DiRecT: Diagnosis and Reconstruction Transformer for Mandibular Deformity Assessment

Source code

Xuanang Xu[1], Jungwook Lee[1], Nathan Lampen[1], Daeseung Kim[2], Tianshu Kuang[2], Hannah H. Deng[2], Michael A. K. Liebschner[3], Jaime Gateno[2], and Pingkun Yan[1] ✉

1. *Department of Biomedical Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*
2. *Department of Oral and Maxillofacial Surgery, Houston Methodist Research Institute, Houston, TX 77030, USA*
3. *Department of Neurosurgery, Baylor College of Medicine, Houston, TX 77030, USA*

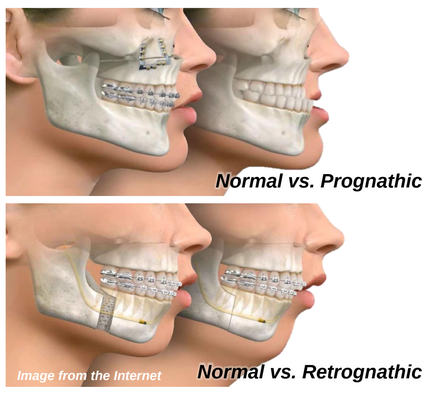
Corresponding author:
Pingkun Yan (yanp2@rpi.edu)

MICCAI 2024 Marrakesh MOROCCO

MICCAI

Introduction

Accurate mandibular deformity diagnosis is crucial for orthognathic surgery but often relies on the clinician's experience, introducing subjectivity and variability. Traditional methods[1,2] using specific bony anatomical landmarks oversimplify the complex facial structures and can be inconsistent. Machine learning approach[3] requires precise segmentation of bony structures from CBCT images, which is labor-intensive, time-consuming, and exposes patients to radiation.



We propose a ***Diagnosis-Reconstruction Transformer*** (**DiRecT**) for diagnosing mandibular deformities with two key contributions:

1. **Simplified process**: Instead of using bony landmarks, we utilize facial soft tissue landmarks that can be easily detected by off-the-shelf models, streamlining the diagnostic process.
2. **Innovative DiRecT network**: Our DiRecT network integrates landmark reconstruction within a teacher-student framework. This reduces reliance on labeled data and achieves performance comparable to even better than traditional methods, while significantly simplifying the diagnostic process.

Method

A. 3D Facial Landmark Extraction using Adapted MediaPipe[4] Model

We introduce a workflow that adapts an off-the-shelf 2D facial landmark detection model (Google's ***MediaPipe***[4]) for 3D facial landmark extraction from CBCT/3dMD images. This is achieved by projecting 3D facial surfaces into 2D images and using ray casting to back-project detected 2D landmarks onto the 3D surface. A final set of **328 stable landmarks** is selected based on detection reliability across subjects.

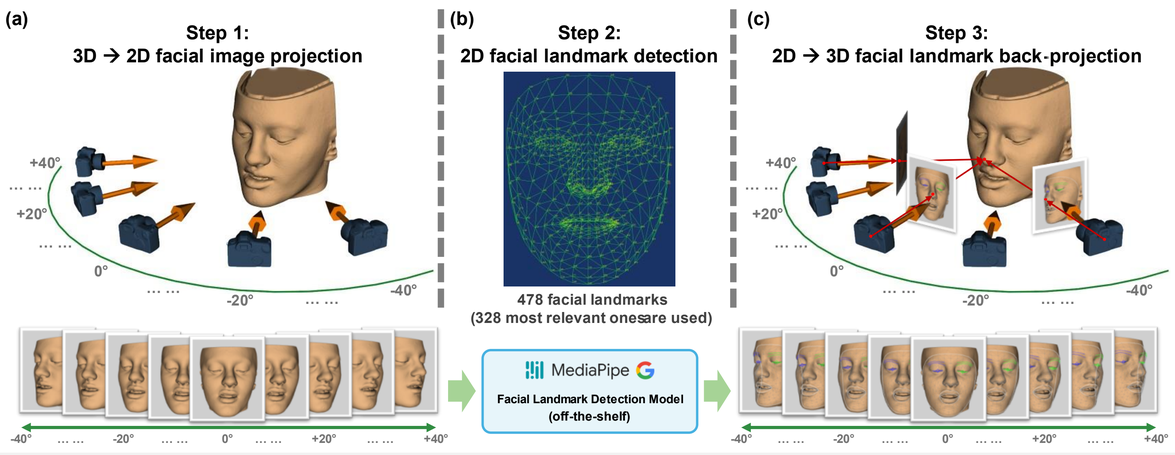


Fig. 1. 3D facial landmark extraction through 2D facial landmark detection model.

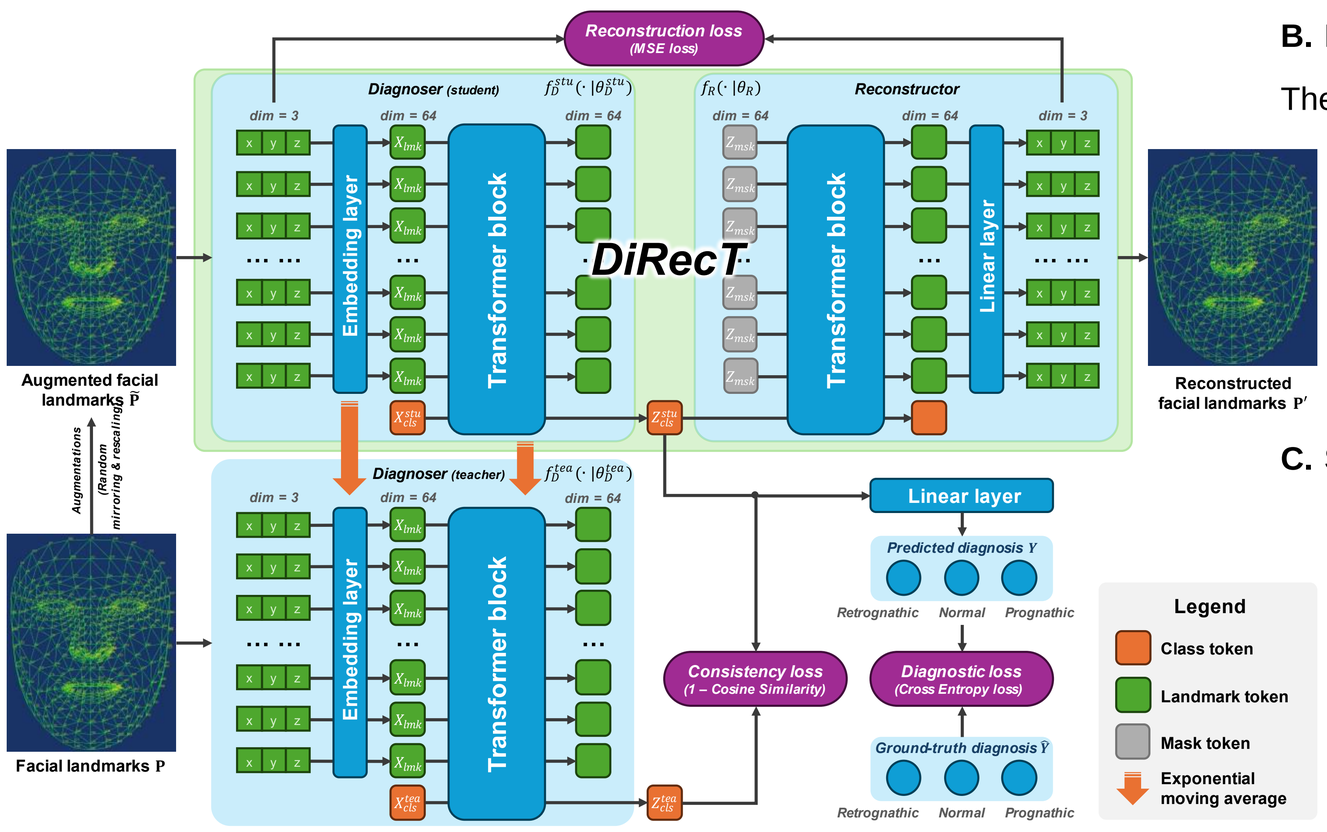


Fig. 2. Scheme of the DiRecT network and teacher-student training framework.

B. DiRecT Network for Mandibular Deformity Diagnosis

The proposed DiRecT network consists of two transformer[5]-based components:

- **Diagnoser:** Takes 3D facial landmarks as input and generates a class token representing the mandibular deformity status (*normal*, *retrognathic*, *prognathic*).
- **Reconstructor:** Uses the class token to reconstruct the 3D facial landmarks, enforcing the class token to encode comprehensive geometric information.

C. Semi-Supervised Learning with Teacher-Student Diagnoser

- To leverage both labeled and unlabeled data, we implement a teacher-student framework, where the teacher network guides the student network using consistency constraint, calculated between class tokens of original and augmented landmarks. This enables training on unlabeled data, expanding the dataset and improving performance.

$$L = L_{diag} + L_{reco} + \lambda \cdot L_{cons}$$

- The overall training objective combines *diagnostic loss* L_{diag}, *reconstruction loss* L_{reco}, and *consistency loss* L_{cons} with a weight λ for the consistency loss increasing linearly during training to avoid instability.

Results

A. Datasets and Metric

We used two datasets: an in-house clinical dataset of **101 subjects** with CBCT images labeled by a senior surgeon into three mandibular deformity categories (*normal*, *retrognathic*, *prognathic*), and the Headspace dataset[6] with **1,519 subjects** of unlabeled 3D head images (**917 subjects** within suitable age range were used). Using our pipeline, we extracted **328 facial landmarks** per subject. We used the Headspace data as unlabeled data for semi-supervised training and conducted 4-fold cross-validation on the clinical dataset, evaluating the model's performance using classification accuracy.

B. Comparison with Other Methods

Methods	Input data	Accuracy [%]			
		Normal	*Retrognathic*	*Prognathic*	**All**
SNB angle	**Bony** landmark	42.11	73.68	88.64	74.26
Facial angle	**Bony** landmark	36.84	84.21	81.82	74.26
MdUL	**Bony** landmark	21.05	78.95	95.45	75.25
MLP[3]	**Bony** landmark	**47.37**	**89.47**	**97.73**	**85.15**
GCN[7]	**Facial** landmark	63.16	84.21	81.82	79.21
GAT[8]	**Facial** landmark	**68.42**	86.84	72.73	77.23
SGC[9]	**Facial** landmark	57.89	84.21	84.09	79.21
GTransformer[10]	**Facial** landmark	63.16	84.21	84.09	80.20
DiRecT (ours)	**Facial** landmark	57.89	**92.11**	**86.36**	**83.17**

C. Ablation Study

Models	Accuracy [%]			
	Normal	*Retrognathic*	*Prognathic*	**All**
L_{diag}	47.37	84.21	81.82	76.24
$L_{diag} + L_{reco}$	52.63	92.11	79.55	79.21
$L_{diag} + L_{cons}$	42.11	92.11	84.09	79.21
$L_{diag} + L_{reco} + L_{cons}$	**57.89**	**92.11**	**86.36**	**83.17**

Acknowledgment

This work was partially supported by NIH under award **R01DE021863**.

References

1. Downs, "*Variations in facial relationships: their significance in treatment and prognosis*," 1948.
2. Anderson et al., "*Development of cephalometric norms using a unified facial and dental approach*," 2006.
3. Xu et al., "*Machine learning effectively diagnoses mandibular deformity using three-dimensional landmarks*," 2024.
4. Lugaresi et al., "*Mediapipe: A framework for building perception pipelines*," 2019.
5. Vaswani et al., "*Attention is all you need*," 2017.
6. Dai et al., "*Statistical modeling of craniofacial shape and texture*," 2019.
7. Kipf and Welling, "*Semi-supervised classification with graph convolutional networks*," 2016.
8. Veličkovićet al., "*Graph attention networks*," 2017.
9. Wu et al., "*Simplifying graph convolutional networks*," 2019.
10. Shi et al., "*Masked label prediction: Unified message passing model for semi-supervised classification*," 2020.