Automatic Detection of Tooth-Gingiva Trim Lines on Dental Surfaces

Geng Chen, Jie Qin, *Member, IEEE*, Boulbaba Ben Amor, *Senior Member, IEEE*, Weiming Zhou, Hang Dai, Tao Zhou, Heyuan Huang, and Ling Shao, *Fellow, IEEE*

***Abstract*— Detecting the tooth-gingiva trim line from a dental surface plays a critical role in dental treatment planning and aligner 3D printing. Existing methods treat this task as a segmentation problem, which is resolved with geometric deep learning based mesh segmentation techniques. However, these methods can only provide indirect results (i.e., segmented teeth) and suffer from unsatisfactory accuracy due to the incapability of making full use of high-resolution dental surfaces. To this end, we propose a two-stage geometric deep learning framework for automatically detecting tooth-gingiva trim lines from dental surfaces. Our framework consists of a trim line proposal network (TLP-Net) for predicting an initial trim line from the low-resolution dental surface as well as a trim line refinement network (TLR-Net) for refining the initial trim line with the information from the high-resolution dental surface. Specifically, our TLP-Net predicts the initial trim line by fusing the multi-scale features from a U-Net with a proposed residual multi-scale attention fusion module. Moreover, we propose feature bridge modules and a trim line loss to further improve the accuracy. The resulting trim line is then fed to our TLR-Net, which is a deep-based LDDMM model with the high-resolution dental surface as input. In addition, dense connections are incorporated into TLR-Net for improved performance. Our framework provides an automatic solution to trim line detection by making full use of raw high-resolution dental surfaces. Extensive experiments on a clinical dental surface dataset demonstrate that our TLP-Net and TLR-Net are superior trim line detection methods and outperform cutting-edge methods in both qualitative and quantitative evaluations.**

***Index Terms*— Dental surface, tooth-gingiva trim line, geometric deep learning, LDDMM, template fitting.**

Manuscript received 23 January 2023; revised 14 March 2023; accepted 19 March 2023. Date of publication 29 March 2023; date of current version 27 October 2023. This work was supported in part by the STI-2030 Major Projects under Grant 2022ZD0209000; in part by the National Natural Science Foundation of China under Grant 62201465, Grant 62276129, and Grant 62172228; in part by the Fundamental Research Funds for the Central Universities under Grant D5000220213; and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220890. *(Corresponding authors: Geng Chen; Jie Qin.)*

Geng Chen is with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: geng.chen.cs@gmail.com).

Jie Qin is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: qinjiebuaa@gmail.com).

Boulbaba Ben Amor is with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates (e-mail: boulbaba.amor@inceptioniai.org).

Weiming Zhou is with Guangzhou HeyGears IMC. Inc., Guangzhou 510705, China, and also with Guangdong Stomatology Hospital, Guangzhou 510280, China (e-mail: 578668032ming@gmail.com).

Hang Dai is with the School of Computing Science, University of Glasgow, G12 8QQ Glasgow, U.K. (e-mail: hang.dai.cs@gmail.com).

Tao Zhou is with the PCA Laboratory, the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 211106, China (e-mail: taozhou.ai@gmail.com).

Heyuan Huang is with Guangzhou HeyGears IMC. Inc., Guangzhou 510705, China (e-mail: hhuang@heygears.com).

Ling Shao is with the UCAS-Terminus AI Laboratory, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: ling.shao@ieee.org).

Digital Object Identifier 10.1109/TMI.2023.3263161

I. Introduction

A KEY step in computer-aided dental treatment planning is to detect the trim line between teeth and gingiva. This step provides valuable information for dentists and dental technicians to conduct a variety of tasks. A typical one is the 3D printing of aligners, where the trim line acts as the key guideline. Other tasks include the visualization of teeth, the use of perio tray model, the fabrication of dental treatment supplies (*e.g.*, dental crown, splint, and nightguard), orthodontics, etc [1], [2], [3]. Therefore, accurate tooth-gingiva trim line detection is of vital importance to successful dental treatment.

Advanced 3D intraoral scanners provide vivid mesh representations for dental surfaces [4]. Thereby, mesh-based trim line detection has become a challenging research problem. In particular, a number of works reformulate trim line detection as a tooth (binary) segmentation problem, where teeth are separated from the gingiva. Traditional tooth segmentation methods often rely on manual or semi-automatic techniques, which are time-consuming and easily biased towards the expert's experience [5], [6], [7]. These issues are effectively resolved by automatic learning-based tooth segmentation approaches, which learn the relationship between dental surfaces and tooth labels. In particular, those based on geometric deep learning (GDL) have shown promising performance due to the powerful ability of GDL in automatic feature learning, which provides more informative features in comparison with the hand-crafted ones. Recently, a number of automatic tooth segmentation methods have been developed [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. For instance, a typical 2D method was proposed by Xu et al. [11], where a patch-based 2D network

is utilized to learn tooth labels from the local geometric features extracted from dental surfaces. However, this method suffers from several limitations, *e.g.*, the time-consuming long pipeline and the overlook of the unique characteristics of meshes. To overcome these limitations, more advanced models, including those based on 3D GDL, are proposed. Lian et al. [15] proposed MeshSNet to segment teeth from dental surfaces in an end-to-end manner. MeshSNet employs graph-constrained learning models to extract the geometric features from dental surfaces and then predict tooth labels using the local-to-global features. In addition, a number of methods have been proposed for tooth segmentation based on PointCNN [12], Mask R-CNN [14], and graph CNN [18]. However, accurate tooth segmentation is still a challenging task due to the quality issue of acquired dental surfaces, differences between individuals' teeth, and tooth abnormalities [2], [11]. In addition, despite their advantages, existing tooth segmentation methods tend to suffer from two major limitations. First, no pooling operation is considered in the network design. This restricts the size of receptive fields, leading to limited context information. Second, existing networks follow very simple architectures, analogous to fully convolutional neural networks. There is a pressing need to consider more suitable segmentation network architectures, *e.g.*, U-Net [20] that has shown remarkable success in numerous medical image segmentation tasks [21].

Although tooth segmentation methods have shown great potentials, it is worth noting that reformulating trim line detection as a segmentation task is not an ideal solution since it suffers from two inherent limitations. First, existing tooth segmentation methods can only work with the low-resolution (LR) dental surface that is simplified from the original one [22]. Although the segmentation result on LR surface is promising, it becomes unsatisfactory when mapping the result back to the original high-resolution (HR) dental surface. In addition, the vital information of HR dental surfaces is not fully utilized in the segmentation models. As shown in Fig. 1, significant information is lost during the simplification of dental surface. Moreover, the tooth-gingiva boundary becomes unclear in the LR dental surface, as the region marked by blue arrow. Second, the tooth-gingiva trim line is the final result demanded by a large number of real applications, e.g., 3D printing of aligners. It allows easy human innervation/correction in comparison with the vertex-level segmentation result. Therefore, an end-to-end solution for extracting tooth-gingiva trim line possesses more significant application value and is greatly desired in practice.

To this end, we propose a novel geometric deep learning framework for tooth-gingiva trim line detection, which is capable of accurately detecting the boundary line between teeth and gingiva by making full use of the HR dental surface data. Our framework consists of two stages, where the first stage, called trim line proposal, aims to generate an initial trim line with a powerful trim line proposal network (TLP-Net) from LR dental surfaces, and the second stage, called trim line refinement, directly works on HR dental surfaces and is able to refine the trim line for the final output. The first stage, TLP-Net, utilizes convolution and pooling operations designed

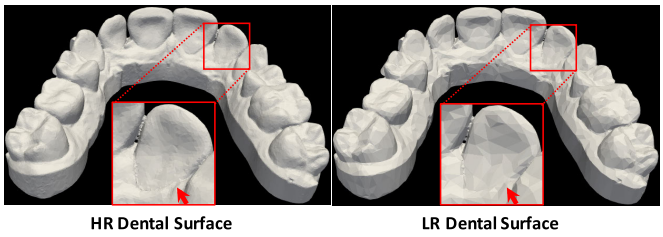

Fig. 1. Comparison of high-resolution (HR) and low-resolution (LR) dental surfaces. As shown in the region marked by a red arrow, the HR dental surface provides more clear boundary information than the LR one.

for meshes to extract robust features from dental surfaces, and then segment the teeth for extracting the initial trim line. It is based on an improved U-Net architecture with various novel features, including a proposed residual multi-scale attention fusion (RMAF) module for integrating the multi-scale features, and multiple feature bridge module for filling the gap between different levels of features. In addition, we propose a trim line loss to force the network to emphasize more on the tooth-gingiva trim line. As a result, our TLP-Net effectively resolves the drawbacks in existing tooth segmentation methods based on the elaborately designed network architecture with large receptive fields and effective mesh pooling/convolution operations. After segmentation, we extract the tooth-gingiva boundary line as the initial line proposal for our second stage, where we refine the trim line with our trim line refinement network (TLR-Net). The TLR-Net is based on a deep learning based LDDMM model, which refines the initial trim line with the edge information extracted from the HR dental surface. Furthermore, we introduce dense connections into TLR-Net for improved performance.

In summary, our contributions lie in four folds.

1) We propose a novel two-stage framework to automatically detect tooth-gingiva trim lines from dental surface. Our framework directly provides high-quality trim line prediction by making full use of the HR dental surfaces, resolving the drawbacks of existing methods effectively.
2) We propose TLP-Net, an effective trim line proposal network that fuses the multi-scale features from a mesh U-Net with the proposed RMAF module. In addition, we propose feature bridge and trim line loss to further improve accuracy of trim line.
3) We refine the initial trim line with our TLR-Net, which is a deep learning based LDDMM model with dense connections and is able to improve the accuracy of trim line by making full use of raw HR dental surfaces.
4) Extensive experiments on a clinical dataset consisting of 200 dental surfaces demonstrate that our method is capable of accurately detecting tooth-gingiva trim line on dental surfaces and outperforms cutting-edge methods. Further ablation studies demonstrate the effectiveness of proposed modules and loss function.

Our paper is organized as follows. In Section II, we review the related work. In Section III, we describe the proposed framework in detail. In Section IV, we present our experimental settings and the results. Finally, we conclude this work in Section V.

II. Related Work

Our framework is built upon geometric deep learning, which is a collection of deep learning techniques designed for non-Euclidean data, such as graphs and manifolds [23], [24]. In general, the geometric deep learning models can be classified into two categories, including those for graphs and the others for manifolds. For the models designed for graphs, graph neural networks (GNNs) are the main approaches due to their superior ability to enhance node representations by performing inter-node information propagation. Scarselli et al. [25] introduced the concept of GNNs for the first time in 2008. In their work, recursive neural networks are extended for processing graphs. Li et al. [26] proposed gated recurrent units (GRUs) to improve the representation capacity of GNNs. Gilmer et al. [27] proposed a message passing neural network to generalize the GNNs. Different GNN models were also proposed to address the challenges associated with graph learning, such as models for hyper-graphs [28], large-scale graphs [29], deep architectures [30], etc. GNNs have also been employed for medical image analysis, especially for disease prediction [31], [32], [33].

For the models designed manifolds, they can be classified into four major categories, including those based on voxels, multiviews, spatial information, and point clouds [24]. Voxel-based methods, such as 3D ShapeNets [34], first voxelize a 3D object into 3D grid-sampled volumetric data and then employ 3D CNNs for feature learning. Similarly, multiview-based methods, such as multiview CNN (MVCNN) [35], convert the 3D object into multiview data for feature learning with CNNs. Spatial-based methods include geodesic CNN (GCNN) [36], the first intrinsic variant of CNN on manifolds, anisotropic CNN (ACNN) [37], a deep learning model that employs anisotropic heat kernels as spatial weighting functions, mixture model network (MoNet) [38], a unified framework that extends CNN to non-Euclidean domains, etc. In particular, MeshCNN [22] was proposed to perform convolution on edges and pooling with an edge collapse mechanism. It shows promising performance in comparison with the cutting-edge models [22], [39]. Point-based methods are designed for point clouds, such as PointNet [40], PointNet++ [41], PointCov [42], etc. These geometric deep learning methods have also been employed for processing and analyzing dental surfaces. For instance, efforts have been dedicated to applying PointCNN [12] and graph CNN [18] to tooth segmentation. Our work differs from existing ones in terms of three aspects. First, we propose a two-stage trim line detection framework, which consists of a TLP-Net and a TLR-Net. Second, our TLP-Net is inspired by the MeshCNN [22], but with three improvements, including a proposed RMAF module, multiple feature bridge modules, and a specially-designed trim line loss. Third, we improve ResNet-LDDMM [43] with a residual dense architecture for TLR-Net, which refines the initial trim line provided by TLP-Net.

III. Method

In this section, we will first present the whole framework, illustrated in Fig. 2, for tooth-gingiva trim line detection. We will then detail the line proposal stage by explaining the mesh convolution/pooling techniques, the architecture of our proposal network, and the trim line loss. Finally, we will provide detailed descriptions for the trim line refinement stage and its role in the whole framework.

A. Proposed Framework

We propose a two-stage framework for tooth-gingiva trim line detection. The overall framework is shown in Fig. 2. In general, the proposed framework consists of two major stages. The first stage, trim line proposal, aims to predict an initial trim line from the LR dental surface with a proposed mesh CNN, which is trained in a supervised manner with the ground truth tooth segmentation results. The second stage, trim line refinement, further refine the initial trim line with the HR dental surface and a proposed deep learning based LDDMM model. Finally, we are able to generate the high-quality trim line prediction, which can be used in a large number of dental applications in clinical settings.

B. Trim Line Proposal

1) Mesh Convolution and Pooling: As shown in Fig. 3(A), any randomly selected edge (marked in red) on the dental surface falls within a "one-ring" structure [22], where this edge and its four neighboring edges constitute two faces. This local structure is invariant on dental surfaces, which, therefore, provides a good basis for performing convolution and pooling on edges. We show the close-up view of the one-ring structure in Fig. 3(B) and denote the target edge and its neighboring edges as e and (a, b, c, d), respectively. However, one potential issue is that the convolution value will be different once the initial ordering of mesh elements changes, *e.g.*, from (a, b, c, d) to (c, d, a, b). This issue is effectively resolved by a symmetric feature mapping, which converts the original edge features $[f(a), f(b), f(c), f(d)]$ to new ones $[f'(a), f'(b), f'(c), f'(d)]$ [22]. Therefore, the convolution for edge e can be defined as

$$C(e) = w_0 f'(e) + w_1 f'(a) + w_2 f'(b) + w_3 f'(c) + w_4 f'(d) \tag{1}$$

where $\{w_k | k = 0, 1, \ldots, 4\}$ are the weights to be learned.

The mesh pooling is based on an edge collapse framework, where the edges with unimportant features are collapsed/removed. This framework effectively preserves the edges with significant features, while reducing the total number of edges. As shown in Fig. 3(B), the edge collapse operation removes the target edge e and converts five original edges (e, a, b, c, d) to two new edges (h, i). The features of (h, i) are defined as

$$f(h) = \frac{1}{3}(f(e) + f(a) + f(b)), \tag{2}$$

$$f(i) = \frac{1}{3}(f(e) + f(c) + f(d)). \tag{3}$$

The unpooling is realized as an inverse procedure of pooling. Fig. 3(B) illustrates how collapsed edges are

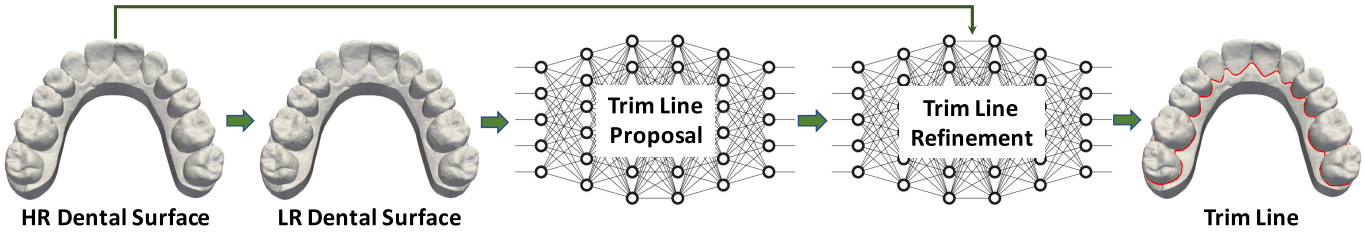

Fig. 2. A diagram of the proposed framework. In general, our framework consists of two stages, including one for trim line proposal and another for trim line refinement. The trim line proposal provides the initial trim line prediction with LR dental surfaces. The resulting trim line is then refined by trim line refinement by making full use of HR dental surface.

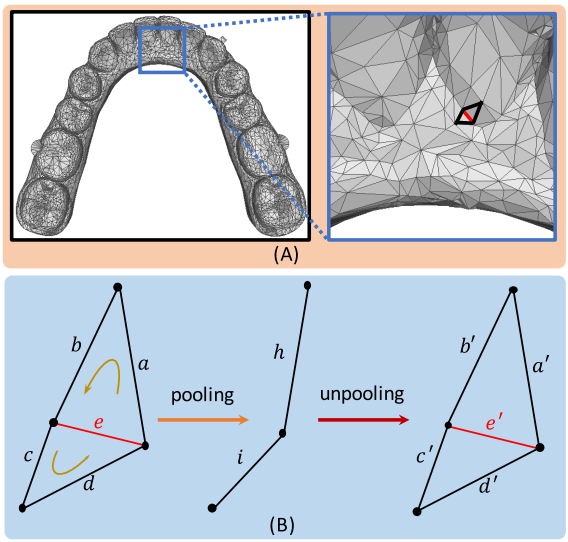

Fig. 3. Mesh convolution and pooling based on one-ring structure [22]. (A) shows the full view of dental surface and the close-up view of a local region. (B) shows the close-up view of the one-ring structure and the procedure of mesh convolution/pooling.

recovered back to (e', a', b', c', d'). The new edge features are defined as

$$f(e') = \frac{1}{2}(f(h) + f(i)), \tag{4}$$

$$f(a') = f(b') = f(h), \tag{5}$$

$$f(c') = f(d') = f(i). \tag{6}$$

2) Network Architecture: An overview of our TLP-Net is shown in Fig. 4. The backbone is a U-Net architecture, which is popular in medical image segmentation tasks. Compared with fully convolutional neural networks, U-Net enlarges the receptive field significantly by the use of pooling and an encoder-decoder architecture. In addition, the by-pass connections concatenate low-level and high-level features, which compensates for the fine-grained details lost during pooling. These two key features improve the network significantly and contribute to the promising performance of U-Net. In addition, we improve the U-Net architecture by introducing residual learning [44], which is particularly effective for resolving the gradient vanishing problem. Specifically, we replace the convolutional layers in the conventional U-Net with multiple residual blocks, where residual skip connections are introduced to ease the training. Instead of using a single residual block, we adopt a cascaded

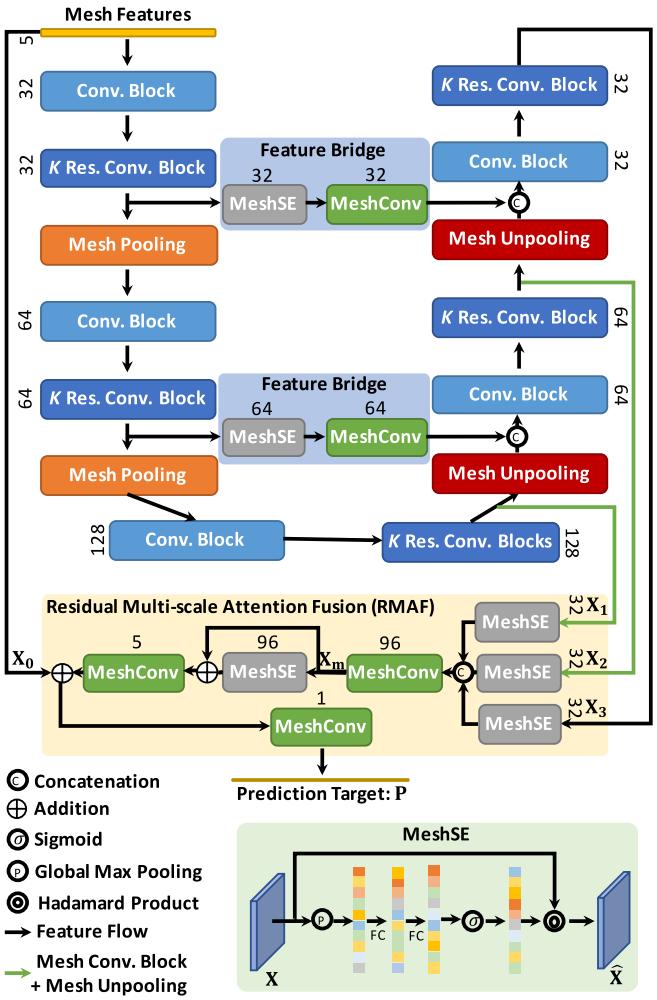

Fig. 4. An overview of the proposed trim line proposal network (TLP-Net). Our TLP-Net is based on a mesh U-Net with the proposed RMAF module for effective fusion of multi-scale feature and feature bridge modules for filling the gap between features at different levels. The number on the top of each block indicates the number of channels of output features.

architecture consisting of K repetitive residual blocks ($K = 2$ in our case). We further improve the network by proposing an effective feature fusion module and the feature bridge modules, which are detailed as follows.

a) Residual multi-scale attention fusion (RMAF): We propose a multi-scale feature fusion module, called RMAF, to integrate the features in the decoder branch so that rich multi-scale features can be used to segment teeth effectively. The RMAF module achieves effective feature fusion by making

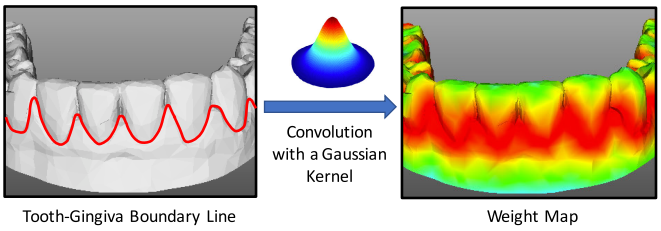

Fig. 5. A weight map generated from the tooth-gingiva trim line. The weight map is generated by convolving a Gaussian kernel with the tooth-gingiva trim line so that large weights are assigned to the regions near the trim line.

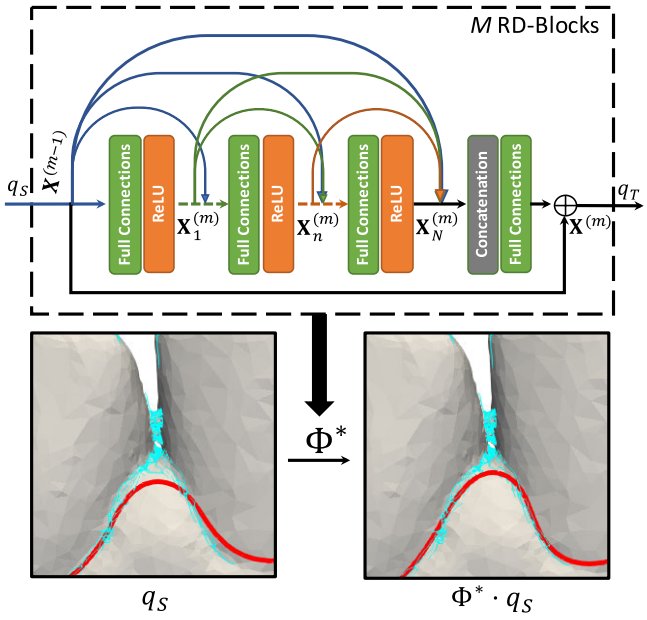

Fig. 6. An overview of our trim line refinement network (TLR-Net). The TLR-Net is composed of $M = 10$ residual dense (RD) blocks. It learns the deformation Φ^* between the smoothed initial trim line q_S (red line shown in bottom left) and the edges q_T (marked in blue color) extracted from the dental surface. Based on Φ^*, we transform q_S and is able to obtain the refined trim line $\Phi^* \cdot q_S$.

full use of the attention mechanism and residual learning. Existing works have demonstrated that the channel relationship can be used to recalibrate the multi-channel features for improved performance [45]. Inspired by this, we proposed a mesh squeeze-and-excitation component, called MeshSE, for exploiting the informative features with an attention mechanism. The refined feature provided by the MeshSE component is defined as follows

$$\hat{\mathbf{X}} = \text{MeshSE}(\mathbf{X}) = \text{Sigmoid}(\text{FC}(\text{FC}(\text{Pool}(\mathbf{X})))) \odot \mathbf{X}, \quad (7)$$

where $\mathbf{X}, \hat{\mathbf{X}} \in \mathbb{R}^{E \times C}$ are the input and output multi-channel features with E and C denoting the number of edges and number of channels. Sigmoid$(\cdot)$, FC$(\cdot)$, Pool$(\cdot)$, and $\odot$ represent a Sigmoid function, a fully-connected layer, a global average pooling layer, and Hadamard product (i.e., element-wise multiplication along the channel dimension), respectively.

In our RMAF module, we first merge the refined multi-scale features with a concatenation operation, i.e.,

$$\mathbf{X}_\text{m} = \text{MeshConv}(\text{Cat}(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3)), \quad (8)$$

where $\mathbf{X}_\text{m}$ is the merged feature, $\{\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3\}$ are the refined features provided by MeshSE components with inputs $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$, and MeshConv$(\cdot)$ denotes a mesh convolutional layer. We then pass $\mathbf{X}_\text{m}$ to a residual MeshSE component followed by a MeshConv layer, a global residual connection, and the final MeshConv layer. Mathematically, the final prediction $\mathbf{P} \in \mathbb{R}^{E \times 1}$ is defined as follows

$$\mathbf{P} = \text{MeshConv}(\text{MeshConv}(\hat{\mathbf{X}}_\text{m} + \mathbf{X}_\text{m}) + \mathbf{X}_0), \quad (9)$$

where $\hat{\mathbf{X}}_\text{m} = \text{MeshSE}(\mathbf{X}_\text{m})$ is the refined feature provided by a MeshSE component, $\mathbf{X}_0 \in \mathbb{R}^{E \times 5}$ is the input mesh features.

b) Feature bridge: U-Net directly concatenates low- and high-level features without considering the gap between different levels of features [46]. To resolve this issue, we first feed low-level feature to a MeshSE component followed by a MeshConv layer before passing it to the decoder branch so that the gap between features at encoder and decoder branches can be filled. Mathematically, given an input feature $\mathbf{X}$, our feature bridge module provides refined feature $\mathbf{X}'$, which is defined as follows

$$\mathbf{X}' = \text{MeshConv}(\text{MeshSE}(\mathbf{X})). \quad (10)$$

This operation fills the gap between different levels of features, which can improve the performance effectively.

3) Trim Line Loss: The main focus of our work is the tooth-gingiva trim line. Therefore, we propose a trim line (TL) loss to force our network to put more emphasis on the trim line. The TL loss is based on a cost-sensitive learning strategy that forces the model to pay more attention to hard/minority samples (i.e., those on trim lines) by enlarging their weights in the loss computation. Specifically, we first create a weight map by convolving the tooth-gingiva trim line with a Gaussian kernel $\mathcal{N}(0, k \times R)$, where $R = \sqrt{\text{SurfaceArea}}$ denotes the radius of a dental surface area [47], and k is a tuning parameter controlling the width of Gaussian kernel. An example of the weight map is shown in Fig. 5. As can be observed, large weights are assigned to the edges near tooth-gingiva trim line. We then apply this weight map to the computation of a weighted cross-entropy loss, *i.e.*,

$$\text{WCE}(p_i, y_i, w_i) = \begin{cases} -w_i \log(p_i) & \text{if } y_i = 1 \\ -w_i \log(1 - p_i) & \text{otherwise,} \end{cases} \quad (11)$$

where $y_i \in \{\pm 1\}$ is the ground truth label at edge $i \in \mathcal{M}$, with $\mathcal{M}$ denoting the mesh surface, $p_i \in [0, 1]$ is the predicted probability for the class with label $y_i = 1$, and w_i is the corresponding weight value in the weight map. Finally, we define our loss function $L(\mathcal{M})$ as

$$L(\mathcal{M}) = \sum_{i \in \mathcal{M}} \text{WCE}(p_i, y_i, w_i). \quad (12)$$

4) Trim Line Extraction: After obtaining the segmentation results, we extract the trim line, which will be used as the input for the next trim line refinement procedure. The trim line extraction is performed in an efficient way, where we detect the vertexes between segmented teeth and gingiva, which are provided by our segmentation map $\mathbf{P}$ in Eq. (9). These vertexes are then sorted and connected one-by-one to form the initial

trim line. Due to the resolution issue associated with LR mesh, the extracted trim line is not smooth and with large corner angles, which is adverse to the dental treatment planning and aligner 3D printing. Therefore, we smooth the trim line with a Gaussian kernel before performing the trim line refinement that is detailed in the following section.

C. Trim Line Refinement

1) Network Architecture: To refine the trim line with the rich information from HR dental surface, we further develop TLR-Net, a deep learning model based on residual dense network (RD-Net) [48] aiming to mimic the LDDMM framework from a deep learning perspective [43]. More specifically, we employ an RD-Net to learn the deformation between the smoothed initial trim line and the edges detected from HR dental surface. Existing works have demonstrated that the edges detected from a dental surface contain rich prior information on the tooth-gingiva boundary line, but suffer from heavy noise [11]. Therefore, we employ the TLR-Net to register the smoothed trim line to the detected edges so that the errors existing in the trim line can be corrected. Thanks to the robust TLR-Net, promising performance is achieved and the disturbance of noise in detected edges is avoided. The overall architecture of our TLR-Net is shown in Fig. 6. Mathematically, TLR-Net aims to minimize a functional (as in LDDMM [43]) as follows,

$$\Theta^*, \Phi^*(1) = \underset{\Theta=\{\theta^t\}_{t\in[0,1]}}{\operatorname{argmin}} \frac{1}{2\sigma^2}\mathcal{D}(\Phi^{\Theta}(1)\cdot q_S, q_T) + \frac{1}{2}\int_0^1 \|f^t(\Phi^{\Theta}(t)\cdot q_S, \theta^t)\|_2^2 dt, \quad (13)$$

where $\Theta = \{\theta^t\}_{t\in[0,1]}$ are the parameters of TLR-Net; $\Phi^{\Theta}(1)$ and $\Phi^{\Theta}(t)$ are the final and intermediate transformations, receptively; $f^t(\cdot, \theta^t)$ are time-dependant velocity vector fields computed by the TLR-Net; q_S and q_T are the smoothed trim line and target edges. σ is a tuning parameter balancing the contributions of two terms; $\mathcal{D}(\cdot,\cdot)$ is a term measuring how the deformed trim line fits the edges detected on the HR mesh, and is defined following the work in [43]. The second term, i.e., the regularization term, allow to achieve an optimal deformation (in terms of kinetic energy).

2) RD-Block: As shown in Fig. 6, the RD-Block consists of multiple fully connected (FC) layers densely connected with each other, a feature fusion unit, and a residual skip connection. Note that we use dash lines to denote the omitted fully connected layers and activation functions. A novel feature of RD-Net is the contiguous memory mechanism, which is achieved by passing the output feature of the last RD-Block to different layers of the current one. Mathematically, this is defined as

$$\mathbf{X}_n^{(m)} = \sigma(\mathrm{FC}(\mathrm{Cat}(\mathbf{X}^{(m-1)}, \mathbf{X}_1^{(m)}, \ldots, \mathbf{X}_{n-1}^{(m)}))), \quad (14)$$

where $\mathbf{X}_n^{(m)}$ denotes the feature at nth layer of mth RD-Block. σ represents the ReLU activation function. Thanks to the dense connections, the output feature of last RD-Block, $\mathbf{X}^{(m-1)}$, is considered in different layers of the current RD-Block, allowing effective contiguous memory.

Features from Eq. (14) are then fed to a feature fusion unit, which consists of a concatenation layer and an FC layer. After that, a residual connection is introduced for better gradient flow, which is able to alleviate the gradient vanishing problem effectively. The final output feature of mth RD-Block, $\mathbf{X}^{(m)}$, is defined as

$$\mathbf{X}^{(m)} = \mathrm{FC}(\mathrm{Cat}(\mathbf{X}^{(m-1)}, \mathbf{X}_1^{(m)}, \ldots, \mathbf{X}_n^{(m)}, \ldots, \mathbf{X}_N^{(m)})) + \mathbf{X}^{(m-1)}. \quad (15)$$

Considering the trade-off between performance and efficiency, we set N to 3 and M to 10.

3) Loss Function: Our loss function consists of two terms, i.e., a data fidelity term and a regularization term. The former one aims to narrow the difference between the initial trim line $\hat{q}_S = \Phi^* \cdot q_S$ and the edges q_T by minimizing the distance between two sets of point clouds. The latter one, defined as the integration of instantaneous kinetic energy terms, controls outputs of the RD-Blocks. Two terms are balanced using a tuning parameter θ, which is defined in Eq. (13).

For the data fidelity term, we use Chamfer's distance (CD), which measures the squared distance between each point from one set to the corresponding nearest neighbor point from the other set and vice versa. Mathematically, it is defined as

$$\ell_{\mathrm{CD}}(\hat{q}_S, q_T) = \sum_{x\in\hat{q}_S} \min_{y\in q_T} \|x-y\|^2 + \sum_{x\in q_T} \min_{y\in\hat{q}_S} \|x-y\|^2. \quad (16)$$

For the regularization term, inspired by the LDDMM algorithm, we use a summation of all kinetic energies at different time steps, i.e.,

$$\ell(f) = \frac{1}{2}\int_0^1 \ell_t\left(f^t\right) dt \quad (17)$$

with

$$\ell_t(f) = \left\| f^t\left(\Phi^{\Theta}(t,.), \theta^t\right)\right\|_{\ell_2}^2, \quad (18)$$

where $\|\|_{\ell_2}^2$ denotes the ℓ_2 norm. Consistent with LDDMM, TLR-Net minimizes the amount of energy spent to draw near to q_T from q_S, while going along the Orbit of q_S. The regularization term allows to fit the initial trim line to the detected edges while preserving its shape properties, i.e., being resistant to the noise.

Our TLR-Net is a revisited variant of diffeomorphic active contours in literature [49] using deep learning. The integration of smooth and regular time-dependant velocity vector fields computed on the original curve (trim line) allows to compute a diffeomorphic deformation of the original trim line to fit the target edges.

IV. Experiments

In this section, we first present the dataset, experimental settings, baseline methods, and evaluation metrics. We then describe the quantitative and qualitative results in detail.

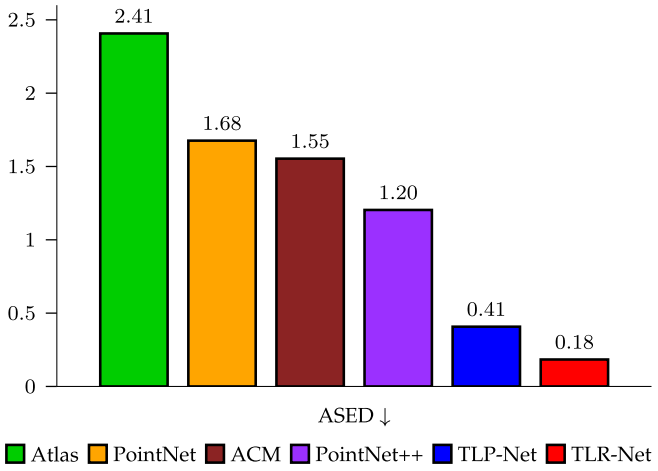

Fig. 7. ASED comparison. We compare our TLP-Net and TLR-Net with competing models, including Atlas, ACM, PointNet, PointNet++. A smaller ASED value indicates better performance.

A. Dataset and Implementation Details

1) Dataset: Our dataset consists of 200 dental surfaces from different subjects. All the surfaces were acquired using a 3D intraoral scanner. The original HR surfaces have around 200,000 faces, providing rich information for tooth-gingiva trim line detection. Due to the memory limit, we simplify each surface to 10,000 faces for the training and testing of line proposal network. The ground truth labels are transformed from the HR ones via a nearest neighbor search strategy.[1] We randomly select 160 surfaces for training and utilize the remaining 40 for testing.

2) Experimental Settings: In all experiments, we train the TLP-Net using an ADAM optimizer [50] with an initial learning rate of 1e-3 and an exponential decay rate of 0.95. We set the batch size to 6 and the number of epochs to 50. For the TLR-Net, we set the learning rate to 1e-5 and the number of epochs to 500. The networks are implemented using Python with deep learning frameworks, including PyTorch (TLP-Net) and TensorFlow (TLR-Net). They are trained using an NVIDIA Tesla V100 GPU with 32GB RAM. It takes about eight hours for training the TLP-Net and two minutes for testing with the trained model. The refinement network takes about ten minutes for inference, which is also efficient. The resources of this work are publicly available at https://github.com/supergengchen/Automatic-Trim-Line-Detection.

B. Baseline Methods and Evaluation Metric

To demonstrate its effectiveness, we compare Proposal-Net with various popular methods, including an atlas-based mesh segmentation method similar to [51], and [52], a 3D activate contour model (ACM) based method [53], and a deep-based point-cloud segmentation network, PointNet [40], and its improved version, PointNet++ [41].

We evaluate the accuracy of predicted trim lines using average squared Euclidean distance (ASED). To establish the correspondence between the point sets of two lines, we first perform Iterative Closest Point (ICP) algorithm with alignICP provided by vedo package.[2] The ASED is then defined as

$$\text{ASED} = \frac{1}{|\Omega|}\sum_{i\in\Omega}||P(i) - G(i)||_2^2, \tag{19}$$

where Ω is the point set, $P(i)$ and $G(i)$ is ith point at predicted and ground truth trim lines, respectively.

C. Experimental Results

1) Quantitative Results: We first evaluate the detected trim lines using ASED. The corresponding results, shown in Fig. 7, indicate that our proposal network, TLP-Net, outperforms the competing methods remarkably. Specifically, compared with the second-best method, PointNet++, our TLP-Net reduces the ASED value by a large margin of 0.79. In contrast, the baseline method, Atlas, provides the worst performance with a large ASED value. PointNet and ACM, which show comparable performance, outperform Atlas, but are still unsatisfactory. After refining the trim line, the result is further improved. Our TLR-Net outperforms TLP-Net in terms of ASED value by a large margin of 0.23. This demonstrates the effectiveness of the refinement network, which is able to make full use of HR dental surfaces and improves the accuracy of detected trim lines effectively. From a clinical perspective, the accurate trim line provided by our model will significantly improve a large number of dental applications, such as the accuracy of 3D printing of aligners, tooth treatment planning, etc.

2) Qualitative Results: We then show the qualitative results for different testing subjects in Fig. 8. As can be observed, both our TLP-Net and TLR-Net provide promising results, which are close to the ground truth. In contrast, the baseline methods, Atlas and ACM, show unstable performance. Atlas tends to fail when there is a large gap between the template and individual dental surface, while ACM suffers from the same issue and is easily influenced by the noisy edges of the dental surface, e.g., the results shown in the third row. Note that the template here is obtained by finding the cluster center of the trim lines of training data, i.e., the trim line that is closest to all other trim lines in terms of ASED. PointNet and PointNet++ improve the stability of the results, but show unsatisfactory performance in terms of the accuracy of trim lines. As shown in Fig. 8, our TLP-Net resolves the drawbacks of baseline methods and significantly improves the accuracy of detected trim lines. Furthermore, our TLR-Net improves the results given by TLP-Net, and is able to improve the smoothness of trim lines in order to make the trim lines fit the dental surfaces well. The final trim lines given by our method are very close to the ground truth, providing valuable information for a number of important tasks in dental treatment.

We also show a number of typical pathological cases in the bottom four rows of Fig. 8. As can be observed, our models consistently achieve the best performance when there exist abnormal teeth, such as teeth with orthodontic brace (fifth

[1] https://www.cs.umd.edu/ mount/ANN/

[2] https://vedo.embl.es/

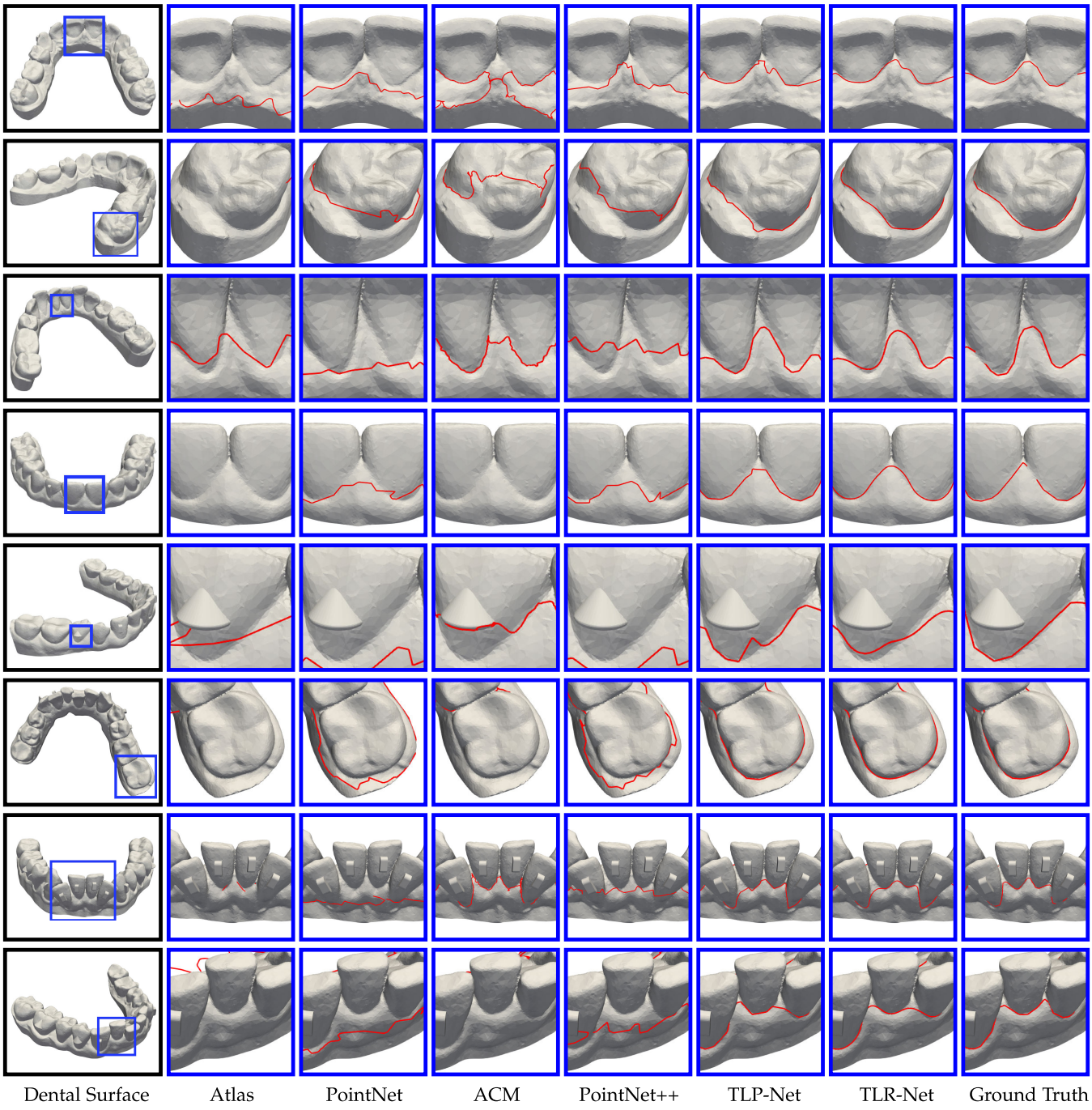

Fig. 8. Visual comparison of trim lines. Our TLP-Net and TLR-Net provide trim lines with significantly improved quality in comparison with competing models, especially for the cases with abnomal teeth.

row), malpositioned teeth (sixth rows), and opsigenes (bottom two rows).

D. Ablation Study

1) Trim Line Proposal Network: We perform extensive ablation experiments to investigate the effectiveness of key components of our TLP-Net, which include RMAF, feature bridge, and trim line (TL) loss. The quantitive ASED results are summarized in Fig. 7. It is worth noting that the model, called "TL Loss ($k = 7\%$)," is the full version of our TLP-Net with default parameters.

a) Effectiveness of RMAF: To investigate the effectiveness of RMAF module, we construct an ablated version, called "w/o RMAF Module," where we remove the RMAF module and directly feed $\mathbf{X_3}$ (i.e., the final feature of U-Net) to a MeshConv layer for the prediction P. The results, shown in Fig. 9, indicate that the full version of our model, "TL Loss ($k = 7\%$)," outperforms "w/o RMAF Module" by a large margin of 0.22, demonstrating that the RMAF module is able to improve the accuracy of detected trim line effectively. The superior abilities of RMAF module owe to its effective fusion of multi-scale features

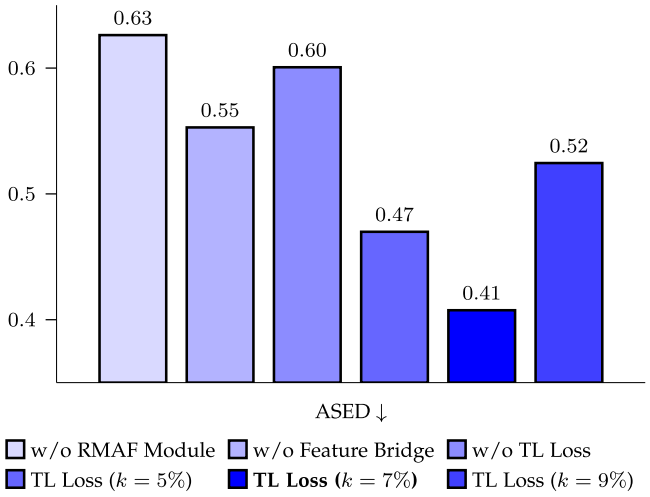

Fig. 9. Ablation results for TLP-Net. The proposed modules contribute to the superior performance of TLP-Net. The TL loss provides the best performance when $k = 7\%$. Note that the version "TL Loss ($k = 7\%$)" is the full version of our TLP-Net with default parameters.

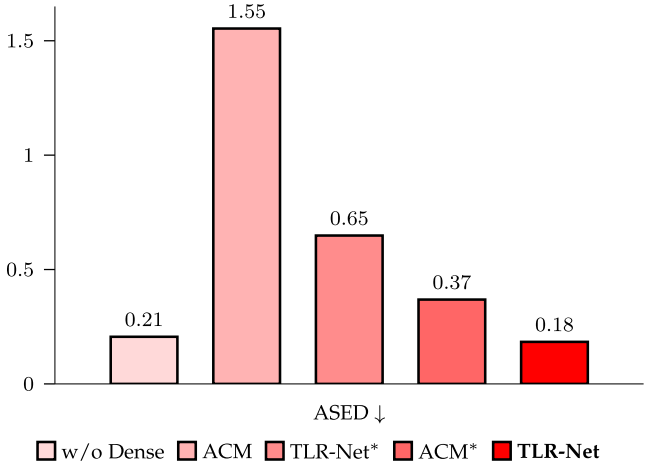

Fig. 10. Ablation results for TLR-Net. Both TLP-Net and TLR-Net are the key factors determining the success of accurate trim line detection.

by making full use of the attention mechanism and residual learning.

b) Effectiveness of feature bridge: We then construct the second ablated version, called "w/o Feature Bridge," to investigate the effectiveness of feature bridge modules. Specifically, we replace the feature bridge modules with direct concatenation optations, which is used in the classical U-Net model. The results, shown in Fig. 9, indicate that the performance decreases after replacing feature bridge with concatenation, i.e., "TL Loss ($k = 7\%$)," outperforms "w/o Feature Bridge." This demonstrates that the feature bridge is an effective module capable of improving performance by filling the gap between different levels of features.

c) TL loss and optimized k: Finally, we evaluate the TL loss and investigate the optimized setting for parameter k, which controls the width of the Gaussian kernel used to generate the weight map. First, we replace the TL loss with a classical binary cross-entropy (BCE) loss, which provides an ablated version, called "w/o TL Loss." As can be observed in Fig. 9, classical BCE loss shows unsatisfactory performance in comparison with TL loss, demonstrating the effectiveness of TL loss sufficiently. By assigning different weights to the dental surface, TL loss is able to enforce the network putting more emphasis on the tooth-gingiva trim line for improved performance.

Furthermore, we study the influence of Gaussian kernel width by performing experiments with different $k = 5\%, 7\%, 9\%$. The results, shown in Fig. 9, indicate that $k = 7\%$ provides the best performance with the lowest ASED value. Therefore, we adopt $k = 7\%$ as the default setting for our TL loss.

2) Trim Line Refinement Network: In this section, we investigate the effectiveness of our TLR-Net by (i) performing experiments with an ablated version without dense connections, (ii) comparing TLR-Net with an improved ACM model, ACM* (i.e., ACM with the initial trim line provided by TLP-Net), and (iii) studying the influence of initial trim lines by comparing TLP-Net with a modified version, TLR-Net*, where the initial trim line is the atlas trim line. The corresponding experimental results are summarized in Fig. 10.

a) Effectiveness of dense connections: To demonstrate the effectiveness of dense connections, we construct an ablated version, called "w/o Dense," by replacing dense connections with a residual skip connection. The results, shown in Fig. 10, indicate that "TLR-Net" outperforms "w/o Dense," implying that dense connections are able to improve the performance effectively. This owes to the superior ability of dense connections in terms of resolving the problem of vanishing gradients.

b) Comparison with ACM: We then investigate the advantage of our model over the ACM model. To establish a fair comparison, we create an ablated version, "ACM*," which shares the same initial trim line (i.e., the one given by TLP-Net) with our model. The results, shown in Fig. 10, suggest two conclusions, including (i) TLP-Net is able to improve the performance of ACM effectively by proving a good initialization for it, and (ii) TLR-Net outperforms ACM remarkably when they are using the same initial trim line from TLP-Net, further demonstrating the effectiveness of both TLP-Net and TLR-Net.

c) Benefit of TLP-net: Finally, we investigate the benefit of TLP-Net by using the results of Atlas as the initialization for TLR-Net. As shown in Fig. 10, the corresponding ablated version, called "TLR-Net*," shows unsatisfactory performance in comparison with the full version of our model. The results demonstrate the benefit of TLP-Net, which is capable of providing a good initialization for the TLR-Net and plays a crucial role in the whole framework.

E. Model Size and Computational Efficiency

Furthermore, we perform experiments to compute the model size, speed, and memory usage of our models and representative deep learning models. The results, as shown in Table I, indicate that our models outperform representative models in terms of both model size and memory usage. Our models also show promising performance in the evaluation of

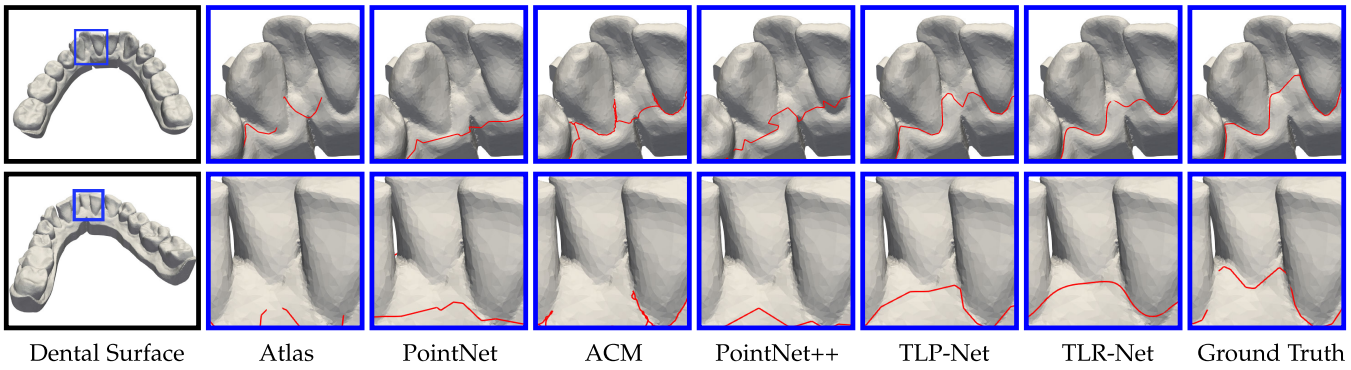

Fig. 11. Failed cases. We show the failed cases for two challenging scenarios here.

TABLE I
COMPARISON OF OUR TLP-NET AND TLR-NET WITH REPRESENTATIVE DEEP LEARNING MODELS IN TERMS OF THE MODEL SIZE, MEMORY, AND SPEED

Method	Model Size (M)	Memory (G)	Speed (FPS)
PointNet [40]	7.14	4.27	129.00
PointNet++ [41]	1.74	9.21	30.40
TLP-Net (Ours)	1.62	2.38	49.43
TLR-Net (Ours)	0.51	1.22	10.20

model speed. Although PointNet is with a high speed, it shows unsatisfactory performance in the detection accuracy of trim lines.

F. Failed Cases

Finally, we show two typical failure cases in Fig. 11. As can be observed, our model may fail in some extremely challenging cases. However, it is worth noting that our model consistently outperforms the existing cutting-edge models in these cases. In the future, we will explore improving the performance of our models with more advanced trim-line-aware network designs and larger-scale datasets covering different challenging scenarios.

V. CONCLUSION

In this paper, we proposed a two-stage deep learning framework for automatic tooth-gingiva trim line detection. Our framework consists of two effective sub-networks, TLP-Net and TLR-Net, for trim line proposal and refinement, respectively. The TLP-Net is based on a mesh U-Net architecture with various novel features, including the RMAF module for fusing multi-scale features, the feature bridge modules for filling the gap between different levels of features, and the trim line loss for forcing the network to specifically focus on the tooth-gingiva trim line. The TLR-Net refines the initial trim line with a deep-based LDDMM model with dense connections. Our method overcomes the limitation of existing methods with a two-stage framework, which is able to make full use of high-resolution dental surface information for generating accurate trim lines. Extensive experiments on a clinical dental surface dataset demonstrate that our method outperforms various cutting-edge segmentation methods, both quantitatively and qualitatively.

REFERENCES

[1] T. Kondo, S. H. Ong, and K. W. C. Foong, "Tooth segmentation of dental study models using range images," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 350–362, Mar. 2004.

[2] C. Lian et al., "MeshSNet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D dental surfaces," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 837–845.

[3] M. P. Muresan, A. R. Barbura, and S. Nedevschi, "Teeth detection and dental problem classification in panoramic X-ray images using deep learning and image processing techniques," in *Proc. Int. Conf. Intell. Comput. Commun. Process.*, 2020, pp. 457–463.

[4] C. B. Martin, E. V. Chalmers, G. T. McIntyre, H. Cochrane, and P. A. Mossey, "Orthodontic scanners: What's available?" *J. Orthodontics*, vol. 42, no. 2, pp. 136–143, 2015.

[5] B. H. Le, Z. Deng, J. Xia, Y.-B. Chang, and X. Zhou, "An interactive geometric technique for upper and lower teeth segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2009, pp. 968–975.

[6] M. Grzegorzek, M. Trierscheid, D. Papoutsis, and D. Paulus, "A multi-stage approach for 3D teeth segmentation from dentition surfaces," in *Proc. Int. Conf. Image Signal Process.* Cham, Switzerland: Springer, 2010, pp. 521–530.

[7] B.-J. Zou, S.-J. Liu, S.-H. Liao, X. Ding, and Y. Liang, "Interactive tooth partition of dental mesh base on tooth-target harmonic field," *Comput. Biol. Med.*, vol. 56, pp. 132–144, Jan. 2015.

[8] H. Lamecker et al., "Automatic detection and classification of teeth in CT data," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2012, pp. 609–616.

[9] J. Keustermans, D. Vandermeulen, and P. Suetens, "Integrating statistical shape models into a graph cut framework for tooth segmentation," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2012, pp. 242–249.

[10] A. Wirtz, S. G. Mirashi, and S. Wesarg, "Automatic teeth segmentation in panoramic X-ray images using a coupled shape model in combination with a neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 712–719.

[11] X. Xu, C. Liu, and Y. Zheng, "3D tooth segmentation and labeling using deep convolutional neural networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 7, pp. 2336–2348, Jul. 2018.

[12] F. G. Zanjani et al., "Deep learning approach to semantic segmentation in (3D) point cloud intra-oral scans of teeth," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2019, pp. 557–571.

[13] S. Tian, N. Dai, B. Zhang, F. Yuan, Q. Yu, and X. Cheng, "Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks," *IEEE Access*, vol. 7, pp. 84817–84828, 2019.

[14] F. G. Zanjani et al., "Mask-MCNet: Instance segmentation in 3D point cloud of intra-oral scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 128–136.

[15] C. Lian et al., "Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3D intraoral scanners," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2440–2450, Jul. 2020.

[16] Y. Liang, W. Song, J. Yang, L. Qiu, K. Wang, and L. He, "X2Teeth: 3D teeth reconstruction from a single panoramic radiograph," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 400–409.

[17] J. Zhang, C. Li, Q. Song, L. Gao, and Y.-K. Lai, "Automatic 3D tooth segmentation using convolutional neural networks in harmonic parameter space," *Graph. Models*, vol. 109, May 2020, Art. no. 101071.

[18] D. Sun et al., "Automatic tooth segmentation and dense correspondence of 3D dental model," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 703–712.

[19] F. Zhang, J. Zhu, P. Hao, F. Wu, and Y. Zheng, "BDU-Net: Toward accurate segmentation of dental image using border guidance and feature map distortion," *Int. J. Imag. Syst. Technol.*, vol. 32, no. 4, pp. 1221–1230, 2022.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[21] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, 2019.

[22] R. Hanocka, A. Hertz, N. Fish, R. Giryes, S. Fleishman, and D. Cohen-Or, "MeshCNN: A network with an edge," *ACM Trans. Graph.*, vol. 38, no. 4, p. 90, 2019.

[23] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.

[24] W. Cao, Z. Yan, Z. He, and Z. He, "A comprehensive survey on geometric deep learning," *IEEE Access*, vol. 8, pp. 35929–35949, 2020.

[25] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2016.

[26] Y. Li, R. Zemel, M. Brockschmidt, and D. Tarlow, "Gated graph sequence neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–20.

[27] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1263–1272.

[28] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3558–3565.

[29] J. Chen, T. Ma, and C. Xiao, "FastGCN: Fast learning with graph convolutional networks via importance sampling," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–15.

[30] Q. Li, Z. Han, and X.-M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.

[31] X. Song et al., "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101947.

[32] S. Yu et al., "Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 228–237.

[33] X. Zhao, F. Zhou, L. Ou-Yang, T. Wang, and B. Lei, "Graph convolutional network analysis for mild cognitive impairment prediction," in *Proc. Int. Symp. Biomed. Imag.*, 2019, pp. 1598–1601.

[34] Z. Wu et al., "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1912–1920.

[35] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 945–953.

[36] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst, "Geodesic convolutional neural networks on Riemannian manifolds," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Jun. 2015, pp. 37–45.

[37] D. Boscaini, J. Masci, E. Rodola, and M. Bronstein, "Learning shape correspondence with anisotropic convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3189–3197.

[38] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model CNNs," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5115–5124.

[39] G. Chen, H. Dai, T. Zhou, J. Shen, and L. Shao, "Automatic Schelling points detection from meshes," *IEEE Trans. Vis. Comput. Graphics*, early access, Jan. 19, 2022, doi: 10.1109/TVCG.2022.3144143.

[40] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.

[41] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5099–5108.

[42] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9621–9630.

[43] B. B. Amor, S. Arguillère, and L. Shao, "ResNet-LDDMM: Advancing the LDDMM framework using deep residual networks," 2021, *arXiv:2102.07951*.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[46] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," *Med. Image Anal.*, vol. 54, pp. 193–206, May 2019.

[47] X. Chen, A. Saparov, B. Pang, and T. Funkhouser, "Schelling points on 3D surface meshes," *ACM Trans. Graph.*, vol. 31, no. 4, p. 29, 2012.

[48] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.

[49] F. Arrate, J. T. Ratnanather, and L. Younes, "Diffeomorphic active contours," *SIAM J. Imag. Sci.*, vol. 3, no. 2, pp. 176–198, 2010.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[51] J. Thariat, L. Ramus, P. Maingon, V. Gregoire, S. Marcie, and G. Malandain, "Construction of a dental atlas for automatic teeth segmentation to improve post-irradiation dental care management," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 78, no. 3, pp. 452–453, 2010.

[52] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-D model-based neuroanatomical segmentation," *Hum. Brain Mapping*, vol. 3, no. 3, pp. 190–208, 1995.

[53] T. Kronfeld, D. Brunner, and G. Brunnett, "Snake-based segmentation of teeth from virtual dental casts," *Comput.-Aided Des. Appl.*, vol. 7, no. 2, pp. 221–233, 2010.