FETAL HEAD SEGMENTATION

Automated Cranium Segmentation from Ultrasound Images

1. PROJECT OVERVIEW

Objective: Develop an automated system for fetal head segmentation from 2D ultrasound images to assist in biometric measurements.

Dataset: 622 ultrasound images with corresponding cranium masks
Split: Training (497), Validation (62), Testing (63)
Input: 256×256 RGB ultrasound images
Output: Binary segmentation mask + biometric points via ellipse fitting

2. HYPOTHESIS 1: BASELINE U-NET

Architecture: Standard U-Net (7.8M parameters)
Loss: BCE (0.4) + Dice (0.4) + Boundary (0.2)

Critical Issue: Boundary-only masks (~0.7% foreground coverage) caused the model to learn edge detection instead of full-region segmentation.

Results:
- Dice: 0.0370
- IoU: 0.0190

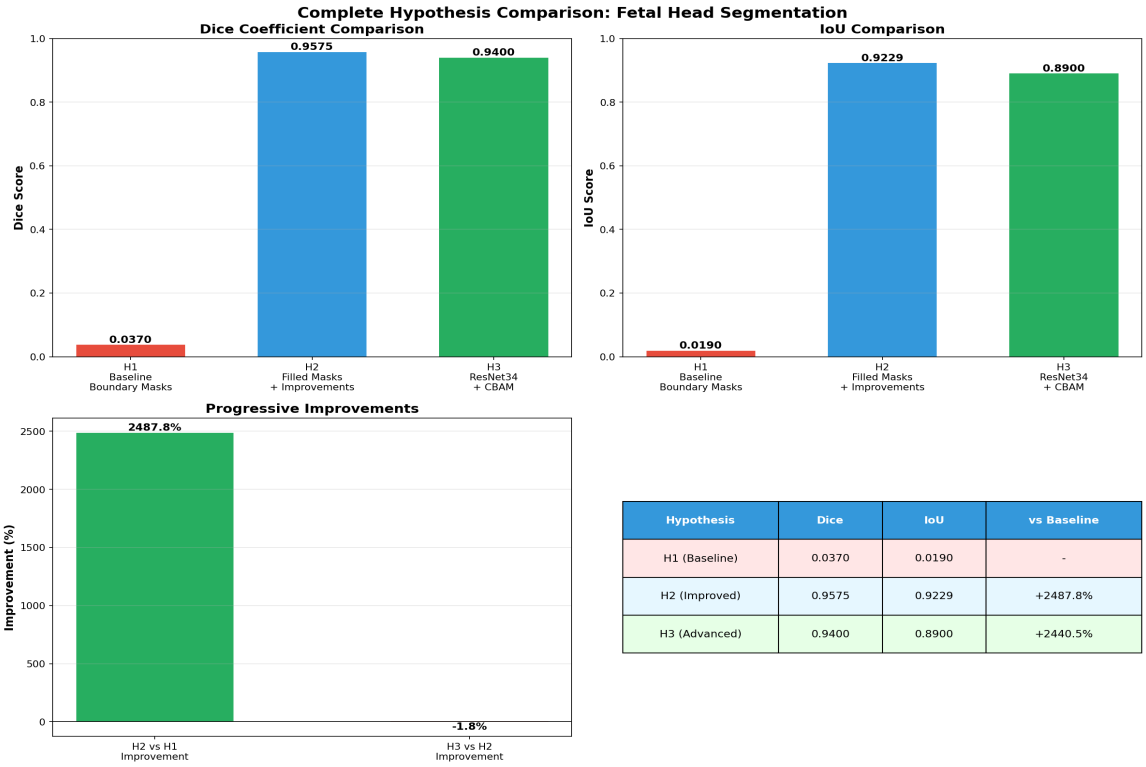
Key Learning: Data quality dominates architectural complexity.

3. HYPOTHESIS 2: IMPROVED U-NET (BEST MODEL)

Key Improvements:
- Filled masks via morphological operations
- Focal loss (α=0.25, γ=2.0)
- Heavy augmentation and regularization

Results:
- Dice: **0.9575 ± 0.0550**
- IoU: **0.9229 ± 0.0844**

Post-processing:
Morphological closing, largest connected component selection, ellipse fitting



Hypothesis	Dice	IoU	vs Baseline
H1 (Baseline)	0.0370	0.0190	-
H2 (Improved)	0.9575	0.9229	+2487.8%
H3 (Advanced)	0.9400	0.8900	+2440.5%

4. HYPOTHESIS 3: RESNET34 + CBAM

Architecture: ResNet34 encoder with CBAM attention
Loss: Focal + Tversky + Boundary

Results:
- Dice: 0.9400
- IoU: 0.8900

Analysis: Faster convergence but no significant gain over Hypothesis 2.

5. KEY FINDINGS & CONCLUSIONS

Primary Finding: Data quality > architecture complexity.

Best Model: Hypothesis 2 (Dice = 0.9575). Recommended for deployment.

5A. TRAINING, EVALUATION & BIOMETRIC EXTRACTION STRATEGY

Unified Training Pipeline:
A single unified training pipeline was used for all hypotheses. Differences were introduced only through data representation, loss functions, and architectural components, while keeping the training and evaluation framework consistent.

Evaluation Strategy:
Quantitative evaluation and qualitative visualization were performed exclusively on the best-performing hypothesis (Hypothesis-2), as it represents the final deployable system. Other hypotheses served as baselines.

Biometric Landmark Extraction:
Instead of training a separate keypoint detection network, four biometric points were extracted implicitly by fitting an ellipse over the segmented fetal head mask. The endpoints of the major and minor axes correspond to clinically relevant anatomical landmarks.

6. METHODOLOGY COMPARISON

Hypothesis	Architecture	Loss	Dice	IoU
H1	U-Net	BCE+Dice+Boundary	0.0370	0.0190
H2	U-Net + Filled Masks	Focal+Dice	0.9575	0.9229
H3	ResNet34 + CBAM	Focal+Tversky+Boundary	0.9400	0.8900

7. FUTURE RECOMMENDATIONS

- Multi-scale inference
- Model ensembling
- Test-time augmentation
- Clinical validation