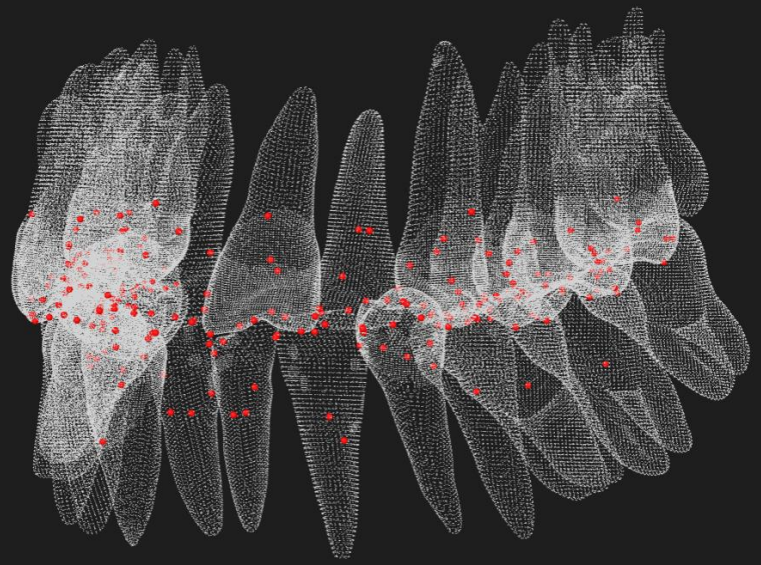
POLITECNICO
MILANO 1863



Bite2Text

ODIN 2026 Challenge!

Algorithm: *Landmarko*

The patient presents a ____ crossbite on the ____ side. From a vertical standpoint, the overbite is ____. The molar relationship is Class ____ on the right and Class ____ on the left, with overjet ____ . The dental midlines are ____. The Curve of Spee is ____ and the Curve of Wilson is ____. Crowding in the upper arch is ____ and in the lower arch is ____.

Structured Slot Prediction and Template-Based Synthesis for Orthodontic Report Generation from Intraoral Scans and Photographs

Ali Shadman Yazdi
PhD candidate

Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico

MICCAI2026

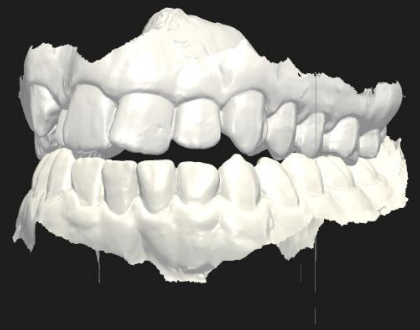
UNIMORE CRNS Radboudumc

Bite2Text TOOTH FAIRY4

odin-workshops.org MICCAI2026 MICCAI2026

The Task: Scans + Photos → Diagnostic Text

3D Intra Oral Scans (IOS)



- Upper + lower 3D mesh per patient
- Shape only — occlusal / bite geometry
- No color, no soft-tissue information

Segmentation & Landmark localization

↓

Clinical metric extraction

↓

Classifier

↓

Text generation

3D

~83 words

Average scan-only report
occlusal findings alone

2D Intra Oral Photographs

- Five 2D photos per patient
- Same bite relationship, visible again
- Plus soft tissue: gums, plaque, fillings, stains

Feature extraction (Res-Net 18)

↓

Classifier

↓

Text generation

2D

~122 words

Average photo report
Occlusal + soft-tissue findings

Turning Free Text into Structured Labels

989 patients total (After outlier removal)
→ split 791 train / 198 frozen test.

1. Seed with an LLM

- **~200 reports** per report-family read by Claude Sonnet
- Converted to JSON slot values → per-slot class labels
- Fixes the slot schema and gives a hand-labeled gold set

2. Deterministic regex parser

- Fixed, hand-written rules, no learned component
- Matches each word to its nearest finding keyword
- Checks negation;
- Written against the gold set, then run over the full corpus
- Applied on the rest of the training set (~591 reports) to get GT labels.

3D

15 labels extracted from the scan reports. Parser agreement is how often the extracted label matches the gold set.

Field	Parser agreement (recall)
Deep bite	100.0%
Midlines (coincident/deviated)	98.5%
Curve of Wilson	98.4%
Crossbite (presence)	97.4%
Midline magnitude	96.6%
Overjet	95.4%
Curve of Spee	94.0%
Overbite	90.7%
Crowding, lower arch	88.5%
Transverse constriction	87.6%
Crowding, upper arch	87.1%
Molar class, left	75.9%
Molar class, right	61.3%
Canine class, left	56.5%
Canine class, right	56.3%

2D

10 labels extracted from the photographic reports.

Field	Parser agreement (recall)
Plaque	100.0%
Calculus	100.0%
Hygiene summary	100.0%
Discoloration	100.0%
Recession	100.0%
Sealants	100.0%
Restorations	99.0%
Gingivitis	89.7%
Caries	87.9%
Enamel defect	100.0%

Ali Shadman Yazdi

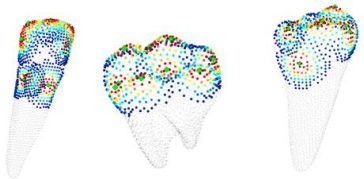
The 3D head – Segmentation, Landmark detection, and measurements

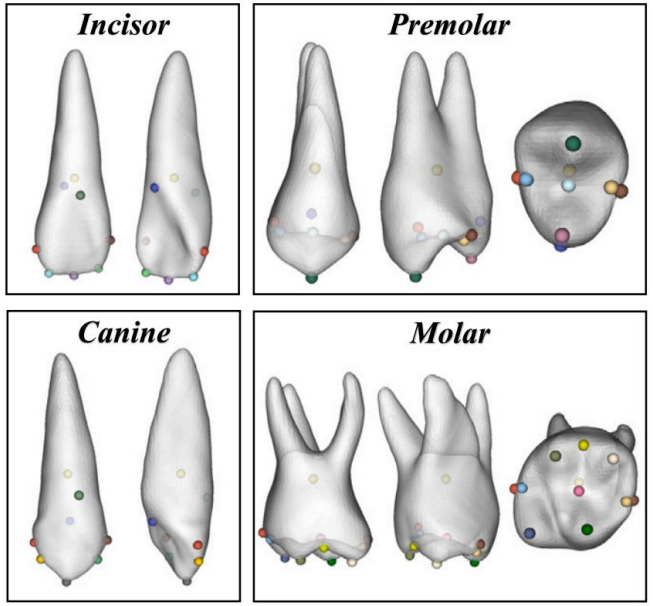
Segmentation
TriDent-GCA [1] + TSGCNet[2] on 3DTeethLand (900 arch pairs), 95.1% / 93.5% point accuracy

Landmark Detection
CLIK-Diffusion (PointMLP)[3], per-tooth model, up to 22 landmarks/tooth

Anatomical Frame
Pose-invariant frame rebuilt per patient from the landmarks

25 Measurements
7 groups: sagittal, transverse, midline, arch width, alignment, curves, dentition

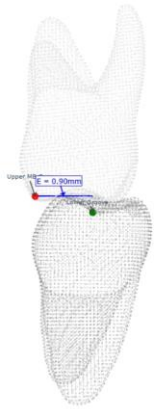




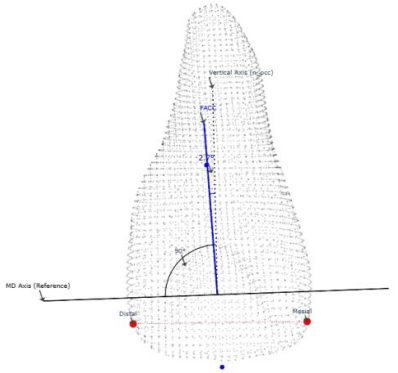
~ 120 landmarks per jaw *Image from [3]*

Andrews' Six Keys to Normal Occlusion[4]:

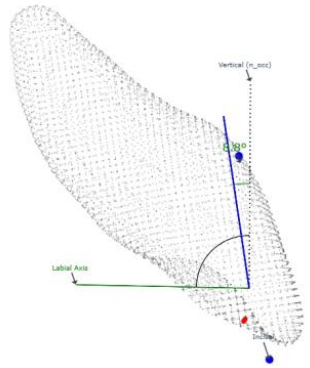
Key I: Molar Relationship



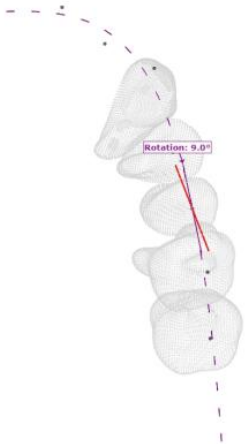
Keys II: Crown Angulation



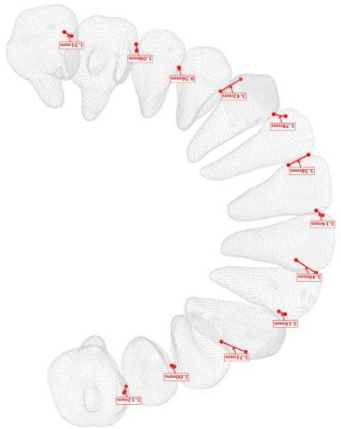
Keys III: Crown Inclination



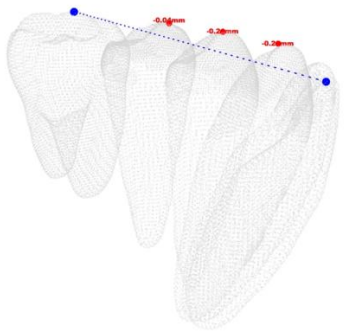
Key IV: Rotations



Key V: Tight Contacts



Key VI: Curve of Spee



The 3D head – Slot prediction

Segmentation
TriDent-GCA [1] + TSGCNet[2] on 3DTeethLand (900 arch pairs), 95.1% / 93.5% point accuracy

↓

Landmark Detection
CLIK-Diffusion (PointMLP)[3], per-tooth model, up to 22 landmarks/tooth

↓

Anatomical Frame
Pose-invariant frame rebuilt per patient from the landmarks

↓

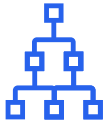
25 Measurements
7 groups: sagittal, transverse, midline, arch width, alignment, curves, dentition

Group	Measurement	Definition	Values
Sagittal	Overjet	distance between the upper/lower central incisors' labial surfaces	1
	Overbite	vertical distance between the upper/lower central incisors' incisal edges	1
	Molar & canine relation	anteroposterior offset to the antagonist in the matching quadrant, right and left	4
Transverse	Molar & premolar offset	sideways offset to the antagonist in the matching quadrant, right and left	4
Midline	Midline deviation	sideways offset between the upper/lower incisors' contact midpoints	1
Arch width	Inter-molar & inter-premolar width	distance between left and right antimeres within a quadrant pair, upper and lower	4
Alignment	Little's Index	sum of gaps between adjacent anterior contact points, upper and lower	2
	Arch crowding	tooth-width sum minus fitted arch-curve length, upper and lower	2
Vert. curves	Curve of Spee	vertical spread of the posterior cusp tips, upper and lower	2
	Curve of Wilson	mean buccal–lingual cusp-height offset, upper and lower	2
Dentition	Tooth count	number of teeth detected present, upper and lower	2
		Total	25

Table 2. The 25 geometric measurements, computed between named crown landmarks.

791 train patients:
- 154 cases used to extract the gold set. (used also as a test set)
- 637 cases used for model fitting. Labels extracted from the Parser

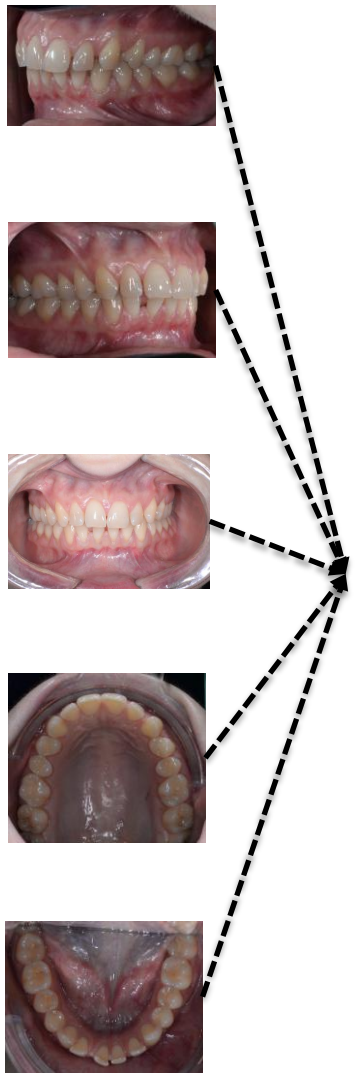
Shallow Decision Tree
Fit only on the features that directly define the slot
learns the clinical threshold



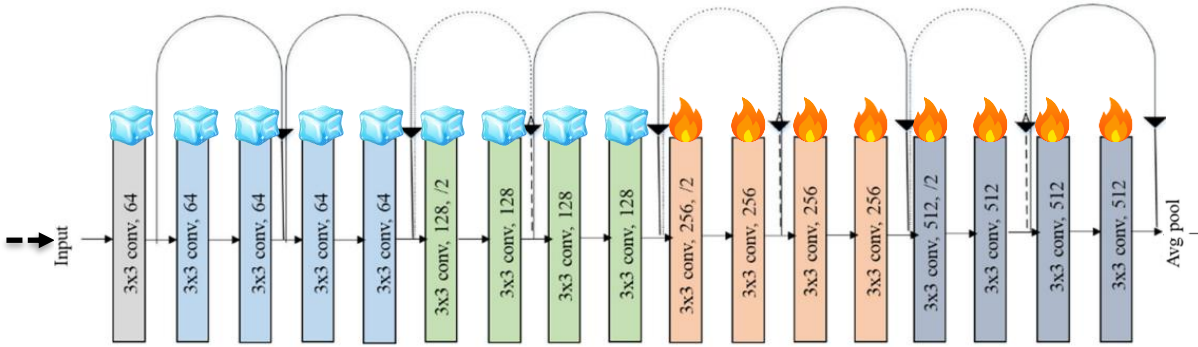
Gradient-Boosted Trees
Fit on all 25 features
Used when a slot isn't determined by any single measurement

- Midlines (coincident/deviated)
- Crossbite (presence)
- Midline magnitude
- Crowding, (U/L) arch
- Deep bite
- Transverse constriction
- Molar class (left/Right)
- Canine class (left/Right)

The 2D head



Preprocessing
Orientation-corrected, resized, ImageNet-normalized.
~Augmentation



LR: 10^{-4}

- Plaque
- Calculus
- Hygiene summary
- Discoloration
- Recession
- Sealants
- Gingivitis
- Caries
- Enamel defect
- Restorations

10 Linear Heads
- *LR: 10^{-3}*
- *Dropout p=0.3*

Five photos per patient go through one shared, ImageNet-pretrained ResNet-18, and the five feature vectors are averaged into one. Ten linear heads then predict the soft-tissue findings from that vector.

791 train patients:
- 93 cases dropped. No 2D photographics66 report
- 200 cases used to extract the gold set. (used also as a test set)
- 498 cases used for model fitting. Labels extracted from the Parser

Results – 2D/3D head

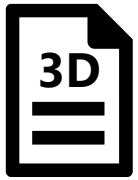


Table 3. 3D slot accuracy on 154 held-out patients. Baseline is the majority-class accuracy for that slot. "Classes" is the number of distinct values observed in the training data.

Slot	Classes	Baseline	Model	Shipped	Constant value
Canine class, left	3	37.8%	**70.9%**	model	—
Transverse constriction	4	32.1%	**59.7%**	model	—
Molar class, left	3	55.2%	**82.5%**	model	—
Overjet grade	5	28.9%	**54.6%**	model	—
Overbite grade	6	29.8%	**49.6%**	model	—
Canine class, right	3	48.6%	**67.1%**	model	—
Molar class, right	3	58.8%	**66.9%**	model	—
Curve of Wilson	3	56.0%	**61.0%**	model	—
Crossbite present	2	74.3%	**75.7%**	model	—
Crowding, upper	6	64.0%	64.0%	constant	mild
Midline deviation	3	75.3%	74.0%	constant	deviated
Crowding, lower	6	50.0%	46.6%	constant	mild
Curve of Spee	7	61.5%	57.4%	constant	increased
Deep bite	2	87.0%	68.8%	constant	false
Midline magnitude	1	100.0%	100.0%	constant	slight

Table 4. Photo-head accuracy on 200 held-out patients.

	Raw accuracy		Balanced accuracy			
Finding	Base	Model	Base	Model	Shipped	Constant value
Gingivitis	42.0%	**48.0%**	25.0%	**45.8%**	model	—
Restorations	51.0%	**59.0%**	33.3%	**47.9%**	model	—
Recession	74.5%	**82.5%**	50.0%	**83.1%**	model	—
Plaque	73.5%	61.0%	50.0%	58.4%	prior	present
Calculus	79.0%	66.5%	50.0%	56.9%	prior	present
Oral hygiene	92.0%	88.5%	50.0%	56.7%	prior	poor
Sealants	72.0%	67.0%	50.0%	61.3%	prior	present
Caries	68.5%	53.5%	33.3%	38.4%	prior	suspected
Discoloration	63.0%	58.0%	25.0%	30.7%	prior	stain
Enamel defect	94.5%	94.0%	25.0%	24.9%	prior	hypoplasia

Results – Challenge Score

Predicted:

The patient presents a transverse constriction of the maxilla in the absence of crossbite. From a vertical standpoint, the overbite is within normal limits. Sagittally, there is a bilateral end-to-end Class II molar and canine relationship with overjet within normal limits. The midlines are slightly deviated relative to each other. The Curve of Spee and the Curve of Wilson are increased. Mild crowding is present in the upper and lower arches. Restorations are present. Fissure sealants are present. The grooves and interproximal contact points should be evaluated for the possible presence of caries/secondary caries. Diffuse extrinsic staining is present. Enamel hypomineralization is present. Buccal gingival recessions are present. The gingivae are inflamed.

Configuration	BLEU-4 [5]	METEOR [6]	Captioning
Constant baseline	0.2411	0.4425	0.3418
3D slots only	0.1760	0.3958	0.2859
+ photo CNN, benched heads silent	0.2194	0.4310	0.3252
+ photo CNN, benched heads → prior	**0.2662**	**0.4594**	**0.3628**
Occlusal + prior alone (no photos read)	0.2735	0.4662	0.3699

Ground truth:

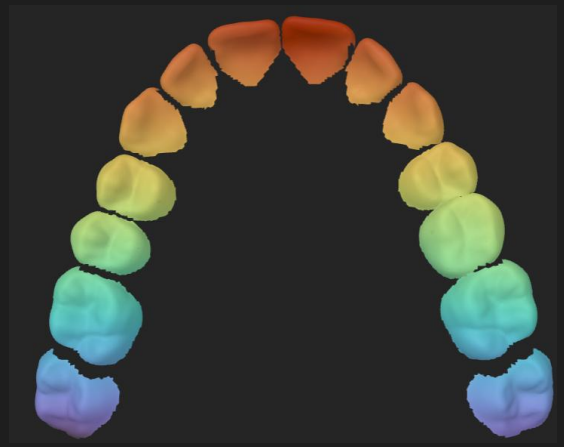
The patient presents a transverse constriction of the maxilla in the absence of crossbite, while from a vertical standpoint the overbite is within normal limits. Sagittally, there is a Class I molar relationship and a canine relationship of Class II edge-to-edge on the right, while on the left the molar relationship cannot be assessed due to the absence of tooth 36 (root fragments remain), whereas the canine relationship is Class II edge-to-edge. The overjet is increased, but there is an anterior crossbite on 12. The dental midlines are not coincident. The Curve of Spee and the Curve of Wilson are increased. Severe crowding in the upper and lower arches. The grooves of premolars and molars are pigmented with possible presence of occlusal and/or interproximal caries; in particular, 16-46-47 should be evaluated. There are buccal gingival recessions on teeth 13-23-24-25-26-35-45-46-47. Visible plaque and calculus are present with inflamed gingivae.

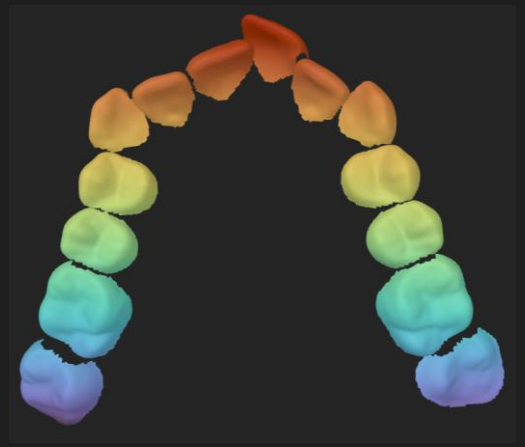
Ranked 28th out of 41 on the Test Phase Leaderboard

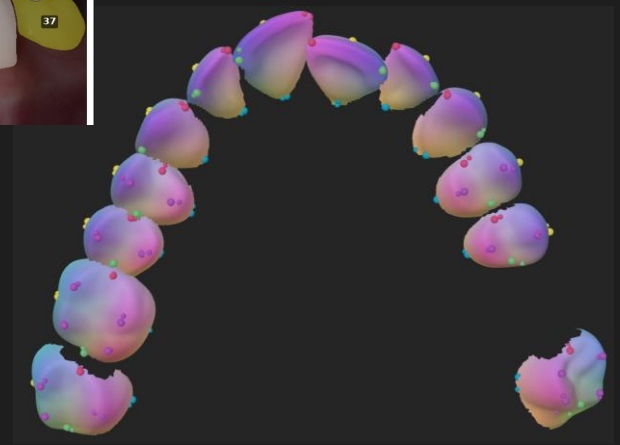
Conclusions and Future Work

Limitations and future Work:

- Template based dental models.
- Recovering failed slots with clinical knowledge-based metrics
- Per tooth analysis
- Using LLMs to turn the same structured slot values into more natural, varied wording







Advantages

- The pipeline is fully modular, different slots can be added or removed.
- Outputs are measurements not black box predictions (3D case only)

Thank you!

References:

[1] Lazarevski, M., Shadman, A., Baselli, G.: A multi-head heatmap architecture for full-mouth 3d dental landmark detection. In: 13th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN). pp. 28–33 (2026). https://doi.org/10.69994/13Ic26006

[2] Zhang, L., Zhao, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D.: Tsgcnet: Discriminative geometric feature learning with two-stream graph convolutional network for 3d dental model segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6695–6704. IEEE (2021)

[3] Dou, Y., Wu, H., Li, C., Wang, C., Yang, T., Zhu, M., Shen, D., Cui, Z.: Clikdiffusion: Clinical knowledge-informed diffusion model for tooth alignment. Medical Image Analysis 106, 103746 (2025). https://doi.org/https://doi.org/10.1016/ j.media.2025.103746, https://www.sciencedirect.com/science/article/pii/ S1361841525002932

[4] Andrews, L.F.: The six keys to normal occlusion. American Journal of Orthodontics 62(3), 296–309 (1972). https://doi.org/https://doi.org/10.1016/ S0002-9416(72)90268-0, https://www.sciencedirect.com/science/article/ pii/S0002941672902680

[5] Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Isabelle, P., Charniak, E., Lin, D. (eds.) Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. pp. 311–318. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA (Jul 2002). https://doi.org/10.3115/1073083.1073135, https: //aclanthology.org/P02-1040/

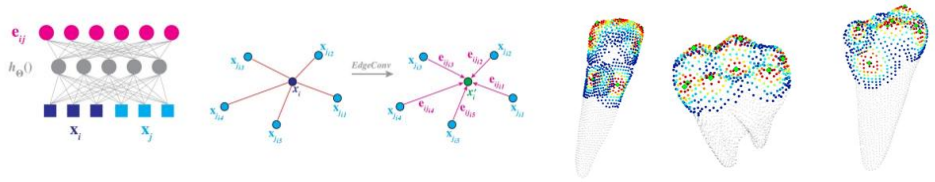
[6] Banerjee, S., Lavie, A.: METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Goldstein, J., Lavie, A., Lin, C.Y., Voss, C. (eds.) Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. pp. 65– 72. Association for Computational Linguistics, Ann Arbor, Michigan (Jun 2005), https://aclanthology.org/W05-0909/

APPENDIX

Automated Orthodontic Diagnosis via Geometric Deep Learning

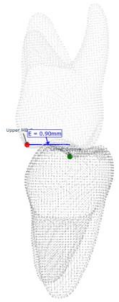
***Abstract*—Precise landmark detection on 3D dental models is a prerequisite for automated orthodontic diagnosis, yet existing methods often rely on computationally heavy models or lack direct clinical interpretability. In this work, we propose a lightweight, end-to-end framework that bridges geometric deep learning with established clinical standards.**

A modified DGCNN to capture 3D geometric topology through dynamic graph updates. To help the network distinguish between similarly shaped teeth, we fuse these geometric features by injecting a learnable tooth-ID token into the prediction head.

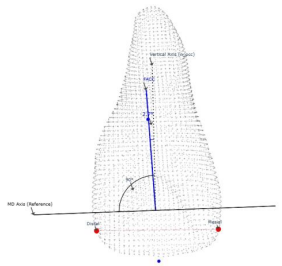


Andrews' Six Keys to Normal Occlusion:

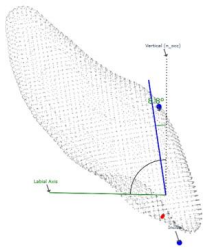
Key I: Molar Relationship



Keys II: Crown Angulation



Keys III: Crown Inclination

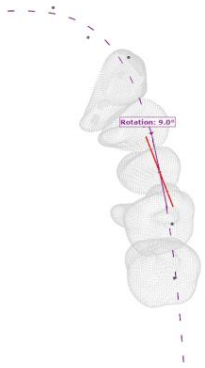


Network Architecture	Model Size (MB)	Inference Time (ms / tooth)
PointMLP (used in CLIK-Diffusion [9])	206.34	575.0
Dense Representative Net [11]	5.40	**415.0**
Tokenized DGCNN (Ours)	**4.63**	493.0

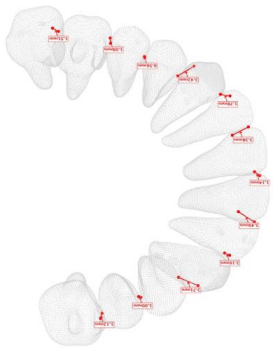
TABLE I: Landmark Detection Accuracy (Mean Euclidean Error in mm) using Top-K Weighted Average ($K = 10$). Comparison against Dense Representative Network across High-fidelity (CLIK), Clinical (TADPM), and Pathological (Teeth3DS) datasets.

	CLIK-Diffusion Dataset [9] (CBCT Segemented Meshes)		TADPM Dataset [14] (IOS)		Teeth3DS [15] (IOS)	
Tooth Class	**Dense Rep. [11]**	**Tok. DGCNN**	**Dense Rep. [11]**	**Tok. DGCNN**	**Dense Rep. [11]**	**Tok. DGCNN**
Incisors	0.764	**0.608**	**0.696**	0.699	1.200	**1.146**
Canines	0.534	**0.508**	0.666	**0.649**	**1.572**	1.762
Premolars	0.450	**0.448**	0.601	**0.600**	**2.177**	2.361
Molars	0.504	**0.471**	**0.741**	0.774	**3.009**	3.283

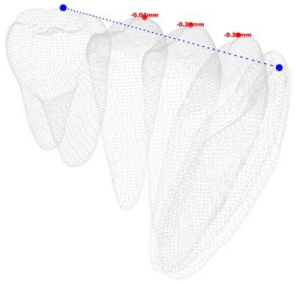
Key IV: Rotations

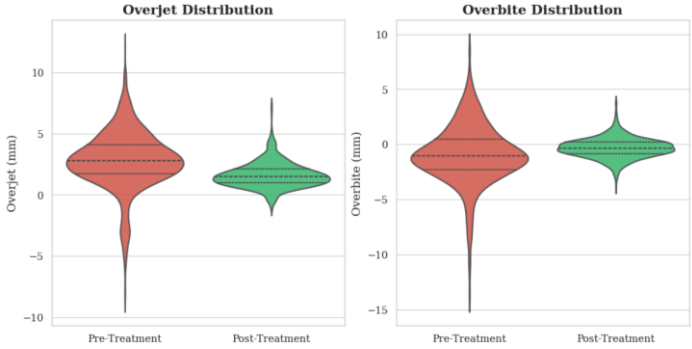


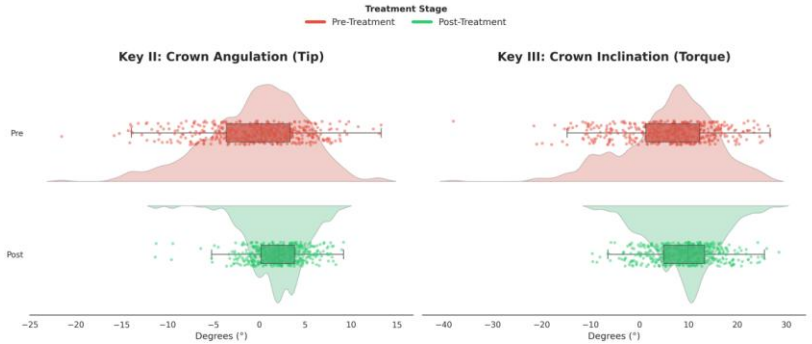
Key V: Tight Contacts

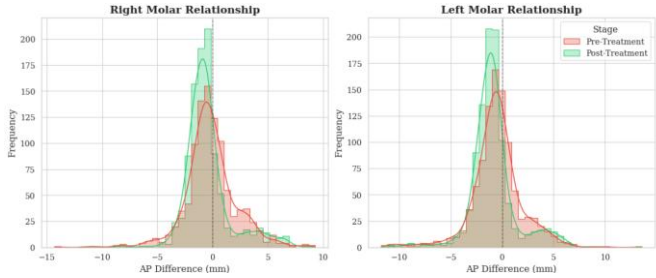


Key VI: Curve of Spee









A Multi-Head Heatmap Architecture for Full-Mouth 3D Dental Landmark Detection

The study proposes a multi-head deep learning architecture that processes full-mouth 3D intraoral scans to jointly predict tooth presence, fixed-count landmarks, and variable-count cusp tips in a single forward pass. A Multi-Head Heatmap Architecture for Full-Mouth 3D Dental Landmark Detection

13th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN), Silver Lake, 8-11 June 2026

A Multi-Head Heatmap Architecture for Full-Mouth 3D Dental Landmark Detection

Marko Lazarevski
DEIB
Politecnico di Milano
Milan, Italy
marko.lazarevski@mail.polimi.it
orcid.org/0009-0003-7689-5576

Ali Shadman Yazdi
DEIB
Politecnico di Milano
Milan, Italy
ali.shadmanyazdi@polimi.it
orcid.org/0009-0002-8317-8332

Giuseppe Baselli
DEIB
Politecnico di Milano
Milan, Italy
giuseppe.baselli@polimi.it
orcid.org/0000-0003-2978-1704

***Abstract*—Accurate 3D dental landmark detection is a foundational step in automated orthodontic planning, aligner design, and occlusal analysis. We present a comparison of three lightweight point-cloud encoders, hierarchical PointNet, PointNeXt, and PointViG, equipped with a unified multi-head architecture for joint full-mouth tooth presence prediction, fixed-count landmark localization, and variable-count cusp tip detection. Both arches are merged into a single full-mouth point cloud that preserves inter-arch geometry. The landmark head uses learned cross-attention queries indexed by tooth slot and landmark category, decoded via a differentiable soft-argmax. Variable-count cusp tips are handled by a dedicated heatmap channel decoded by k-means clustering, where the number of clusters is provided by an auxiliary count regression head. Evaluated under 5-fold cross-validation on the 3DTeethLand benchmark, PointViG achieves a mean radial error of 1.78 mm and both mean average precision and mean average recall of 0.548, reaching competitive leaderboard performance without ensembling or per-tooth spatial conditioning. Among the three backbones, PointViG's dynamic graph convolution in feature space proves decisive: its ability to aggregate information from semantically related but spatially distant points yields consistent gains across all landmark (missing, extra, or partially erupted teeth), distinguishing anatomically relevant landmarks per tooth type, and balancing multiple supervision signals during training.**

To date, IOS landmarking methods have predominantly targeted tooth-level or single-arch settings, often within benchmarks like Teeth3DS+/3DTeethLand [5] that define the task at the level of individual jaws. While several frameworks process entire intraoral scans and then crop or segment them into tooth-wise regions [4], [6], to the best of our knowledge no prior work reports a single end-to-end model operating on a full-mouth IOS (both arches in occlusion) to jointly predict tooth presence and per-tooth landmarks in one multi-head forward pass. Full-mouth input preserves inter-arch geometry — the spatial arrangement of teeth across both jaws — which is a powerful contextual cue that isolated tooth-level models discard.

This paper makes the following contributions:

- A unified full-mouth IOS multi-head architecture operating over a fixed schema of 32 tooth slots, jointly predicting tooth presence, fixed-count landmark

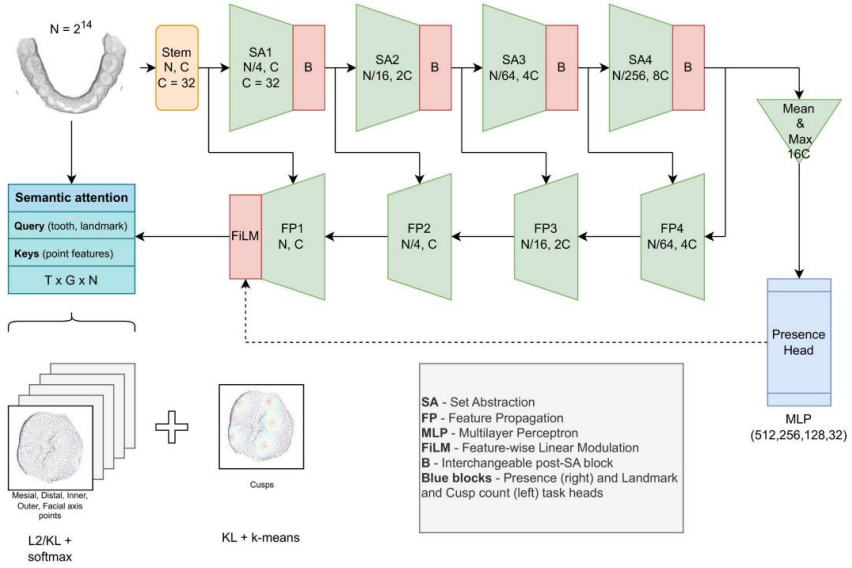


Fig. 1. Overview of the shared multi-head baseline architecture. A stem MLP lifts raw coordinates to C = 32 features, followed by 4 Set Abstraction stages (encoder) and 4 Feature Propagation stages (decoder) with skip connections. The interchangeable post-SA block (B) is the only component that varies across PointNet++, PointNeXt, and PointViG. Three task heads operate on the decoded per-point features and global descriptor.

TABLE I. OVERALL TEST METRICS (5-FOLD CV MEAN) ON 3DTEETHLAND. MRE IN MM (↓), SDR IN % (↑), CUSP COUNT MEAN ABSOLUTE ERROR (MAE, ↓).

Backbone	MRE	$SDR_{1\,mm}$	$SDR_{2\,mm}$	Cusp MRE	Cnt MAE
PointNet++	*2.07*	*30.3*	*68.6*	*2.15*	*0.33*
PointNeXt	*2.10*	*30.3*	*67.3*	*2.21*	*0.34*
PointViG	1.78	42.2	78.5	1.83	0.32

TABLE II. PER-TOOTH-GROUP MEAN RADIAL ERROR (MM, ↓) ON 3DTEETHLAND (5-FOLD CV MEAN).

Group	PointNet++	PointNeXt	PointViG
Incisors	*1.81*	*1.90*	1.69
Cuspids	*1.91*	*1.95*	1.69
Premolars	*1.92*	*1.91*	1.61
Molars	*2.63*	*2.59*	2.07
Anterior	*1.84*	*1.92*	1.69
Posterior	*2.24*	*2.22*	1.82

TABLE IV. COMPARISON AGAINST THE 3DTEETHLAND CHALLENGE LEADERBOARD [5], SORTED BY MAP. OUR BASELINE (POINTVIG) IS REPORTED AS 5-FOLD CV MEAN ± STD.

Rank	Team	mAP	mAR
1	Radboud	0.785	0.656
2	ChohoTech	0.775	0.634
3	YY-LAB	0.719	0.579
4	YN-LAB	0.643	0.527
5	IGIP-LAB	0.590	0.466
6	3DIMLAND	0.578	0.438
—	PointViG (ours)	0.548 ±0.036	0.548±0.034

Thank you!