DiRecT: Diagnosis and Reconstruction Transformer for Mandibular Deformity Assessment

Xuanang Xu[1 [0000−0002−6045−8457]], 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 (✉) [0000−0002−9779−2141]]

[1] Department of Biomedical Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
yanp2@rpi.edu
[2] Department of Oral and Maxillofacial Surgery, Houston Methodist Research Institute, Houston, TX, 77030, USA
[3] Department of Neurosurgey, Baylor College of Medicine, Houston, TX 77030, USA

Abstract. In the realm of orthognathic surgical planning, the precision of mandibular deformity diagnosis is paramount to ensure favorable treatment outcomes. Traditional methods, reliant on the meticulous identification of bony landmarks via radiographic imaging techniques such as cone beam computed tomography (CBCT), are both resource-intensive and costly. In this paper, we present a novel way to diagnose mandibular deformities in which we harness facial landmarks detectable by off-the-shelf generic models, thus eliminating the necessity for bony landmark identification. We propose the Diagnosis-Reconstruction Transformer (DiRecT), an advanced network that exploits the automatically detected 3D facial landmarks to assess mandibular deformities. DiRecT's training is augmented with an auxiliary task of landmark reconstruction and is further enhanced by a teacher-student semi-supervised learning framework, enabling effective utilization of both labeled and unlabeled data to learn discriminative representations. Our study encompassed a comprehensive set of experiments utilizing an in-house clinical dataset of 101 subjects, alongside a public non-medical dataset of 1,519 subjects. The experimental results illustrate that our method markedly streamlines the mandibular deformity diagnostic workflow and exhibits promising diagnostic performance when compared with the baseline methods, which demonstrates DiRecT's potential as an alternative to conventional diagnostic protocols in the field of orthognathic surgery. Source code is publicly available at https://github.com/RPIDIAL/DiRecT.

Keywords: Mandibular deformity diagnosis· Orthognathic surgical planning · Transformer · Semi-supervised learning.

1 Introduction

Orthognathic surgery, a critical intervention aimed at correcting jaw deformities, relies heavily on accurate preoperative diagnosis to ensure optimal treatment planning and patient outcomes. Mandibular deformities, characterized by deviations in the lower jaw's position relative to the cranial base, can significantly impact an individual's functional and aesthetic aspects. However, the accuracy of diagnosing these deformities often depends on the clinician's experience, introducing subjectivity and variability in the assessment process. The traditional approach to diagnosing mandibular deformities involves cephalometry (or cephalometric measurement analysis), which was introduced by Downs in 1948 for clinical use [5]. This analytical method, integral to orthodontic diagnosis and treatment planning, utilizes specific bony anatomical landmarks to calculate various measurements, including distances, angles, and ratios, aiming to assess facial symmetry, growth patterns, and skeletal relationships. Although the cephalometry approach is widely adopted in clinical practice and have shown acceptable validity [1,21,11,19,2,16], it faces criticism for oversimplifying the complex anatomical structures of the face, thereby failing to capture the intricate three-dimensional relationships essential for a comprehensive mandibular deformity assessment. The recent advancements in computer vision and deep learning have opened new avenues for integrating artificial intelligence (AI) into orthognathic surgical planning [6,9,15,8,4,7,14,17], including the diagnostic process for mandibular deformities. These models, capable of analyzing vast datasets and identifying complex patterns, propose a revolutionary paradigm in craniofacial morphology analysis. Initial attempts to leverage deep learning for this purpose, such as the use of a 6-layer multilayer perceptron (MLP) [27], have demonstrated the potential for significant improvements in diagnostic accuracy and efficiency. However, these methods often necessitate accurate segmentation of bony structures and precise identification of dense bony landmarks in the volumetric Cone Beam Computed Tomography (CBCT) images, which is often labor-intensive, time-consuming, and radioactive risky. Moreover, the efficacy of deep learning approaches is contingent on the availability of substantial labeled data for training, a requirement challenging to meet within the medical domain due to the inherent scarcity of annotations and concerns of data privacy.

In light of these challenges, we introduce an innovative approach, termed Diagnosis-Reconstruction Transformer (DiRecT), for diagnosing mandibular deformities that does not rely on the bony anatomical landmarks and can efficiently leverage the large-size unlabeled non-medical data along with a small portion of labeled medical data for training through a semi-supervised learning manner. The major contribution of our DiRecT approach can be summarized in twofolds: 1) In stead of using the anatomical bony landmarks as the hint to assess the mandibular deformity status, we propose to leverage the facial soft tissue landmarks that can be easily detected by the pre-trained off-the-shelf models. By redirecting our approach's input from the bony anatomical landmarks to the facial soft tissue landmarks, we explored a new way that significantly simplifies the diagnostic process of mandibular deformities. 2) We developed an innova-

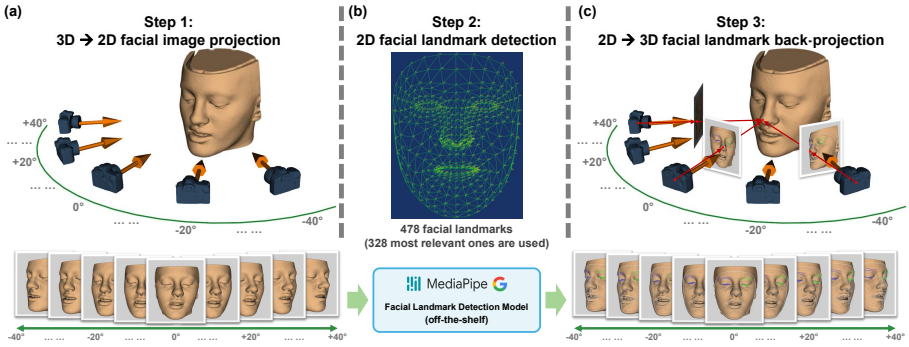

Fig. 1. 3D facial landmark extraction through 2D facial landmark detection model.

tive DiRecT network to address the task of mandibular deformity diagnosis using automatically detected facial landmarks. Our DiRecT network is enhance with an auxiliary task of facial landmark reconstruction and a teacher-student diagnoser framework, both of which are demonstrated effective in not only improving the learned representation quality but also fostering a semi-supervised learning paradigm, substantially reducing our method's dependence on labeled medical datasets while showing promising results compared with the conventional methods. We trained our model using a small set of labeled medical data (101 subjects) together with a large-size non-medical purposed public dataset (1,519 subjects). The experimental results demonstrate the data and label efficiency of our approach while exhibiting promising diagnostic performance.

2 Method

Sec. 2.1 outlines a 3D facial landmark extraction pipeline using off-the-shelf 2D landmark detection models, whose result serves as the input for our DiRecT network, which is elaborated in Sec. 2.2 and Sec. 2.3.

2.1 Facial landmark extraction using the MediaPipe framework

Traditional diagnostic approaches for mandibular deformities predominantly depend on the identification of bony landmarks within volumetric CBCT images—a process that is both laborious and time-consuming. In this study, we introduce an alternative methodology that leverages facial landmarks, rather than bony landmarks, to evaluate mandibular deformities. Facial landmark detection has been extensively studied within the field of general computer vision, yielding several high-performance, ready-to-use models, as well as access to substantial public datasets. Nonetheless, a significant challenge arises from the dimensionality discrepancy between the 2D natural images used in these models and the volumetric nature of CBCT images or the 3dMD head camera images, making a direct application of these models to our medical data infeasible. To bridge this gap, we

devised an innovative workflow that effectively extracts 3D facial landmarks from CBCT/3dMD head images utilizing an off-the-shelf facial landmark detection model trained on large-scale 2D natural images. The procedure of this workflow is depicted in Fig. 1 and encompasses three primary stages. Initially, we capture a sequence of images projecting the patient's 3D facial surface—obtained with ease from CBCT images via thresholding or from the 3dMD head camera—and document the corresponding camera parameters. Subsequently, these 2D images are processed through the chosen facial landmark detection model (in our case, Google's MediaPipe [18]) to identify facial landmarks on each photograph. Finally, we implement ray casting, utilizing the preserved camera parameters, to back-project the identified 2D landmarks onto the 3D facial surface. By averaging the back-projected points from all images, we compute the final set of 3D facial landmarks. The orignal MediaPipe model can identify 478 facial landmarks covering the whole face. Given that some landmarks are less relevant to mandibular deformty form, out of CBCT imaging scope, or failed in detection process, we finally select N=328 facial landmarks (see Fig. 1b) those are stable and shared across all the experimental subjects. The above innovative procedure effectively transforms the problem of 3D facial landmark detection into a 2D challenge, well within the capability range of existing off-the-shelf models. Such an approach circumvents the need to identify anatomical bony landmarks, thereby streamlining the diagnostic process for mandibular deformities significantly.

2.2 Mandibular deformity diagnosis using DiRecT network

Capitalizing on the transformative success of Transformer architectures in analyzing non-image data, we introduce the Diagnosis-Reconstruction Transformer (DiRecT) network, which is illustrated in Fig. 2. DiRecT is comprised of two synergistic components: the diagnoser network $f_D(\cdot|\theta_D)$ and reconstructor network $f_R(\cdot|\theta_R)$. The diagnoser $f_D(\cdot|\theta_D)$ ingests 3D facial landmark coordinates $\mathbf{P}$=$\{P_i \in \mathbb{R}^3\}_{i=1}^N$, which are initially embedded into 64D landmark tokens $\{X_{lmk_i} \in \mathbb{R}^{64}\}_{i=1}^N$. Subsequently, these tokens, concatenated with a learnable 64D class vector $X_{cls} \in \mathbb{R}^{64}$, are processed through a Transformer block, utilizing the self-attention mechanism to distill semantic information. The resultant class token $Z_{cls} \in \mathbb{R}^{64}$, extracted as the final embedding feature, undergoes transformation by a linear layer to yield a probability distribution $Y \in [0,1]^3$ across three classes—normal, retrognathic, and prognathic—indicative of the mandibular deformity diagnosis. Given the ground-truth diagnosis label $\hat{Y} \in \{0,1\}^3$, one of the training objective of the diagnoser network is to minimize the diagnosis loss $L_{diag} = CrossEntroy(Y, \hat{Y})$ by tuning the network parameters θ_D. In pursuit of enriching the semantic content encapsulated within the output class token Z_{cls} and ensuring its discriminative capacity, we instituted a reconstructor network $f_R(\cdot|\theta_R)$ in sequence with the diagnoser. Input to the reconstructor consists of the class token Z_{cls}, derived from the diagnoser, concatenated with N duplicates of a learnable mask token $Z_{msk} \in \mathbb{R}^{64}$ equalling the landmark tokens in quantity. The intended output of the reconstructor is the reconstructed 3D facial landmark coordinates $\mathbf{P'}$=$\{P'_i\}_{i=1}^N$. The reconstructor network's parameters θ_R are

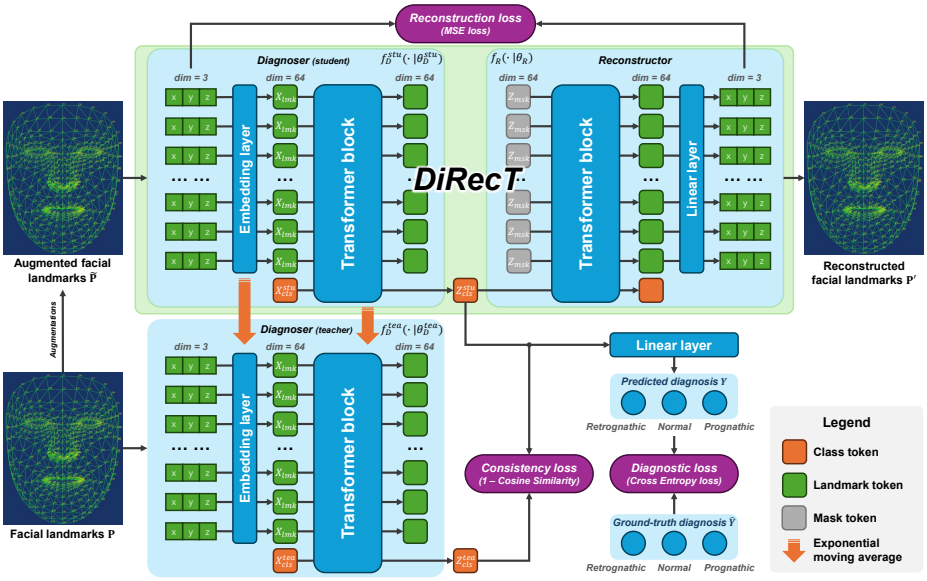

Fig. 2. Scheme of the proposed DiRecT network.

optimized by minimizing the reconstruction loss $L_{reco} = MSE(\mathbf{P}', \mathbf{P})$. Given that the reconstructor is informed solely by the class token Z_{cls} emanating from the diagnoser, exacting reconstructed facial landmarks—aligned with the original input—compels the class token Z_{cls} to assimilate comprehensive geometric information of the entire facial landmark set $\mathbf{P}=\{P_i\}_{i=1}^{N}$. This process, in turn, aids the diagnostic task predicated on the class token Z_{cls}'s informative nature.

2.3 Semi-supervised learning through teacher-student diagnoser

Given the scarcity of medical data juxtaposed with the abundance of publicly available facial model data, we advocate for the integration of semi-supervised learning to fortify the training regimen of our DiRecT. For this purpose, we duplicate the diagnoser network to function as a teacher $f_D^{tea}(\cdot|\theta_D^{tea})$, guiding the student $f_D^{stu}(\cdot|\theta_D^{stu})$ diagnoser network's training. The teacher diagnoser's parameters θ_D^{tea} are meticulously updated via the exponential moving average algorithm, based on the corresponding parameters of the student diagnoser θ_D^{stu}. Throughout the training phase, original facial landmarks $\mathbf{P}$ are subjected to random transformations, encompassing random mirroring and rescaling (without impacts on the diagnostic outcome), to create augmented counterparts $\tilde{\mathbf{P}}$. These original and augmented landmarks are then respectively fed to the teacher and student diagnosers. Despite discrepancies in their inputs, the output class tokens Z_{cls}^{tea} and Z_{cls}^{stu} from both diagnosers should exhibit consistency, reflecting identical semantic information relative to facial form. This process can be modeled by minimizing the consistency loss $L_{cons} = 1 - CosineSimilarity(Z_{cls}^{stu}, Z_{cls}^{tea})$.

This approach circumvents the need for ground-truth diagnostic labels, thereby permitting the application of this training schema to unlabeled data. Consequently, this expands our dataset breadth and markedly elevates the trained model's performance. The overall training objective of our DiRecT network is $L = L_{diag} + L_{reco} + \lambda L_{cons}$, where the weight of the consistency loss term λ linearly increases from 0 to 1 during the training phase to avoid misleading information from the teacher diagnoser at the beginning of the training.

2.4 Implementation details

Our approach is implemented using the Visualization Toolkit (VTK) [22] and PyTorch [20]. The model parameters are initialized using Xavier algorithm [10] and optimized by Adam optimizer [12] with a base learning rate of 0.001 for 800 epochs with a batch size of 64 (32/32 labeled/unlabeled samples). The input CBCT images are center-aligned with the nose tip of the nose (i.e., the facial landmark #4 detected by MediaPipe). The landmark coordinate values are normalized by dividing a rescale factor of 100.0 mm after the center-alignment. The Transformer block involved in our DiRecT network follows the original literature [24] with a feed-forward network dimension of 128. Softmax and Sigmoid activation functions are used as the output layers to normalize the output of the diagnoser and reconstructor networks, respectively. For better reproducibility, our source code is released at https://github.com/RPIDIAL/DiRecT.

3 Experiments

3.1 Dataset and metric

We conducted experiments using an in-house clinical dataset containing 101 subjects and the public Headspace dataset [3] of 3D human head images containing 1,519 subjects without mandibular deformity labels. Each subject from the in-house clinical dataset has a whole head CBCT image, from which we extracted the 3D facial soft tissue surface by applying a simple thresholding and the Marching Cube algorithm. A senior oral and maxillofacial surgeon with more than 30 years of clinical experience classified each subject's anteroposterior mandibular position as normal (19 subjects), retrognathic (38 subjects), or prognathic (44 subjects), which were used as the ground truth in the following experiments. Each subject from the Headspace dataset has a 3D head surface acquired by a 3dMD head camera and most of them are rendered with the facial appearance textures. From this dataset, we selected 917 subjects that fall within the age brackets considered optimal for orthognathic intervention (Females: 15-45 years; Males: 18-45 years) to create a representative sample reflective of the target population for orthognathic surgery. We applied our facial landmark extraction pipeline (introduced in Sec. 2.1) on both datasets, resulting in the automated identification of $N = 328$ key facial landmarks for each subject. We evenly divided our in-house clincal dataset into four parts (including 25/25/25/26 subjects, respectively) and performed 4-fold cross-validation (three/one folds for

Table 1. Comparison with different mandibular deformity diagnosis methods.

Methods	Input	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 [27]	Bony landmark	47.37	89.47	97.73	85.15
GCN [13]	Facial landmark	63.16	84.21	81.82	79.21
GAT [25]	Facial landmark	68.42	86.84	72.73	77.23
SGC [26]	Facial landmark	57.89	84.21	84.09	79.21
GTranformer [23]	Facial landmark	63.16	84.21	84.09	80.20
DiRecT (ours)	Facial landmark	57.89	92.11	86.36	83.17

training/testing, respectively) for evaluation. The Headspace dataset was served as unlabeled data for the semi-supervised training of our DiRecT network. Classification accuracy was used as the primary metric to quantitatively assess the model's diagnostic performance.

3.2 Comparison with other methods

We first compared the proposed method with the baseline methods for mandibular deformity diagnosis. The competing methods encompass conventional cephalometric measurements, such as the Sella-Nasion-B (SNB) angle, facial angle, and mandibular unit length (MdUL), and advanced computational models, including an MLP model employing 50 bony landmarks [27], as well as contemporary neural network architectures adapted to our input modality, such as Graph Convolutional Networks (GCNs) [13,25,26] and Transformers [23]. The comparison results, as summarized in Table 1, illustrate that DiRecT outperforms the traditional cephalometric methods overall, showcasing its capability to surmount the inherent limitations of the traditional measurements. Notably, DiRecT demonstrates superior accuracy over other neural network-based methods that utilize facial landmarks, reinforcing the benefits of the landmark reconstruction task and teacher-student diagnoser framework within our approach. While the MLP-based diagnosis method [27] demonstrates higher accuracy than our method, it is essential to underscore that DiRecT can achieve its results without relying on the bony landmarks annotated in CT/CBCT images, thus mitigating exposure risks and reducing procedural costs. These findings validate the proposed DiRecT network as a robust and less invasive alternative for mandibular deformity diagnosis, offering significant workflow enhancements without compromising diagnostic accuracy.

Table 2. Ablation study on the proposed method.

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} (ours)	57.89	92.11	86.36	83.17

3.3 Ablation study

We conduct ablation studies on the proposed method to justify the effectiveness of the key components within our proposed DiRecT network. Specifically, four ablation models are included in this study: 1) the base diagnoser network trained with only the diagnosis loss (L_{diag}), serving as the foundation of our architecture, 2) an enhanced diagnoser network appended with the reconstructor network and trained by minimizing both diagnosis and reconstruction losses ($L_{diag} + L_{reco}$), 3) an expanded diagnoser model incorporating the teacher-student consistency loss ($L_{diag}+L_{cons}$), which adopts the semi-supervised learning framework, and 4) the full DiRecT network amalgamating all three loss components ($L_{diag} + L_{reco} + L_{cons}$), representing our complete model as introduced in this study. The results of the ablation study are shown in Table 2. Using the diagnoser network alone exhibits the lowest accuracy than other ablation models. By introducing the reconstructor and teacher-student consistency regularization into our framework, the diagnosis accuracy increased gradually, finally achieving our proposed DiRecT network that exhibited the highest accuracy. This result demonstrates the effectiveness of our two key designs, i.e., 1) the reconstruction regularization enhancing the semantic information captured by the class token from the diagnoser network and 2) the teacher-student diagnoser framework allowing the semi-supervised learning on large-size unlabeled data.

4 Conclusion

This study explored a novel way for the diagnosis of mandibular deformities, which is pivotal to the success of orthognathic surgical planning. The proposed approach pivots from the conventional reliance on bony landmarks (which are obtained through time-intensive annotation and radiation-based imaging modalities like CT or CBCT) to the utilization of facial landmarks detectable by off-the-shelf generic models. This strategic shift not only facilitates a reduction in resource consumption but also aligns with contemporary efforts to streamline medical workflows. Our main contribution, the Diagnosis-Reconstruction Transformer (DiRecT), benefits from the automated detection of 3D facial landmarks, effectively employing them as hint for mandibular deformity diagnosis. The auxiliary task of landmark reconstruction alongside the innovative teacher-student

diagnoser framework enhances DiRecT's performance, allowing it to leverage the knowledge from unlabeled data while learning more discriminative representations. The comprehensive experimental evaluation using both an in-house clinical dataset and a publicly available non-medical dataset has evidenced the efficacy of DiRecT. With performance that is competitive with established methods, DiRecT presents a compelling alternative that streamlines the diagnostic process, mitigates the need for specialized radiographic imaging, and opens the door to more accessible orthognathic surgery planning. While the results are promising, future work may focus on extending DiRecT's application to a wider range of craniofacial anomalies, exploring the integration of additional non-invasive diagnostic modalities, and validating the approach in multi-center clinical trials to establish its efficacy in diverse populations.

Acknowledgments. This work was partially supported by NIH under awards R01 DE021863.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anderson, G., Fields, H.W., Beck, M., Chacon, G., Vig, K.W.: Development of cephalometric norms using a unified facial and dental approach. The Angle Orthodontist **76**(4), 612–618 (2006)
2. Baik, C.Y., Ververidou, M.: A new approach of assessing sagittal discrepancies: the beta angle. American journal of orthodontics and dentofacial orthopedics **126**(1), 100–105 (2004)
3. Dai, H., Pears, N., Smith, W., Duncan, C.: Statistical modeling of craniofacial shape and texture. International Journal of Computer Vision **128**(2), 547–571 (Nov 2019). https://doi.org/10.1007/s11263-019-01260-7, https://doi.org/10.1007/s11263-019-01260-7
4. Deng, H., Liu, Q., Chen, A., Kuang, T., Yuan, P., Gateno, J., Kim, D., Barber, J., Xiong, K., Yu, P., Gu, K., Xu, X., Yan, P., Shen, D., Xia, J.: Clinical feasibility of deep learning-based automatic head cbct image segmentation and landmark detection in computer-aided surgical simulation for orthognathic surgery. International Journal of Oral and Maxillofacial Surgery **52**(7), 793–800 (2023)
5. Downs, W.B.: Variations in facial relationships: their significance in treatment and prognosis. American journal of orthodontics **34**(10), 812–840 (1948)
6. Fang, X., Deng, H.H., Kuang, T., Xu, X., Lee, J., Gateno, J., Yan, P.: Patient-specific reference model estimation for orthognathic surgical planning. International Journal of Computer Assisted Radiology and Surgery pp. 1–9 (2024)
7. Fang, X., Kim, D., Xu, X., Kuang, T., Deng, H.H., Barber, J.C., Lampen, N., Gateno, J., Liebschner, M.A., Xia, J.J., et al.: Deep learning-based facial appearance simulation driven by surgically planned craniomaxillofacial bony movement. In: International conference on medical image computing and computer-assisted intervention. pp. 565–574. Springer (2022)

8. Fang, X., Kim, D., Xu, X., Kuang, T., Lampen, N., Lee, J., Deng, H.H., Gateno, J., Liebschner, M.A., Xia, J.J., Yan, P.: Soft-tissue driven craniomaxillofacial surgical planning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 186–195. Springer (2023)
9. Fang, X., Kim, D., Xu, X., Kuang, T., Lampen, N., Lee, J., Deng, H.H., Liebschner, M.A., Xia, J.J., Gateno, J., et al.: Correspondence attention for facial appearance simulation. Medical Image Analysis **93**, 103094 (2024)
10. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
11. Gupta, P., Singh, N., Tripathi, T., Gopal, R., Rai, P.: Tau angle: A new approach for assessment of true sagittal maxillomandibular relationship. International Journal of Clinical Pediatric Dentistry **13**(5), 497 (2020)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
14. Lampen, N., Kim, D., Fang, X., Xu, X., Kuang, T., Deng, H.H., Barber, J.C., Gateno, J., Xia, J., Yan, P.: Deep learning for biomechanical modeling of facial tissue deformation in orthognathic surgical planning. International journal of computer assisted radiology and surgery **17**(5), 945–952 (2022)
15. Lampen, N., Kim, D., Xu, X., Fang, X., Lee, J., Kuang, T., Deng, H.H., Liebschner, M.A., Xia, J.J., Gateno, J., Yan, P.: Spatiotemporal incremental mechanics modeling of facial tissue change. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 566–575. Springer (2023)
16. Lee, J., Kim, D., Xu, X., Kuang, T., Gateno, J., Yan, P.: Predicting optimal patient-specific postoperative facial landmarks for patients with craniomaxillofacial deformities. International Journal of Oral and Maxillofacial Surgery (2024)
17. Lee, J., Xu, X., Kim, D., Deng, H.H., Kuang, T., Lampen, N., Fang, X., Gateno, J., Yan, P.: Large language models diagnose facial deformity. In: CARS 2024—Computer Assisted Radiology and Surgery Proceedings of the 38th International Congress and Exhibition, Barcelona, Spain (2024)
18. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong, M.G., Lee, J., et al.: Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172 (2019)
19. Neela, P.K., Mascarenhas, R., Husain, A.: A new sagittal dysplasia indicator: the yen angle. World journal of orthodontics **10**(2) (2009)
20. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
21. Poosit, M., Basafa, M., Ahrari, F., Movahedian, A.R.: Sensitivity and specificity of snb and facial angles in diagnosis of mandibular anteroposterior position in class ii patients. Iranian Journal of Orthodontics **2**(1), 54–60 (2007)
22. Schroeder, W., Martin, K., Lorensen, B.: The Visualization Toolkit (4th ed.). Kitware (2006)
23. Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y.: Masked label prediction: Unified message passing model for semi-supervised classification. arXiv preprint arXiv:2009.03509 (2020)

24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
25. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
26. Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., Weinberger, K.: Simplifying graph convolutional networks. In: International conference on machine learning. pp. 6861–6871. PMLR (2019)
27. Xu, X., Deng, H.H., Kuang, T., Kim, D., Yan, P., Gateno, J.: Machine learning effectively diagnoses mandibular deformity using three-dimensional landmarks. Journal of Oral and Maxillofacial Surgery **82**(2), 181–190 (2024)