ODIN 2026 · Oral and Dental Image aNalysis · Lecture Notes in Computer Science (LNCS)

Attention Voting for 3D Cephalometric Landmark Detection

Giacomo Melacini [1] · **Carlotta Bortolami** [2] · **Chiara Gulotta** [2] · **Giovanni Badiali** [2,3] · **Giuseppe Lisanti** [1] · **Luigi Di Stefano** [1] · **Samuele Salti** [1]

[1] DISI, University of Bologna, Italy | [2] DIBINEM, University of Bologna, Italy | [2] Oral and Maxillofacial Surgery Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Italy

CODE

1. INTRODUCTION

Motivation

Landmark detection is a critical task in maxillofacial surgical planning and orthodontic diagnosis.
Traditionally done manually, which is time consuming.
Achieving high localization accuracy is challenging, particularly in 3D medical volumes.

Work and findings

We propose ***Attention Voting (AV)***, a modified cross-attention mechanism, for landmark detection.

Landmark detection as a matching task → Matching between learned landmark embeddings and spatial features → Less hyperparameter tuning → Matches or outperforms existing methods

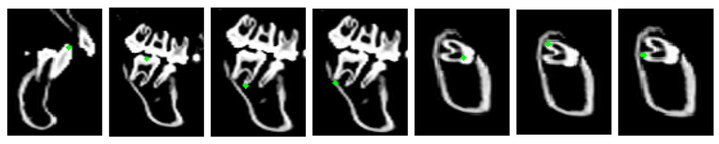
We modify a three-stage detection pipeline to make it anatomically consistent.

2. DATASET

The models are trained on two different datasets: the public **MMLD dataset** and a **private dataset** from Bologna's Sant'Orsola hospital.

MMLD

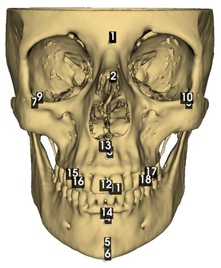
658 CT samples (458-100-100) | **Fixed size** (512 × 512 × 256) **Varying spacing** | **14** mandibular landmarks



(a) Cusp (b) 2nd C (c) 2nd R1 (d) 2nd R2 (e) 3rd C (f) 3rd R1 (g) 3rd R2

Private dataset

127 samples (100-7-20) | **83 CT 44 CBCT** | **18** cephalometric landmarks | **Dysmorphic patients Varying size Varying spacing**



Landmarks			
1	Nasion	10	Porion L
2	Sella	11	Coronal Up
3	A (subspinale)	12	Coronal Lo
4	B (supramentale)	13	Apical Up
5	Pogonion	14	Apical Lo
6	Menton	15	Molar Up R
7	Condylon R	16	Molar Lo R
8	Condylon L	17	Molar Up L
9	Porion R	18	Molar Up R

3. ATTENTION VOTING

Attention Voting enables to model landmark detection as a matching problem.

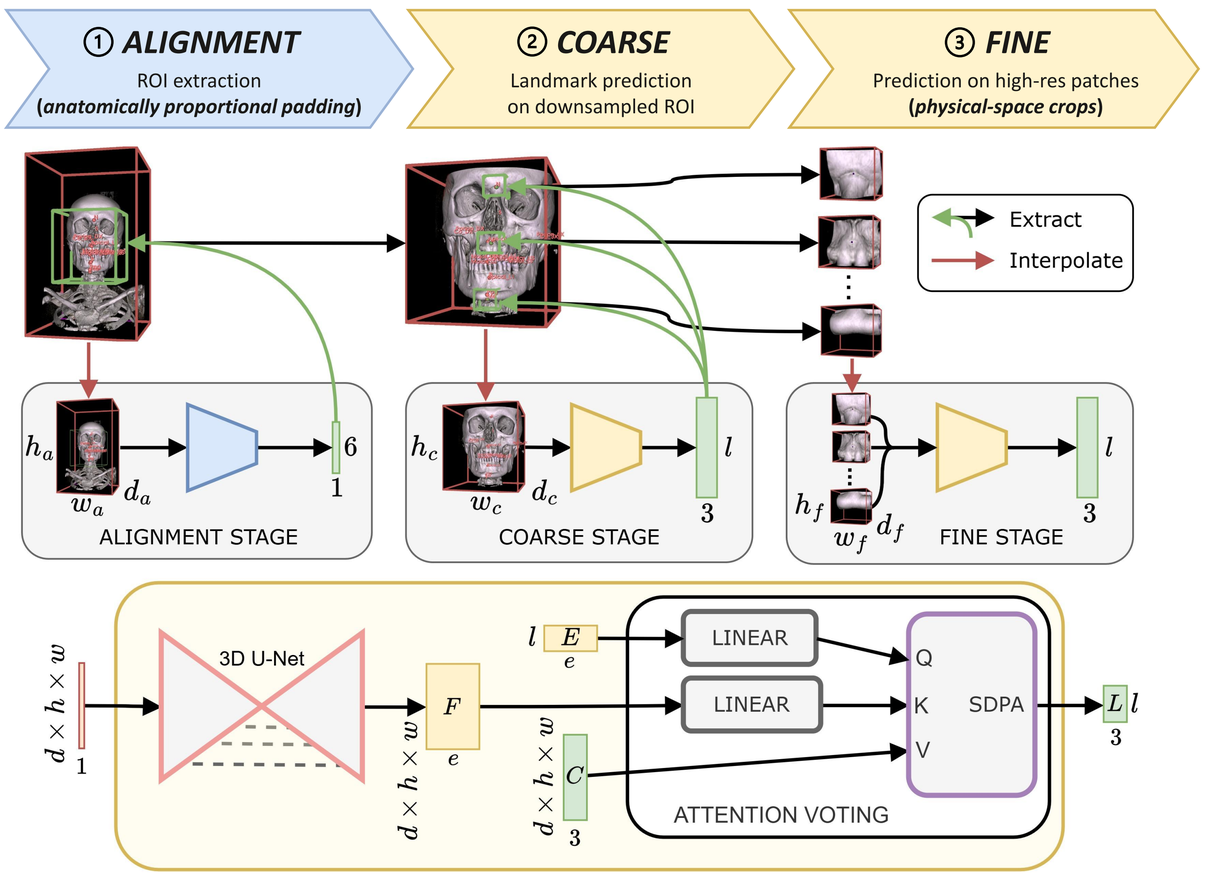
1. Learn a set of landmark specific embeddings $E \in \mathbb{R}^{l \times e}$ - one per landmark.
2. Embeddings attend to dense spatial features $F \in \mathbb{R}^{(d \times h \times w) \times e}$ to produce attention weights.
3. Landmark coordinates are predicted by aggregating the 3D spatial coordinates $C \in \mathbb{R}^{(d \times h \times w) \times 3}$ according to the attention weights.

$$L = AV(E,F,C) = softmax_{axis=-1}\left(\frac{EW^Q(FW^K)^T}{\sqrt{e}}\right) \cdot C$$

** The spatial coordinates C are not projected before computing SDPA.*
*** The output L represents the coordinates and does not need to be reprojected.*

4. PIPELINE

① ***ALIGNMENT*** ROI extraction (***anatomically proportional padding***) → ② ***COARSE*** Landmark prediction on downsampled ROI → ③ ***FINE*** Prediction on high-res patches (***physical-space crops***)



**** ROI padded proportionally to minimal bounding box → consistent spatial resolution across all samples in stages ② and ③.*

4. RESULTS

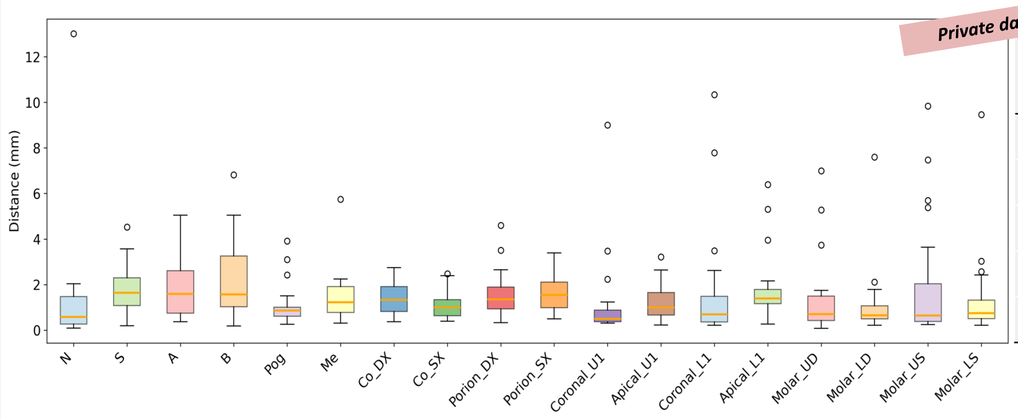
MMLD

Method	MRE ± std ↓ (mm)	SDR (%) ↑ 2mm	2.5mm	3mm	4mm
3D U-Net	2.54 ± 0.83	45.29	59.59	70.91	86.36
Two Stage	1.78 ± 0.81	68.93	82.48	88.68	95.37
ABRM	1.66 ± 0.92	79.26	88.43	92.64	**96.78**
M-GLIA	**1.52 ± 0.61**	**80.86**	**88.82**	**92.90**	95.91
AV (ours)	1.59 ± 0.64	80.47	87.21	90.83	95.79

On MMLD Attention Voting is on par with the state of the art, with a MRE gap of less than *0.1mm*.

We evaluated the proposed method on the public MMLD dataset and on the private Sant'Orsola's dataset measuring Mean Radial Error (MRE) and Successful Detection Rate (SDR) at different thresholds.

AV outperforms competing methods on the private dataset, with a significantly better SDR within *2mm*.



Private dataset

Method	MRE ± std ↓ (mm)	SDR (%) ↑ 2mm	2.5mm	3mm	4mm
3D U-Net	3.07 ± 1.8	30.83	43.06	56.94	78.06
SCN	2.47 ± 1.58	45.83	62.2	73.33	86.67
ABRM	2.47 ± 1.48	46.39	60.83	73.61	85.83
M-GLIA	2.44 ± 7.15	66.94	78.33	84.17	**91.67**
AV (ours)	**1.84 ± 3.38**	**78.61**	**82.50**	**86.11**	**91.67**

5. ABLATION STUDIES

Ablations have been carried out to assess the contribution of AV and validate the anatomically consistent pipeline.

To isolate the contribution of AV, we performed a **head replacement experiment** in which the coarse and fine stages of the model were trained with a heatmap regression (HR) head.

MMLD

Anatomical Alignment	Consistent Fine Crop	Head	MRE ± std (mm)	SDR (%) 2mm	2.5mm	3mm	4mm
✗	✗	AV	1.63 ± 0.63	79.55	86.87	91.16	**96.04**
✗	✓	AV	1.62 ± 0.62	78.96	**87.63**	**91.75**	95.54
✓	✓	AV	**1.59 ± 0.64**	**80.47**	87.21	90.83	95.79
✓	✓	HR	2.06 ± 2.29	77.02	85.35	88.72	92.51

Both Attention Voting and the physically consistent pipeline contribute to improving the model's overall performance.

6. LIMITATIONS & FUTURE WORK

Current limitations

1. Outlier predictions remain a limitation, preventing full automation of the task in clinical practice.
2. The model is unable to identify the absence of anatomical landmarks, a situation that is not uncommon in clinical settings.

Future work

1. Modify the Attention Voting mechanism to enable the recognition of absent landmarks.
2. Exploit the flexibility of Attention Voting to design a predictor capable of detecting an arbitrary number of landmarks, specified at test-time by a clinical expert.

ACKNOWLEDGEMENTS

We acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support.

REFERENCES

1. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. pp. 424–432. Springer International Publishing, Cham (2016)
2. Zhu, H., Li, X., He, T., Liu, B., Tang, W., Guo, J.: Enhanced landmark detection in oral and maxillofacial ct and cbct scans: A multistage global–local integration approach. IEEE Transactions on Instrumentation and Measurement 74, 1–17 (2025).
3. He, T., Xu, G., Cui, L., Tang, W., Long, J., Guo, J.: Anchor ball regression model for large-scale 3d skull landmark detection. Neurocomputing 567, 127051 (2024).
4. He, T., Yao, J., Tian, W., Yi, Z., Tang, W., Guo, J.: Cephalometric landmark detection by considering translational invariance in the two-stage framework. Neurocomputing 464, 15–26 (2021).
5. Payer, C., Štern, D., Bischof, H., Urschler, M.: Integrating spatial configuration into heatmap regression based cnns for landmark localization. Medical Image Analysis 54, 207–219 (2019).