Put your heatmaps on! Let's find the facial biomarkers!

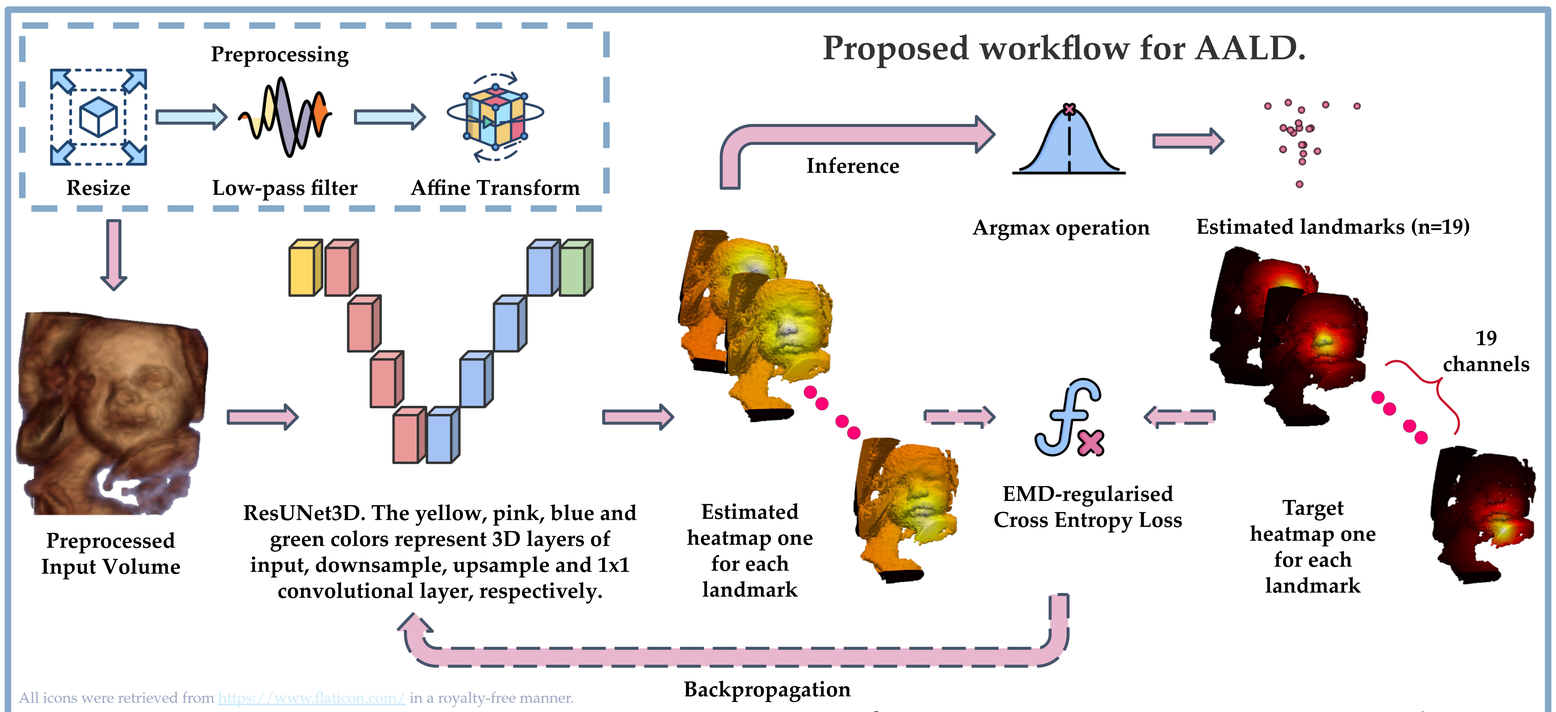
Yusuf B. Tanrıverdi, Federico Sukno, Gemma Piella

Towards an AI-assisted 3D fetal face reconstruction for craniofacial assessment: EMD -regularised heatmap regression for automated anatomical landmark detection (AALD)

Landmarking leads to a standardised analysis![1]

Evaluating fetal facial morphology remains observer-dependent, even with the recent 3D ultrasound trend[2].

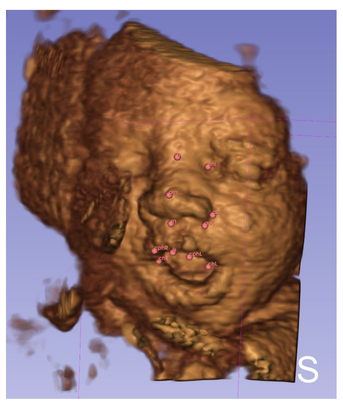
Automated anatomical landmark detection is essential to overcome characteristic limitations of 3D fetal ultrasound, as it serves as the key step toward objective, standardized analysis.



Proposed workflow preprocesses images with SimpleITK and Python. Image sizes are 128^3[3]. Rotation algorithm was implemented in previous work[4]. 19 anatomical landmarks were considered. 3D images are visualised by Slicer3D and PyVista software. All code is open-source accessible by scanning the QR.

Effective landmark detection requires <u>high-precision probability peaks</u> for every anatomical landmark.[5]

3D fetal ultrasound has limited acoustic views and facial features might be occluded[3].



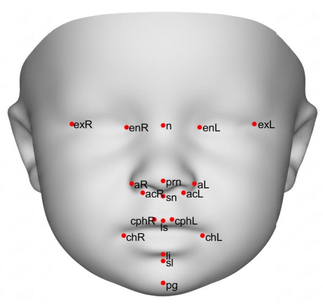
- Thus, a generalisable landmark detector should consider bias and ensure impartiality in the data.
- We collected data from different hospitals in Barcelona: Hospital del Mar, and Hospital Universitari Dexeus. **This amounted to 726 images.**
- Moreover, our data are spread across gestational ages (GAs), covering a broad range of embryonic development.

Heatmap regression suits the task better than coordinate regression to reduce false positive candidates. Essentially, it

- provides probability regions,
- captures correlation and local features,
- shows interpretability on results[5].

How do we generate ground truth?

Target landmarks are annotated by three different researchers. Invisible facial markers were completed by symmetry and visual comparison.



19 landmarks shown on perinatal face morphology.

For each landmark (d_0, h_0, w_0), a 3D Gaussian heatmap is generated:

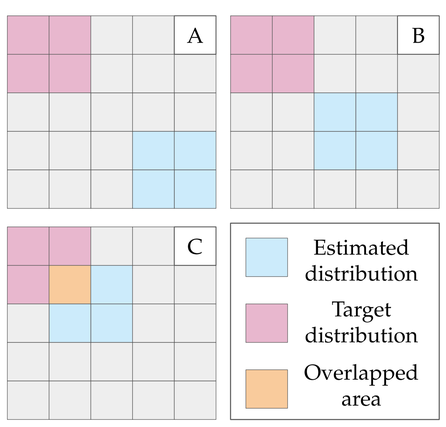
$$H(d,h,w) = \exp\left(-\frac{\sqrt{(d-d_0)^2+(h-h_0)^2+(w-w_0)^2}}{2\alpha^2}\right)$$

where d, h and w denote a voxel coordinate. $\propto$ regulates the distribution spread.

What should be the optimisation criteria?

Mean Squared Error (MSE) as a cost function works fine at the voxel level, but does not capture global spatial structure. To address this, **Earth Mover's Distance (EMD)** was introduced[6].

From A to C, the optimal transfer flow would constantly decrease. While, MSE would only decrease in Figure C. This does not account for the similarity of shapes in distributions.



$$\mathcal{E}(\mathbf{p}^l, t^l) = -\lambda_1 \mathbf{H}_{t^l} \log \mathbf{p}^l + \lambda_2 \mathbf{p}^l \mathbf{D}^{\omega}_{t^l}$$

Regularised loss | Softmax Cross-Entropy | Distance-weighted spatial enforcer

Distance matrix as EMD[2] regularisation.

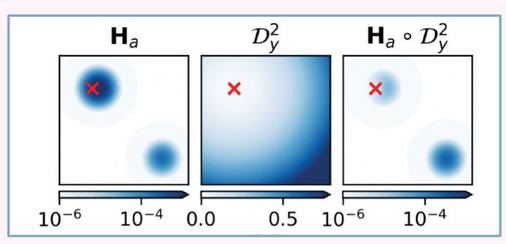


Figure is retrieved directly[6].

How to evaluate?

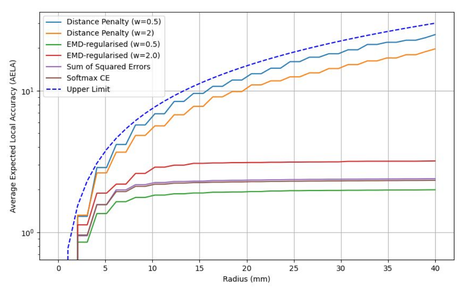
Besides mean distance, we use **Expected Local Accuracy (ELA)** that checks accuracy relative to distance to target.

We outperform the baseline study!

Our model achieves **lower error scores** than the baseline study[7] where they use Region Proposal Networks (RPNs) formulating "tiny" object detection for **only** 5 landmarks.

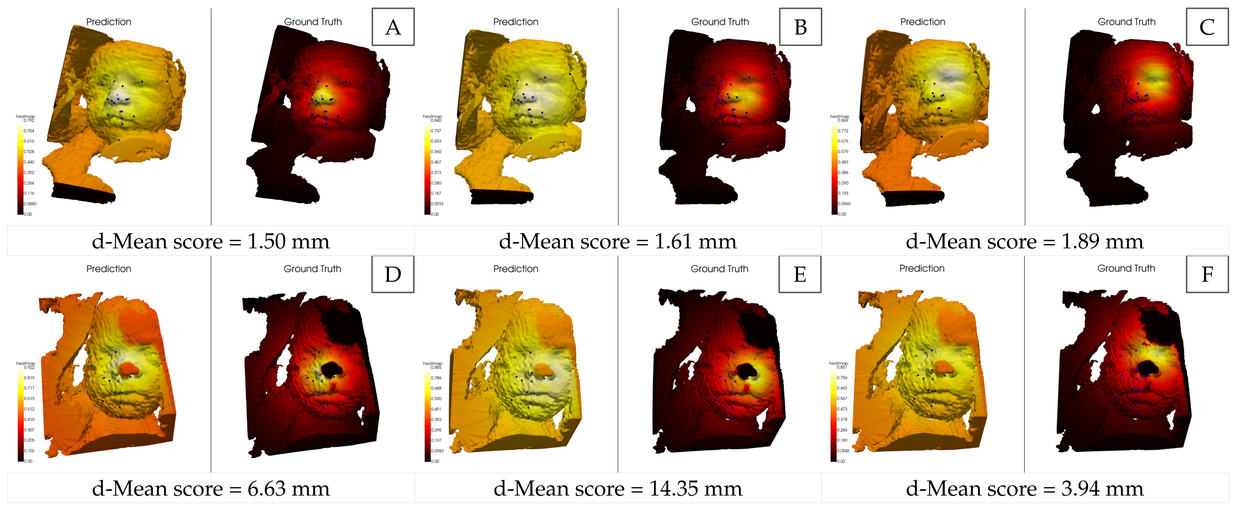
Table of overall distance scores of ResUNet3D trained with loss functions and baseline study. † marks that the setting was subject to cross-fold validation (k=4).

Setting	w	d-Mean (mm)
Softmax CE †	-	2.21
Distance Penalty	2	3.24
	0.5	2.37
Sum of Squared Errors	-	2.24
EMD-regularised †	2	3.03
	0.5	**1.81**
RPN[7]	-	4.38



Expected local accuracy curves for each setting. A random guesser limit was introduced. The chart indicates that within a 40 mm radius, the detector does not change the voxel candidate (**argmax**) in major misdirection for regularised loss functions.

What really happens in the "black box"?



A qualitative comparison of model performance on 29-weeks (A-C) and 26-weeks (D-F) fetuses. Target landmarks are prn, aL, and enL (left-to-right). The segmentations are done in 3D Slicer before visualising via PyVista.

Let's sum up.

Heatmap regression achieves low distance errors and our model can capture facial features while providing clinical interpretation. We have two **limitations**: Data is imbalanced in GA-distribution and the model is prone to error propagation from rotational alignment. **Future work** may focus on enhancing the precision and two-stage learning that includes confidence for landmark visibility.

For references, please scan the QR code!

This work is supported by the predoctoral program AGAUR-FI ajuts (2025 FI-3 00065) Joan Oró of the Department of Research and Universities of the Generalitat of Catalonia, European Social Plus Fund and Grant PID2024-161292OB-I00 funded by both ICIU/AEI/10.13039/50110 0011033 and "ERDF - A way of making Europe".