results
The most important takeaway from this work is that problem formulation matters more than model complexity. Reformulating the task to match the physical imaging signal had a greater impact than any architectural change. Visual inspection is essential in medical imaging, geometry-aware pipelines are

more interpretable and reliable, and not all robustness needs to be learned. Simple rule-based postprocessing can significantly improve safety and stability.