Fetal Biometric Landmark Detection (Hourglass Model)

1. Problem Understanding & Overall Approach

The task is to automatically localize four fetal cranial biometric landmarks corresponding to the **BPD and OFD measurements** in axial ultrasound images. Unlike segmentation-based approaches, this problem requires identifying **precise anatomical points** rather than full regions. Therefore, my approach focuses on **landmark heatmap regression** instead of pixel segmentation.

My main objective was to design a model that:

- works reliably on **2D ultrasound images**
- learns **structural context of the fetal skull**
- remains lightweight and robust given the dataset size (~500–600 images)

Ultrasound data contains **speckle noise, intensity variations, and partial skull boundaries**, so I prioritised:

- architectures that preserve **spatial structure**
- representations that encode **local + global shape context**
- stable training rather than chasing raw accuracy

After reviewing alternatives, I chose a **Single-Hourglass CNN architecture with heatmap outputs** for final submission because it provided:

- interpretable landmark probability maps
- good balance between complexity and performance
- stable convergence with limited data

The model outputs **4 Gaussian heatmaps**, one for each landmark, and the final landmark locations are decoded from the peak responses.

---

2. Data Pre-Processing & Augmentation (with Rationale)

Pre-processing

- Converted images to **grayscale** → ultrasound data is single-channel
- Resized images to **256 × 256** with proportional coordinate scaling
- Normalized pixel intensity to **[0,1]**
- Converted to **3-channel tensors** for CNN compatibility

Rationale:
Normalization and resizing ensure consistent geometry and stable learning. Since fetal head orientation varies, consistent scaling prevents landmark drift across samples.

Landmark Heatmap Generation

For each landmark:

- Generated a **Gaussian heatmap** centered at ground-truth point
- Heatmaps serve as **soft spatial supervision**

Rationale:
Direct coordinate regression can be unstable in small datasets. Heatmaps allow the model to learn **spatial confidence patterns** instead of exact coordinates.

Data Augmentation

- Small rotations (±10°)
- Horizontal flips (limited cases)
- Slight scaling / translation
- Kept landmarks synchronized using keypoint-aware augmentation

Rationale:

- Ultrasound probes vary in angle and head alignment

- Mild geometric augmentations improve **pose robustness**
- Augmentations were intentionally **conservative** to avoid anatomical distortion

---

3. Other Models & Experiments (Hypotheses and Outcomes)

I explored multiple architectural directions to understand what works best for this task.

(A) Baseline: Single-Hourglass Heatmap Network (Final Selected Model)

- Encoder–bottleneck–decoder structure
- Outputs **4 landmark heatmaps**
- Trained with **MSE heatmap regression loss**

Hypothesis:
Hourglass preserves both **local contour detail** and **global cranial symmetry**, which is essential for anatomical landmark localization.

Outcome:
Provided stable training and good qualitative predictions. Selected as final model.

(B) Hybrid Model - Heatmap + Coordinate Regression Head

- Same backbone
- Two branches:
  - Heatmap output
  - Fully-connected coordinate regression head
- Joint loss = heatmap loss + coordinate loss

Hypothesis:
Combining **coarse spatial probability (heatmaps)** with **precise point refinement (coordinates)** may improve localization.

Observation:

- Loss values improved numerically
- But regression head was **more sensitive to noise**
- Pixel-error improvement did not generalize consistently

Takeaway:
Hybrid methods need **larger datasets** or landmark-specific constraints. Interesting direction but less stable than single-hourglass for now.

(C) Stacked Hourglass Model

- Multiple hourglass stages stacked sequentially
- Intermediate supervision applied

Hypothesis:
Iterative feature refinement could improve localization on subtle skull boundaries.

Outcome:

- Training became **complex and unstable**
- Model occasionally over-smoothed heatmaps
- Pixel errors increased for some landmarks
- The gains did not justify added complexity for dataset size

Conclusion:
Stacked refinement is powerful but **data-hungry**. Single-hourglass was more efficient and reliable in this setting.

---

4. If Given More Time - Future Improvements

If more time and compute were available, I would explore:

- **Self-supervised pre-training on large ultrasound datasets**
  - Improve feature generalization
- **Attention-based Hourglass Variants**
  - Learn to focus on cranial boundary curves
- **Bone-Contour Assisted Landmark Detection**
  - First detect skull ellipse
  - Then constrain landmark coordinates on ellipse boundary
- **Uncertainty-aware Landmark Prediction**
  - Predict confidence variance to handle ambiguous scans
- **Multi-view or meta-learning approaches**
  - Model robustness across machines and acquisition settings

These ideas align with clinical context and could meaningfully improve performance beyond brute-force scaling.

---

5. Key Takeaways from My Approach

- **Heatmap-based landmark regression** is well-suited for fetal ultrasound because it:
  - handles noisy textures
  - learns structural context
  - remains interpretable
- **Model simplicity matters in limited-data regimes**
  - Single-Hourglass performed more reliably than heavy stacked networks
- **Data augmentation must respect anatomy**

  - Aggressive augmentation degraded localization
- **Hybrid regression models are promising**
  - but require **more training samples** for stability
- The project strengthened my understanding of:
  - ultrasound image characteristics
  - coordinate-vs-heatmap supervision
  - architectural tradeoffs in medical landmark detection

Most importantly - the challenge was engaging and helped me approach the task like an applied research problem rather than just optimizing a score.

Model / Hypothesis	Description	Mean Pixel Error	Median Pixel Error	Notes
H1 - Single-Hourglass (Final Model)	Heatmap-based landmark regression with one hourglass backbone	**≈ 24 px**	**≈ 4 px**	Stable, consistent, best generalization
H2 - Hybrid Model (Heatmaps + Coordinate Regression Head)	Dual-branch output (heatmaps + FC coordinate head)	≈ 35–40 px	≈ 25–27 px	Sensitive to noise, inconsistent refinement gains
H3 -Stacked Hourglass	Multi-stage refinement with intermediate supervision	≈ 80–120 px	≈ 80–90 px	Training instability, over-smooth heatmaps