Bridging the Patient Distribution Gap for Robust 3D Tooth Segmentation

Hao Chen[*1] Haoc.19@intl.zju.edu.cn
[1] *Zhejiang University-University of Illinois at Urbana-Champaign Institute, Zhejiang University*

Jianfei Yang[*2] yang0478@e.ntu.edu.sg
[2] *School of Electrical and Electronic Engineering, Nanyang Technological University*

Yang Feng[3] fengyang@angelalign.com
[3] *Angelalign Inc.*

Jin Hao[4] jin_hao@g.harvard.edu
[4] *Department of Stem Cell and Regenerative Biology, Harvard University*

Huimin Xiong [1] huimin.21@intl.zju.edu.cn
Zuozhu Liu [1] zuozhuliu@intl.zju.edu.cn

Editors: Under Review for MIDL 2022

Abstract

Empowered by deep learning, 3D semantic segmentation algorithms have been applied to computer-aided dental systems in recent years. Existing models yield satisfactory performance for 3D point cloud data. However, there exists a common assumption that the data distribution of training and testing scenarios is the same, which limits the model capacity to deal with cross-domain situations. In real-world clinical scenarios, the patients for evaluation could be quite different from those for training in terms of their teeth symptoms, such as teeth defects and eruption. To deal with this problem, we borrow the idea of Domain Adaptation (DA) and propose a Domain-Invariant Tooth Segmentation (DITS) framework that bridges the clinically symptomatic domain gap. DITS leverages a dynamic graph convolutional neural network for 3D semantic segmentation backbone, while maximizing the probabilities of well-classified samples during evaluation by maximum square loss, thus adapting the 3D segmentation model to a realistic domain with different teeth symptoms. In the experiment, the real-world datasets are collected including 4272 3D IOS scans which are annotated with tooth-ID and three common tooth symptoms by experts. Extensive experiments have shown that DITS leads to a significant improvement for the large-scale cross-domain 3D tooth segmentation.

Keywords: Tooth Segmentation, Domain Adaptation, 3D Point Cloud

1. Introduction

In recent years, deep learning has enabled more and more applications on tooth data analysis due to the prominent performance of deep models and increasing demands for dental health (Cui et al., 2021). Computer-aided Dental Systems (CDS) can not only save the time of the dentist doing laborious work, but also enhance the treatment quality by providing precise clinical information (Eid et al., 2019). In automatic dental systems, 3D tooth semantic

[*] These authors contributed equally.

segmentation is one of the most important problems for its wide applications in dental diagnosis and orthodontics.

The semantic segmentation of 3D intra-oral scanner (IOS) mesh data is to identify different categories of teeth and gingiva under pixel level (Hao et al., 2021). Different appearances of teeth make great difficulties in automatic tooth segmentation. For example, it is very common to see that the dental arch forms are not consistent across different patients. Some complex cases like tooth shape alternation and tooth size abnormalities are harder to deal with. Besides, 3D IOS mesh data leads to tremendous computational complexity (Rodrigues et al., 2018), so we usually convert them into point cloud data for segmentation algorithm development.

Though there have existed some deep models such as Dynamic Graph CNN (DGCNN) (Wang et al., 2019) that yield excellent performance, in practice the unseen tooth with various shapes and diseases still hinder the availability of deep segmentation models. The reason is that deep models follow the same fundamental assumption that the training and testing data are individually independently distributed (i.d.d.) (Pan and Yang, 2009). This assumption is not always true in realistic applications, especially for 3D teeth data. It's not surprising that patients have their unique conditions (i.e. tooth shape and disease such as tooth eruption and tooth defect), but the training data may not have sufficient samples of these diverse tooth data, leading to distribution difference. Such difference results in the decreasing performance of deep tooth segmentation models, as proved in our experiments. In statistical learning, it is termed as domain shift or data set bias between a source domain (i.e. training data) and a target domain (i.e. testing data) (Ben-David et al., 2010). This problem also occurs in many computer vision problems (Yang et al., 2021). Domain adaptation methods set out to deal with such problems (Long et al., 2015), which utilizes the labeled source domain and the unlabeled target domain for model adaptation. In our scenario, the source domain is the well-annotated training data collected by the lab developers, while the target domain is unlabeled and hence available when the CDS is manipulated by dentists in the real world.

In this paper, we develop a robust Domain-Invariant Tooth Segmentation (DITS) method to bridge the gap for 3D tooth segmentation. The DITS can adapt to the new scenarios where various teeth with different symptoms appear in an unsupervised manner. More specifically, it leverages a novel DGCNN segmentation model as backbone, and borrows the idea of semi-supervised learning (Grandvalet et al., 2005) to achieve robust segmentation across patients with various symptoms and appearances. A maximum square loss is employed to learn new tooth appearances by leveraging gradients of the confident samples in the target domain, which adapts the model to new patients. In this fashion, DITS continues to increase its performance by domain adaptation with the unlabeled data increasing during realistic dental diagnosis. In our experiments, it is demonstrated that our proposed DITS outperforms the state-of-the-art DC-Net segmentation model. The contribution of this paper can be concluded as:

- We demonstrate that the distribution divergence exists in real-world tooth data with respect to different tooth symptoms and appearances, and we formulate it to be a domain adaptation problem.

- We propose a novel DITS framework based on DGCNN segmentation backbone and a novel semi-supervised loss to bridge the gap of diverse patient distributions. To the best of our knowledge, DITS is the first system that tackles the domain shift of different tooth symptoms and empowers robust 3D tooth segmentation.
- To verify the effectiveness, we collect the first large-scale 3D tooth dataset with expert annotations of three tooth symptoms. Extensive experiments have been conducted on three domains where DITS achieves state-of-the-art performance.

2. Related Work

For tooth segmentation, there are plenty of existing works that try to do semantic segmentation on 3D IOS tooth data. Some basic models like U-net (Ronneberger et al., 2015) for 2D using geometric based methods are the first step, and then extra hand-crafted features such as curvatures of the 3D mesh tooth data are used to improve the performance of the model (Tzeng et al., 2014; Fan et al., 2014; Kumar et al., 2011; Wang and Li, 2016). However, these models are restricted to some special teeth due to the geometric based method is not suitable for tooth variation. Thanks to deep learning techniques, some advanced methods (Lian et al., 2019; Xu et al., 2018) can be applied on a wide range of data. Notwithstanding, these methods only show satisfactory results with large amounts of training resources and they are limited in scales and scopes. Recently, the DC-Net (Hao et al., 2021) based on DGCNN (Wang et al., 2019) show a great performance on this problem. However, all these models assume that train data and test data have the same distribution which is not always the case for clinics. Every patient has a unique tooth shape and tooth disease like tooth defect, which will cause the divergence of tooth mesh data. Even for patients who have the same symptom, the degree of dental defects will make the performance of previous models decrease greatly.

To tackle this problem, unsupervised domain adaptation (UDA) aims transfer the knowledge from a label-rich source domain to an unlabeled target domain (Pan and Yang, 2009). This can be achieved by minimizing the feature distributions by some measures such as maximum mean discrepancy (Long et al., 2015; Tzeng et al., 2014). Recently, adversarial based methods (Ganin et al., 2016; Yang et al., 2020) learn domain-invariant features by a tailored minimax game (Zou et al., 2019). They show great performance but require enormous computations. Another stream of UDA encourages the model to get familiar with the target domain by learning confident samples (Chen et al., 2019; Grandvalet et al., 2005), which inspires DITS that learns confident target samples during dental diagnosis for model adaptation.

3. Method

3.1. Problem Formulation

For tooth semantic segmentation tasks, the objective is to learn a model Φ that segments each tooth at the pixel level. In our problem, the training and testing data have different data distributions due to symptoms and shapes of tooth. In domain adaptation, the semantic segmentation model can access to a labeled source domain $D_S = \left\{(x_S^i, y_S^i)\right\}_{i=1}^{N_S}$ and an

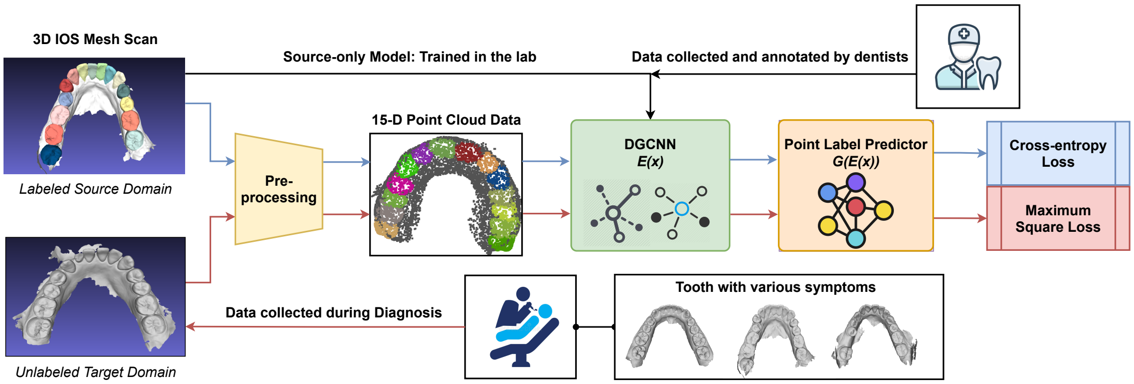


Figure 1: Our model consists of a feature extractor and a label predictor. The mesh data is firstly pre-processed to 15-D Point Cloud data, and then the DITS achieves robust semantic segmentation by supervised learning in the source domain and unsupervised domain adaptation in the target domain.

unlabeled target domain $D_T = \{x_T^i\}_{i=1}^{N_T}$, where N_S and N_T denote the number of samples in two domains. Each input sample $x \in \mathcal{R}^{15\times N_p}$ includes the point cloud data (i.e. xyz coordinates), the normal vector of the mesh and the 9-dim shape descriptor of the mesh, detailed in the data setup. The corresponding label y is the teeth ID ranging from 0 to 32 (i.e. totally 33 labels). The source domain can be any perfectly-annotated data in the lab, which can be used to train a source-only model Φ_S. During realistic clinics, the users can conveniently collect massive unlabeled data, i.e. the target domain, which is leveraged for domain adaptation to obtain the final model Φ.

3.2. Domain-Invariant Tooth Segmentation

The proposed DITS tackles the domain shift caused by different tooth appearances for tooth semantic segmentation, which takes labeled instances from source domain and unlabeled instances from target domain as inputs. As shown in Figure 1, DITS consists of a feature extractor $E(\cdot)$ that learns robust representations for 3D point cloud data, and a label predictor $G(\cdot)$ that classifies each pixel to a specific teeth ID. In DITS, the state-of-the-art DGCNN (Wang et al., 2019) is utilized as the feature extractor. The label predictor is composed of fully-connected layers.

Having the labeled source domain, DITS firstly learns the semantic segmentation by fitting into the source domain, which can be achieved by minimizing the well-known cross-entropy loss:

$$\mathcal{L}_{CE}(x_S, y_S) = -\frac{1}{N}\sum_{m=1}^{N}\sum_{c=1}^{C} y_S^{m,c} * log(p^{m,c}), \tag{1}$$

where m represents a pixel point, $p^{m,c}$ is the prediction probability of point m for class c that is calculated by $G(E(x_S))$. By cross-entropy loss, a single-domain segmentation model is obtained but its performance can dramatically decrease when confronting data of other distributions.

To tackle the 3D domain adaptation problem, many methods in computer vision resort to self-supervised learning (Achituve et al., 2021), adversarial training (Ganin et al., 2016) and statistical domain alignment (Long et al., 2015). Despite their significant improvement of the performance, these methods either require a large number of samples from the target domain, or demand powerful computational resources, which is not applicable in practice. In DITS, we find that such a domain gap can be mitigated well by minimizing the conditional entropy of the target domain prediction. Entropy minimization facilitates the semi-supervised learning by enforcing the classifier to follow the cluster assumption and go through low-density regions (Grandvalet et al., 2005). This method also ameliorates the cross-domain semantic segmentation method. However, in tooth semantic segmentation, the points and their predictions are usually imbalanced since the teeth with bigger volume have more points than the small ones. Furthermore, the teeth with more points usually have a better chance to be correctly classified. If we directly apply entropy minimization, the gradients from these bigger teeth will dominate the unsupervised training.

To remedy such a problem, the Maximum Square Loss (MSL) (Chen et al., 2019) is leveraged in DITS, which only processes the target domain samples in the user side and is computational-efficient. The MSL is calculated by

$$\mathcal{L}_{MSL}(x_T)) = -\frac{1}{2N}\sum_{n=1}^{N}\sum_{c=1}^{C}(p^{n,c})^2, \tag{2}$$

where $L_{MSL}(x_T)$ denotes the Maximum Square loss of target samples. Note that the MSL only requires the target domain samples without labels, and it encourages the semantic segmentation model to learn confident samples in the target domain while leveraging the effective gradients from the imbalanced categories of teeth data, thus adapting the model to the new tooth data in an unsupervised manner.

Regarding the MSL as an unsupervised loss in the training procedure, the optimization of DITS is summarized as

$$\min_{E,G}\mathcal{L}_{DITS}(x_S, x_T, y_S) = \mathcal{L}_{ce}(x_S, y_S) + \lambda\mathcal{L}_{MSL}(x_T) \tag{3}$$

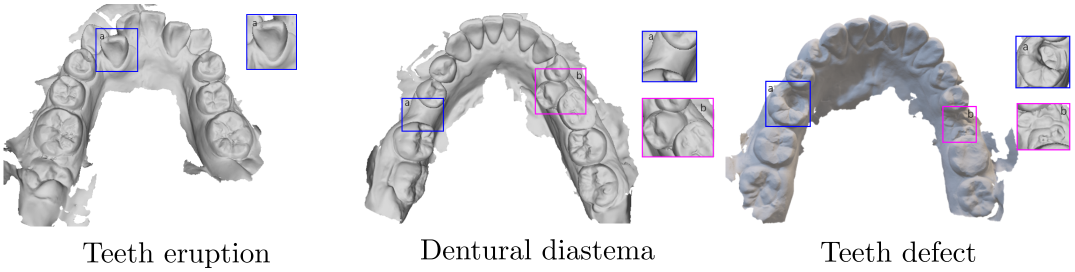
where λ balances the two loss terms and prevents the unsupervised loss dominates the training. When $\lambda = 0$, the training procedure is a standard semantic segmentation method.

In the real-world applications, tremendous tooth data annotated in the lab is used as the source domain. Most of these teeth are in good conditions. When deploying the model at the dentist side, the user can easily collect many unlabeled teeth during dental diagnosis. Then DITS learns to adapt by unsupervised domain adaptation, increasing the semantic segmentation performance in practice.

4. Experiment

4.1. Setup

Dataset To facilitate automatic dental diagnosis research, we collect a large-scale dataset that consists of 4272 IOS mesh data labeled by dentists experts with equal numbers of the upper jaw and the lower jaw. Each tooth has been labeled according to FDI notation by



Teeth eruption

Dentural diastema

Teeth defect

Figure 2: Visualization of Tooth Symptoms

experts. For each face f_i, there are 32 teeth ID and the tooth symptoms are recorded by their dentists in the medical record. Moreover, each sample is annotated by multi-labels: whether one has teeth defects, dentural diastema and tooth eruption, as shown in Figure 2. These form three transfer tasks: from the normal case (i.e. domain label 'no') to the diseased case (i.e. domain label 'yes'). The detailed number of samples for each domain and train-test split is summarized in the appendix. For comparison, we also create a "i.i.d. case" by splitting the source and target domain by random shuffling, which creates the identically distributed training and testing set.

Criterion For evaluation, we employ the mean Intersection-over-Union (mIoU) of all the classes as our evaluation criterion. For a specific class c, the sets of prediction and ground truth is denoted as P_l and T_l. Then the IoU can be calculcated by

$$\text{IoU} = \frac{P_l \cap T_l}{P_l \cup T_l} \tag{4}$$

where c consists of 32 teeth ID. mIoU actually represents the overall segmentation accuracy of pixel level. Another evaluation criterion is the pixel accuracy (ACC_p) that calculates the percentage of correctly predicted pixels on the whole point cloud pixels.

Implementation Details We first transform each 3D IOS mesh scan, which contains 100,000-350,000 triangular faces, to a point cloud with 10,000 points through a uniform random sampling strategy over face centers. For each point, we compute three predefined features, including the location of face center $h_c = [x_c, y_c, z_c] \in \mathbb{R}^3$ with x_c, y_c, z_c as corresponding 3D coordinates, the face normal vector $h_n \in \mathbb{R}^3$, and a face shape descriptor $h_s \in \mathbb{R}^9$. In particular, for each face with three vertices $v_i = [x_i, y_i, z_i]_{i=1}^3$ and a face center h_c, the face shape is defined as $h_s = \text{Concat}([v_i - h_c]_{i=1}^3)$. $Concat()$ is the concatenate operation for vectors, leading to a 9-dimensional shape feature for each face. In consequence, the final output after data preprocessing for each IOS mesh scan is a point cloud with 10,000 points, each associated with a 15-dimensional feature vector $h = \text{Concat}(h_c, h_n, h_s) \in \mathbb{R}^{15}$. This preprocessing procedure can be finished with a modern computer on-the-fly.

In DITS, the baseline model is the state-of-the-art DC-Net (Hao et al., 2021) that only leverages the source domain for training, denoted as "source-only" model. For DITS, we always use $\lambda = 1$ for all experiments, which are obtained empirically. Better results may be produced by tuning λ. In the training procedure, we use AdamW as our optimizer with learning rate = 0.001 and the model is trained with 24 epochs. A server with 4 NVIDIA RTX3090 is used for experiments. The batch size is set to 2 per GPU. The result of the last

criterion(%) case model	i.i.d. case		teeth eruption		dentural diastema		teeth defects	
	mIoU	ACC$_p$	mIoU	ACC$_p$	mIoU	ACC$_p$	mIoU	ACC$_p$
DC-Net	87.74	94.81	85.27	93.69	84.48	94.73	85.27	94.67
DITS	/		**90.30**	**96.53**	**86.67**	**95.44**	**89.30**	**96.56**

Table 1: The mIoU and pixel accuracy (ACC$_p$) on three transfer tasks.

epoch is reported as we have no idea when to stop in the real applications. All experiments are run three times and the mean results are reported.

4.2. Overall Results

We firstly evaluate the DC-Net on the i.i.d. case. As shown in Table 1, the backbone model of DITS (i.e. DC-Net) can achieve a 87.74% mIoU and 94.81% pixel accuracy. Compared to the i.i.d. case, it is found that the cross-domain evaluations on three tooth symptoms obtain decreasing performance by 2-3%, which is caused by the domain shift. Then the DITS can adapt the model to the target domain by only learning around 200 samples for three tooth symptoms, and the performances of 3D semantic segmentation are boosted significantly. It is seen that the mIoU increases by 5.03%, 2.19% and 4.03% for teeth eruption, dentural diastema and teeth defects, respectively. This demonstrates that the DITS is able to adapt the model to a target domain by only a few samples.

Domain adaptation to teeth eruption data								
λ	0	0.5	1	2	4	6	8	10
mIoU (%)	85.27	89.34	90.30	90.56	89.84	89.32	**90.70**	90.30
ACC$_p$(%)	93.69	95.96	**96.53**	96.26	96.06	96.11	96.44	96.13

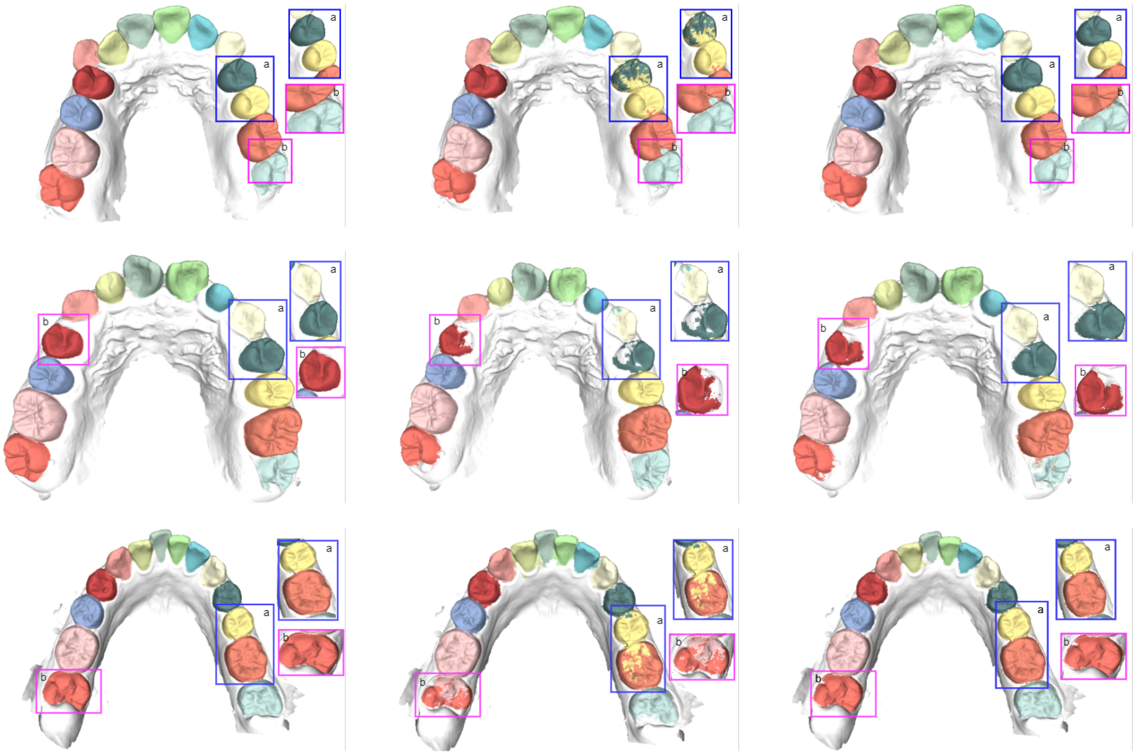
Table 2: hyper parameter sensitivity analysis

4.3. Hyper-parameter Sensitivity

In DITS, the hyper-parameter λ is important as it balances the supervised and unsupervised loss. We study the sensitivity of λ and find that it is not quite sensitive. We choose the transfer task of teeth eruption and evaluate the DITS while λ varies from 0 to 10. Note that the case of $\lambda = 0$ is the source-only model. As reported in Table 2, it is found that all settings of λ lead to some improvements. Thus the DITS is robust to the hyper-parameter even though it is not manually tuned. The best mIoU is 90.7% when $\lambda = 8$ and the best pixel accuracy is 96.53% when $\lambda = 1.0$. Empirically, we can set $\lambda = 1.0$ that is used for all experiments in Table 3, and better results can be achieved by tuning λ.

4.4. Visualization

The superiority of our DITS method is further demonstrated with 3 case visualizations in Figure 3. The first case illustrates that the DC-Net wrongly identifies the erupted teeth, while DITS can tackle this much better. The second case shows that the DC-Net fails to completely recognize the first premolar of the jaw's left side due to teeth defects. Moreover,



(a) Ground truth (b) DC-Net (c) DITS

Figure 3: Experimental results of DC-Net and our model for 3D Tooth Segmentation. Each row contains three segmentation results of one sample with a specific tooth symptom or disease, where the first row is teeth eruption, the second row is teeth defects and the last row is dentural diastema. (a) The ground truth. (b) The result of train from DC-Net. (c) The result of DITS.

it mislabels the canine tooth and the first premolar on the other side, which is largely improved by our DITS. In the third case, the DC-Net commits mistakes on the dentural diastemai.e. between the molars as well as between the second premolar and the first molar. In stark contrast, our DITS is superior to the baseline on these complicated cases, which has a promising future on the clinical application.

5. Conclusion

In our paper, we bridge the gap between the existing 3D semantic tooth segmentation model and the realistic dental scenarios where many unseen tooth with diverse shapes and diseases appear. Such gap is formulated as a domain adaptation problem, which inspires us to develop the DITS model. Our model learns robust tooth representations by leveraging gradients of confident samples in the target domain. It achieves state-of-the-art performance on realistic 3D tooth data. We believe that its clinical application has a promising future that releases humans from laborious work with an end-to-end structure to improve the efficiency of treatment.

References

Idan Achituve, Haggai Maron, and Gal Chechik. Self-supervised learning for domain adaptation on point clouds. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 123–133, 2021.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2090–2099, 2019.

Zhiming Cui, Changjian Li, Nenglun Chen, Guodong Wei, Runnan Chen, Yuanfeng Zhou, Dinggang Shen, and Wenping Wang. Tsegnet: an efficient and accurate tooth segmentation network on 3d dental model. *Medical Image Analysis*, 69:101949, 2021.

Rita Eid, Jelena Juloski, Hani Ounsi, Munir Silwaidi, Marco Ferrari, and Ziad Salameh. Fracture resistance and failure pattern of endodontically treated teeth restored with computer-aided design/computer-aided manufacturing post and cores: A pilot study. *Journal of Contemporary Dental Practice*, 20(1):56–63, 2019.

Ran Fan, Xiaogang Jin, and Charlie CL Wang. Multiregion segmentation based on compact shape prior. *IEEE Transactions on Automation Science and Engineering*, 12(3):1047–1058, 2014.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

Yves Grandvalet, Yoshua Bengio, et al. Semi-supervised learning by entropy minimization. *CAP*, 367:281–296, 2005.

J Hao, W Liao, YL Zhang, J Peng, Z Zhao, Z Chen, BW Zhou, Y Feng, B Fang, ZZ Liu, et al. Toward clinically applicable 3-dimensional tooth segmentation via deep learning. *Journal of dental research*, page 00220345211040459, 2021.

Yokesh Kumar, Ravi Janardan, Brent Larson, and Joe Moon. Improved segmentation of teeth in dental models. *Computer-Aided Design and Applications*, 8(2):211–224, 2011.

Chunfeng Lian, Li Wang, Tai-Hsien Wu, Mingxia Liu, Francisca Durán, Ching-Chang Ko, and Dinggang Shen. Meshsnet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3d dental surfaces. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 837–845. Springer, 2019.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

Rui SV Rodrigues, José FM Morgado, and Abel JP Gomes. Part-based mesh segmentation: a survey. In *Computer Graphics Forum*, volume 37, pages 235–274. Wiley Online Library, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014.

Hao Wang and Zhongyi Li. Tooth separation from dental model using segmentation field. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5616–5619. IEEE, 2016.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

Xiaojie Xu, Chang Liu, and Youyi Zheng. 3d tooth segmentation and labeling using deep convolutional neural networks. *IEEE transactions on visualization and computer graphics*, 25(7):2336–2348, 2018.

Jianfei Yang, Han Zou, Yuxun Zhou, Zhaoyang Zeng, and Lihua Xie. Mind the discriminability: Asymmetric adversarial domain adaptation. In *European Conference on Computer Vision*, pages 589–606. Springer, 2020.

Jianfei Yang, Han Zou, Yuxun Zhou, and Lihua Xie. Robust adversarial discriminative domain adaptation for real-world cross-domain visual recognition. *Neurocomputing*, 433: 28–36, 2021.

Han Zou, Yuxun Zhou, Jianfei Yang, Huihan Liu, Hari Prasanna Das, and Costas J Spanos. Consensus adversarial domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5997–6004, 2019.

Appendix A. Dataset Summary

Tooth Symptoms	Domain label	Total	Train	Test
Teeth eruption	no	2394	2394	-
	yes	968	242	726
Teeth defects	no	2300	2300	-
	yes	590	160	430
Dentural diastema	no	2145	2145	-
	yes	591	133	398

Table 3: Dataset Summarization