IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 25, 2019, accepted June 18, 2019, date of publication June 21, 2019, date of current version July 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924262

Automatic Classification and Segmentation of Teeth on 3D Dental Model Using Hierarchical Deep Learning Networks

SUKUN TIAN[1], NING DAI[1], BEI ZHANG[1], FULAI YUAN[1], QING YU[2], AND XIAOSHENG CHENG[1]
[1]College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[2]Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing 210008, China

Corresponding author: Ning Dai (dai_ning@nuaa.edu.cn)

This work was supported in part by the Funding from the National Key R&D Projects, China, under Grant 2018YFB1106903, in part by the National Natural Science Foundation of China under Grant 51775273, in part by the Natural Science Foundation of Jiangsu Province, China, under Grant BK20161487, in part by the Six Talent Peaks Project in Jiangsu Province, China, under Grant GDZB-034, in part by the Jiangsu Province Science and Technology Support Plan Project, China, under Grant BE2018010-2, and in part by the Jiangsu Provincial Health and Family Planning Commission's 2017 Research Project, China, under Grant H201704.

ABSTRACT To solve the problem of low efficiency, the complexity of the interactive operation, and the high degree of manual intervention in existing methods, we propose a novel approach based on the sparse voxel octree and 3D convolution neural networks (CNNs) for segmenting and classifying tooth types on the 3D dental models. First, the tooth classification method capitalized on the two-level hierarchical feature learning is proposed to solve the misclassification problem in highly similar tooth categories. Second, we exploit an improved three-level hierarchical segmentation method based on the deep convolution features to conduct segmentation of teeth-gingiva and inter-teeth, respectively, and the conditional random field model is used to refine the boundary of the gingival margin and the inter-teeth fusion region. The experimental results show that the classification accuracy in Level_1 network is 95.96%, the average classification accuracy in Level_2 network is 88.06%, and the accuracy of tooth segmentation is 89.81%. Compared with the existing state-of-the-art methods, the proposed method has higher accuracy and universality, and it has great application potential in the computer-assisted orthodontic treatment diagnosis.

INDEX TERMS Tooth segmentation, CNN, sparse voxel octrees, hierarchical classification.

I. INTRODUCTION

Segmentation of individual tooth from 3D dental models is a key technique in computer-aided orthodontic systems. Malocclusion is a common oral disease with a high prevalence, and its prevalence is about 50% [1]. The dental models play an important role in the clinical orthodontic diagnosis. They can truly show the 3D anatomy structure of patients with malocclusion, as well as the shape and position distribution of the teeth, and assist dentists to design an efficient and accurate dental treatment plan by extracting, moving and rearranging the teeth from the dental models [2], [3]. Therefore, tooth segmentation is a core step in many oral medical research processes and is the basis for computer- aided dental diagnosis and treatment.

With the improvement of computer hardware and software technology, many commercial CAD/CAM software for orthodontics, such as 3Shape, Implant3D and OrthoCAD, have emerged, and which can realize the automatic tooth segmentation to a certain extent. However, due to the complexity of interactive operation and the high degree of manual intervention in orthodontic CAD systems, their segmentation efficiency are lower. Orthodontic patients usually have symptoms such as odontoloxy, crowded dislocation between adjacent teeth, missing teeth and indistinctive tooth boundary. To solve this problem, researchers try to achieve automatic tooth segmentation by improving and optimizing algorithms, but these algorithms lack robustness and the segmentation effect is less than ideal. Kondo *et al.* [4] extracted the segmentation contour of the tooth based on the range image information calculated by using the 3D dental model and the shape of the dental arch. Grzegorzek *et al.* [5] presented a multi-stage approach for teeth segmentation from dentition surfaces based on a 2D model-based contour retrieval algorithm. Poonsri *et al.* [6] proposed a method to segment teeth from a panoramic dental x-ray image by means of

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

tooth area identification and template matching. However, these methods are unreliable for segmenting teeth in terms of irregular teeth arrangements and indistinctive tooth boundary, and which cannot accurately express the 3D shape of teeth. Therefore, many segmentation methods based on the three-dimensional dental model have emerged. Li *et al.* [2] proposed an interactive tooth segmentation method based on fast marching watersheds and threshold- controlling method. Wu *et al.* [7] and Yuan *et al.* [8] adopted morphologic skeleton operations to optimize the feature regions and identify the tooth boundaries, and achieved good segmentation results. Kumar *et al.* [9] used shortest path search algorithm and surface curvature field to separate gums and individual teeth from dental models, but this method is not suitable for the case of tooth loss or dislocation. For the complicated dental model with severe malformed and indistinctive tooth boundaries, Zou *et al.* [10] and Li and Wang [11] proposed an interactive tooth segmentation method based on harmonic fields, but which required too many surface points as prior. Fan *et al.* [12] presented an approach to segment and optimize the tooth boundaries using anisotropic filtering and agglomerative clustering, and which is based on the compact shape prior technique. Despite extensive research on tooth segmentation, the automatic and accurate tooth segmentation has become extremely difficult due to the specificity of tooth morphology, the tight line between the teeth and gingival tissues, and the tight fusion between the teeth.

In recent years, with deep learning becoming a research hotspot in the field of computer vision. The typical deep learning model, represented by CNN, has strong robustness and universality, and made breakthroughs in the field of medical diagnosis [13], [14]. Wang *et al.* [15] presented a supervised learning method based on CNN and ensemble random forests for automatic retinal blood vessel segmentation. According to the characteristic expression ability of CNN and stacked convolutional auto-encoders, Kallenberg et al. [16] constructed a convolutional sparse autoencoder (CSAE) model to apply to breast density segmentation and mammographic risk scoring. Shen *et al.* [17] proposed a multi-scale convolutional neural networks (MCNN) architecture for lung nodule diagnostic classification, and achieved 86.84% accuracy on nodule classification. Kamnitsas *et al.* [18] presented a fully automatic approach for brain lesion segmentation in multimodal brain MRI by the use of parallel convolutional pathway architecture, and the segmentation results exceeded the level of human expert delineation. Because deep learning has the ability to automatically learn high-level and more discriminative features from the sample dataset, researchers have begun to focus on the application of convolutional features in the intelligent diagnosis of oral diseases. Based on GoogLeNet Inception v3 architecture, Lee *et al.* [19] introduced a method to evaluate the efficacy of deep CNN algorithms for detection and diagnosis of dental caries. Miki *et al.* [20] employed the pre-trained AlexNet network to classify tooth types on dental cone-beam CT images, and the ROIs including single teeth were successfully extracted from CT slices. Xu *et al.* [21] designed a DCNN architecture based on [22] for 3D dental model segmentation, and which improved the tooth segmentation boundary accuracy by combining boundary-aware simplification method and fuzzy clustering algorithm.

Based on the study of 3D dental segmentation methods, an automated segmentation and classification method for teeth models is presented in this paper, in which sparse voxel octrees and conditional random field (CRF) model are used. It not only has high accuracy and robustness, but also has less manual intervention and parameter adjustment. To the best of our knowledge, it is the first attempt that makes use of hierarchical feature learning framework based on 3D CNN for automatically extracting high-dimensional features from 3D teeth models to segment and classify the tooth types. More importantly, the proposed method is robust to the patient's dental model with various complex symptoms such as odontoloxy, feature-less regions, crowding teeth, as well as missing teeth. Specifically, the main contributions of this study include the following:

- A general and robust tooth segmentation and classification framework which achieves 88.06% accuracy for highly similar tooth categories, and 89.81% for individual tooth segmentation.
- An optimized design of two-level hierarchical classification network architecture can solve the misclassification problem in highly similar tooth categories.
- An improved three-level hierarchical segmentation network based on conditional random field to refine the segmentation boundary, and which is robust to various complex malformations in patient teeth.
- Last but not least, the proposed model is flexible and extendible and can be trained again to generalize for the new dental samples, which allows the framework to improving the intelligence level of orthodontic CAD systems.

II. MATERIAL AND METHODS

A. OVERVIEW OF THE PROPOSED METHOD

The proposed methods for tooth segmentation and classification based on 3D CNN mainly consists of three steps as shown in Fig. 1. (1) Data preprocessing stage, the sparse octree partitioning method is used to label the original dental model, and then the tooth model is labeled as the input model (octree model). (2) Tooth classification stage, the preprocessed octree models are input into two-level hierarchical feature learning network for training to perform the classification task. Level-1 network is a 4-label classification network, which is used to distinguish the types of incisors, canines, premolars, and molars. Level-2 network is a 2-label classification network, which is used to distinguish the types of central and lateral incisors, first and second premolars, and first and second molars respectively. (3) Dental model segmentation stage, the designed three-level hierarchical network with similar layers is trained to obtain three caffe-models with different weight parameters respectively.

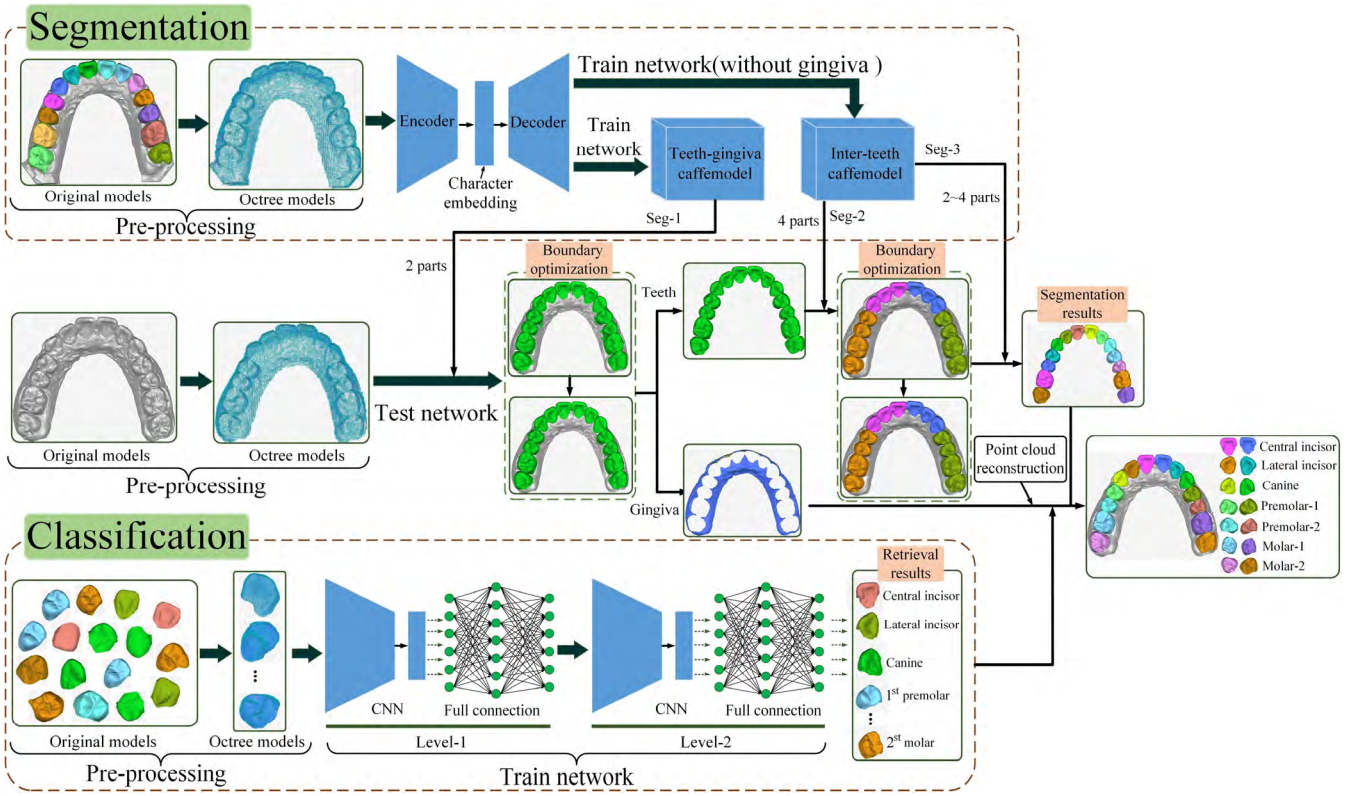


FIGURE 1. **Outline of the proposed hierarchical feature learning method for tooth segmentation and classification.**

The segmentation of the teeth and gingiva is completed by the teeth-gingiva caffemodel, and then the other two inter-teeth caffemodels are used to complete the segmentation of the individual teeth. Finally, the conditional random field model is used to optimize the boundary of the gingival margin and the inter-teeth fusion region.

B. DATASET PRE-PROCESSING

In order to verify the validity and accuracy of the proposed method for automatic segmentation and classification of dental models. A high-resolution laser scanner (3Shape D700, Denmark) is used to obtain a 3D digital dental model that has reliable quality and is represented as PLY format, and the scanned dental samples are randomly divided into three subsets: training, validating, and testing set.

Convolutional neural networks usually rely on a supervised learning method based on gradient descent method for network training. When the training dataset is enough large, the network can perform a stronger feature expression ability and self-learning function. But when the training dataset is limited, the network has strong sample-dependent characteristics [14]. Therefore, the virtual sample generation technology based on perturbation method mentioned in the literature [23], [24] is used to realize sample expansion, which means that we can increase the number of samples by rotating each dental model along the upright direction uniformly to improve the generalization ability of the network.

The point cloud model has the characteristics of irregular spatial relationship in the segmentation or classification process, so the existing network model cannot be directly applied to the point cloud model. Traditional voxel-based 3D CNN usually use full voxels in space to express the 3D shape information of the model, and which adopts the adaptive spatial partitioning method of octree to improve space utilization. However, since the memory and computation cost grow as the depth of the octree increases (Fig. 2), these methods become prohibitively expensive for high-resolution 3D models [23]. These full-voxel-based CNNs are limited to low resolution 3D models due to the high memory and computational cost. Therefore, according to the high-efficiency of hash method in data storage and computation, we construct an octree sparse expression model based on hash table to process the dental model into a sparse point cloud with labels, and the labels are set for the point cloud model of teeth and gingiva and each point in the model respectively. At the same time, we take the average normal vectors of dental model sampled in the finest leaf octants of the octree structure as input, and the efficiency of the method is reflected.

C. NETWORK ARCHITECTURE

In recent years, deep CNNs have been developed in a deeper, more flexible and effective training direction. The purpose is to reduce the computational complexity and memory consumption of the network, and to obtain better feature

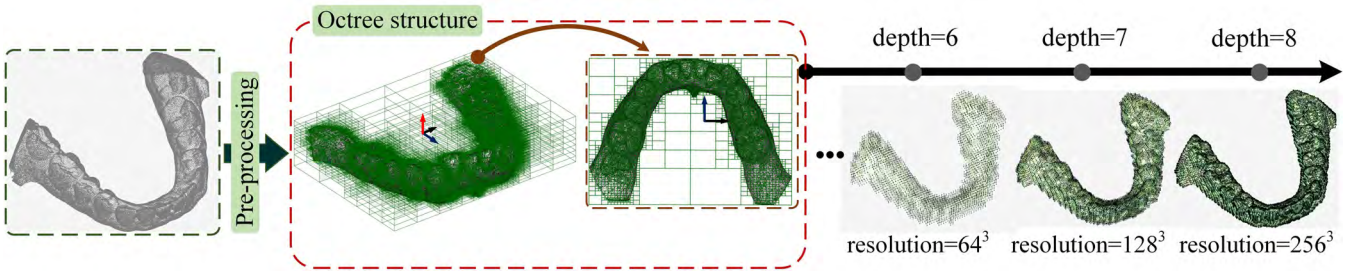


FIGURE 2. Dental model preprocessing, and the low-resolution octree model is gradually refined to a desired high-resolution dental model.

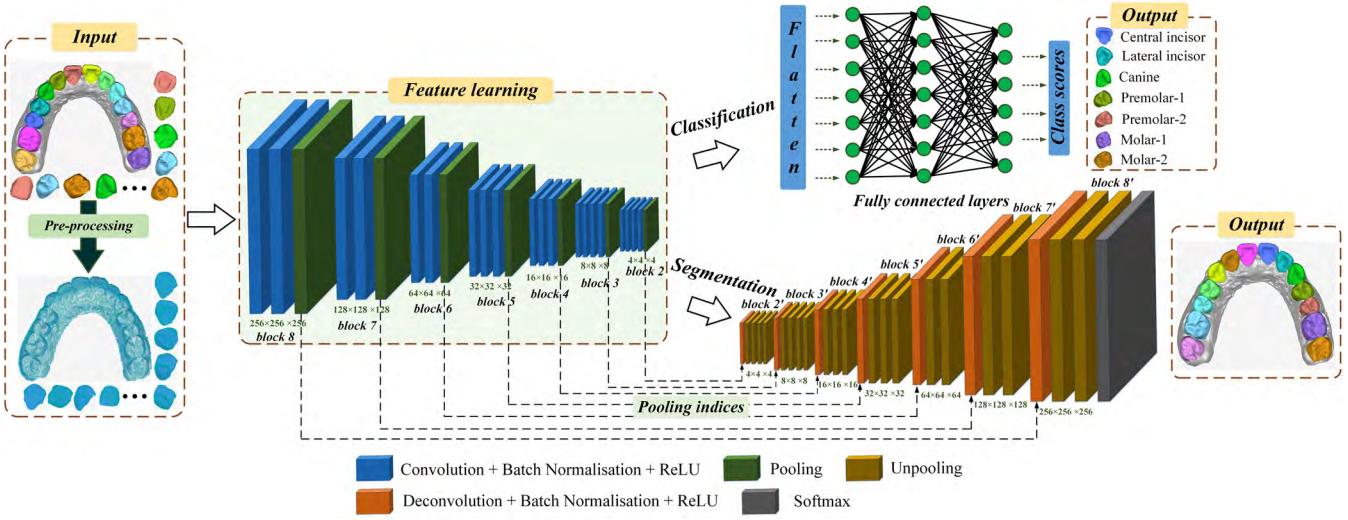


FIGURE 3. Overall architecture of the segmentation and classification network model.

representations, so numerous deep learning frameworks based on voxelization methods of point cloud models have emerged. As a typical deep CNN structure, O-CNN network has made a major breakthrough in 3D model classification, retrieval and semantic segmentation [23]. Therefore, by modifying the size of the filter, expanding the depth of the O-CNN network model, and optimizing the network structure and parameters, this paper effectively improves the segmentation and recognition ability of the network model for the dental models. The specific structure is shown in Fig. 3.

1) TOOTH CLASSIFICATION NETWORK BASED ON TWO-LEVEL HIERARCHICAL FEATURE LEARNING

Although some achievements have been made in point cloud classification based on deep convolution neural network, there are still quite a few problems that need to be further improved. Aiming at the problem that the sample misclassification is easy to be produced by the categories with higher similarity, a tooth classification method based on two-level hierarchical feature learning is proposed to further extract the different features between different tooth types with higher similarity to improve the accuracy of recognition, the network structure is shown in Fig. 1 and Fig. 3. To reflect the superiority of the hierarchical structure of the network, this network consists of alternating convolutional layers, pooling layers and fully connected layers. At the same time, we construct the basic unit block structure in the order of "convolution + batch normalization (BN) + rectified linear unit (ReLU) + pooling" to speed up the training process of separator network, and which is denoted by $block_d$ if the convolution is applied to the d-th depth octants. The network structure is defined as follows:

$$Input \rightarrow block_d \rightarrow block_{d-1} \rightarrow \cdots \rightarrow block_2 \rightarrow Class\ scores \rightarrow Output$$

The convolutional layer is the core of the whole network structure. According to its characteristics of local connection and weight sharing, it can effectively reduce the complexity of the network and the number of training parameters. Meanwhile, the convolution calculation is limited to the leaf octants of the octree, and the 3D convolution expression is defined as:

$$\mathbb{C}_{i,j,k}^{x,y,z} = f\left(\sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1}\omega_{p,q,r}^{(m)}\cdot T^{(m)}\left(O_{i,j,k}\right)+b_{i,j}\right) \quad (1)$$

where $f(\cdot)$ is the nonlinear function, $P_i \times Q_i \times R_i$ is kernel size, $\omega \in \Re^{P\times Q\times R}$ represents the weight vector of the kernel at the (p,q,r), $O_{i,j,k}$ represents a neighboring octant of octree, $T^{(m)}(\cdot)$ denotes the m-th channel of the feature vector, and $b_{i,j}$ denotes the bias term. To effectively alleviate the gradient dispersion phenomenon, we use the ReLU function to increase the sparsity of the network, and add BN processing, so that the network converges at a faster speed when the training gradient descend, and then solves the gradient disappearing problem.

As an important operation in the CNN, the pooling layer can reduce the dimensionality of the 3D geometric features and maintain the local translation/rotation invariance of the features to a certain extent. The max-pooling is the most common form, and which is defined in $\Omega(S_1, S_2, S_3)$ as:

$$\mathbb{S}_{x,y,z} = \beta_{x,y,z} \max_{(i,j,k)\in\Omega(S_1,S_2,S_3)} \left(\mu_{x\times l+i,y\times m+j,z\times n+k}\right) \quad (2)$$

where $\beta_{x,y,z}$ is the downsampling coefficient, μ is the value at$(x \times l + i, y \times m + j, z \times n + k)$, (l,m,n) denotes strides in the (x,y,z).

Different from local connection in convolution layer, the last layers of deep CNNs are usually fully connected, referred to as full-connected layers, which are used to connect classifier layers. To avoid overfitting, Dropout technology is usually used in the full-connected layers to reduce the correlation between features to improve the generalization ability of the model, and then the features transferred to the output layer are fed into classifier for classification and prediction. After the predicted labels of teeth obtained by using the Softmax classifier, the deviation between the predicted label and the ground truth is calculated by using the cross-entropy loss function.

$$H\left(p_i, q_i\right) = \sum_{i=1}^{n} p_i\left(x\right) ln\left(q_i\left(x\right)\right) \quad (3)$$

where $p_i(x)$ is the actual probability distribution; $q_i(x)$ is the predicted probability distribution. Finally, the parameters of the neural network are adjusted by the back propagation algorithm to extract the features of the tooth dataset and complete the construction of the tooth classification model.

2) THREE-LEVEL HIERARCHICAL TOOTH SEGMENTATION MODEL BASED ON DEEP CONVOLUTION FEATURES

In recent years, with the development of deep learning, semantic segmentation technology based on deep CNNs has attracted extensive attention. It mainly uses neural networks to complete segmentation and classification of point cloud models, so as to segment elements with semantic information. With the increase of semantic abstraction and non-linearity, the spatial shape information of model will be lost layer by layer along with feature extraction [25]. To alleviate the issue of gradient vanishing in deep networks, Noh *et al.* [26] proposed a semantic segmentation system based on deep encoder-decoder network. The encoder network performs convolution with a filter bank to extract the semantic features and produce a set of feature maps. The decoder network upsamples (via deconvolution) its input feature maps using the pooling indices from the corresponding encoder feature map. Deconvolution network [27] is an unsupervised learning model consisting of alternating unpooling layer (the reverse operation of pooling), deconvolution layer (the transpose of convolution kernel), and it is also a shape generator that produces object segmentation from the multidimensional features extracted from the convolution network [26]. Given that the l-th sample in the dental model dataset (N) is $x^{(l)}$, and the cost function of the m-th channel is obtained by deconvolution operation of the 3D feature z_k^i and the convolution kernel $f_{k,m}$ in the hidden layer:

$$\mathbb{C}_l^{'}\left(x^{(l)}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{m=1}^{M_{l-1}}\left\|\sum_{k=1}^{M_l} g_{k,m}^{i}\left(z_{k,l}^{i} \oplus f_{k,m}^{l}\right) - z_{c,l-1}^{i}\right\|_2^2 + \sum_{i=1}^{N}\sum_{k=1}^{M_l}\left|z_{k,l}^{i}\right|^p \quad (4)$$

where $g_{k,m}^{i}$ is a binary matrix, which is used to connect 3D features between neighboring layers, p denotes regularization parameter, $p = 1$.

Hierarchical network framework has the advantages of scalability and high efficiency, and which can maximize the utilization of training dataset to alleviate the imbalanced data problem [28]. By referring to the construction ideal of DeconvNet [29] and O-CNN structure [23], this paper constructs a hierarchical segmentation network architecture based on encoder-decoder structure, which is more suitable for dental model segmentation. The network structure is shown in Fig. 3. The segmentation network is defined by a structure similar to the classification model as follows:

$$\underbrace{Input \rightarrow block_d \rightarrow block_{d-1} \rightarrow}_{Encoder} \ldots \underbrace{\rightarrow block_2^{'} \rightarrow block_3^{'} \rightarrow block_d^{'} \rightarrow Output}_{Decoder}$$

Here $block_{\mathrm{d}}^{'}$ denotes the deconvolution operation, that is, "unpooling + deconvolution + BN + ReLU". Although the deconvolution operation and feature index compensation technology are used to solve the problem of network gradient disappearance in decoding network, the feature compensation continuity during the upsampling operation is lower, which leads to the appearance of sawtooth edges in gingival margin and the inter-teeth fusion region.

D. SEGMENTATION BOUNDARY REFINEMENT AND POST-PROCESSING

The target region segmentation algorithm based on deep CNNs can obtain the high-level semantic information of the point cloud model, which shows that the main part of the target objects can be segmented from the segmentation results. However, since the local detail features of the model are not considered, the prediction result is prone to produce rough

and inaccurate segmentation edges. The existing segmentation methods based on deep CNNs are inefficient for point cloud model boundary segmentation. To solve this problem, we exploit the dense-CRF technique [30] to refine the boundary of the gingival margin and the inter-teeth fusion region, and obtain the local detail features of the segment-ation region.

We denote a dental point cloud as $\{p_i\}_{i=1}^{N}$, and the normal of each point as $\{n_i\}_{i=1}^{N}$. The label set L consists of the label l_i of each point, and $l_i \in (0, \ldots k)$, $k \leq 4$. The corresponding Gibbs energy function is constructed as:

$$E(L) = \sum_i \varphi_i(l_i) + \sum_{i<j} \varphi_{ij}(l_i, l_j) \tag{5}$$

where $i, j \in [1, N]$, $\phi_i(l_i) = -log(p(x_i))$ is the unary energy, which is used to indicate the category of the corresponding point, where $p(x)$ is the label probability produced by the neural network. $\varphi_{ij}(l_i, l_j)$ is the pairwise potential energy to incorporate neighbor data information to refined the output:

$$\varphi_{ij}(L_i, L_j) = \mu(x_i, x_j)\left(\omega^1 \mathrm{k}^1(p_i - p_j) + \omega^2 \mathrm{k}^2(p_i - p_j)\,\mathrm{k}^3(n_i - n_j)\right) \tag{6}$$

where $\mu(x_i, x_j)$ is the label compatibility function, k^m is Gaussian function, ω^m is the linear combination weights corresponding to k^m. Finally, we do energy function optimization to smooth the segmentation boundary, followed by point cloud reconstruction and back-projection to get the final segmentation results.

III. EXPERIMENTS AND RESULTS

A. PARAMETER SETTINGS

In order to objectively record the surface morphological characteristics of the dental models while ensuring high graphics memory and computational efficiency, we exploit an 8-depth octree structure constructed to express the 3D details of the dental model. Seven basic unit blocks are constructed for dental feature extraction, where the kernel size for convolution and deconvolution is set to $3 \times 3 \times 3$ and applied with a stride of 1, the max-pooling size are set to $2 \times 2 \times 2$. Moreover, the ReLU function and BN are adopted following each convolution to improve the ability of network feature learning. Referring to the hierarchical tooth classification network structure in Fig. 1, the number of neurons in the fully connected layers of the Level-1 network are set to 128 and 4, and the number of neurons in the last block of the Level-2 network are respectively set to 128, 128 and 2. The high dimensional feature representation at the output of the fully connected layers are input to a Softmax classifier to predict the class probabilities of the tooth structure. The number of segmentation categories of the dental models is set to $2 \sim 4$ with reference to the dental segmentation framework. Moreover, skip-layer connection is added to the segmentation network to transfer the geometric details of 3D convolution feature to the corresponding deconvolution layer, and the weight of the deconvolution operation is tied with that of corresponding convolution layer.

TABLE 1. Numbers of three sample datasets by data augmentation.

Group		Training	Validation	Test
DatasetI	Segmentation	500	50	50
	Classification	1000	50	100
DatasetII	Segmentation	3000 (500×6)	300 (50×6)	50
	Classification	6000 (1000×6)	300 (50×6)	100
DatasetIII	Segmentation	5000 (500×10)	500 (50×10)	50
	Classification	10000 (1000×10)	500(50×10)	100

As introduced in the dataset pre-processing section, we have a total number of 600 dental models extracted from the scanned dentition plaster models, and the teeth segmented from each dental model are used as training sample datasets of the classification network. To augment the sample datasets, we further rotated the training and validation dataset from 0 to 360° in 60° and 36° steps along the z-axis respectively. We compared the segmentation and classification results using DatasetI without rotation, DatasetII with 60° rotation, and DatasetIII with 36° rotation. The number of three sample datasets is listed in Table 1.

In this paper, we adopt the stochastic gradient descent algorithm to train the 3D CNN network model, so that the learning rate can be adjusted adaptively with the gradient of the network, where we set the maximum number of iterations to 10000, the initial learning rate 0.01, and decreased by a factor of 10 after every 1000 iterations, the dropout ratio is 0.5, and the batch size is 12.

B. EXPERIMENTAL RESULTS AND EVALUATION

We conduct a number of experiments to verify the superiority of our approach, our network models are trained with Caffe framework [31] on an Intel (R) Platinum 8168 CPU (2.70 GHz) together with a GeForce GTX 1080Ti GPU.

1) CLASSIFICATION PERFORMANCE WITH DIFFERENT NETWORK LAYERS

Traditional binary classifiers usually adopt three evaluation indexes, accuracy (A), specificity (S) and recall (R) to validate the classification results. For the k-classification problem of imbalanced tooth datasets, the feasibility and effectiveness of the tooth classification network are verified by means of the macro-accuracy (MA), macro-specificity (MS) and macro-recall (MR).

$$MA = \frac{1}{k}\sum_{i=1}^{k} A_i = \frac{1}{k}\sum_{i=1}^{k} \frac{TP + TN}{TP + TN + FP + FN}$$

$$MS = \frac{1}{k}\sum_{i=1}^{k} S_i = \frac{1}{k}\sum_{i=1}^{k} \frac{TN}{TN + FP}$$

$$MR = \frac{1}{k}\sum_{i=1}^{k} R_i = \frac{1}{k}\sum_{i=1}^{k} \frac{TP}{TP + FN}$$

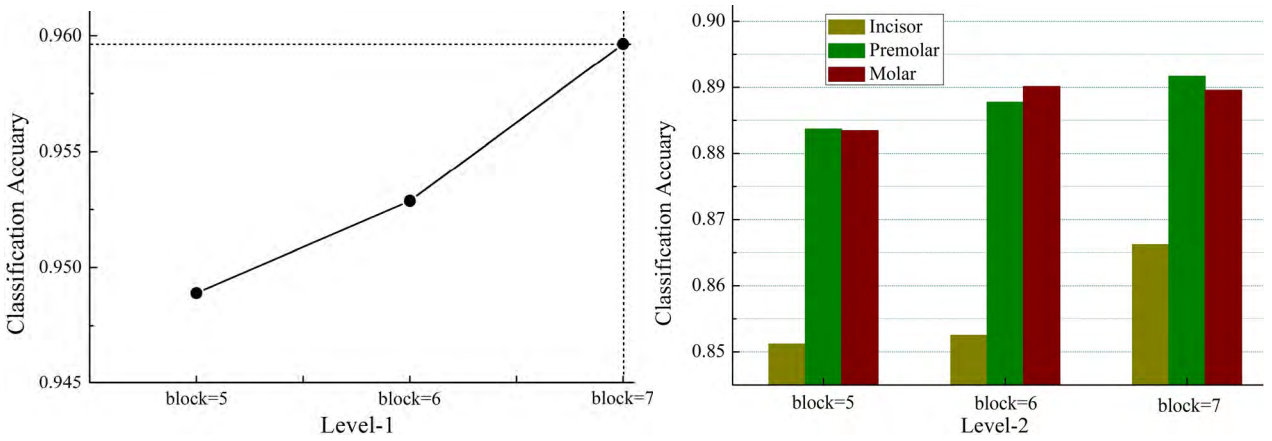


FIGURE 4. The accuracy comparison of classification results with different network layers.

where True Positive (TP) denotes the number of tooth types correctly identified as target type, False Positive (FP) denotes the number of tooth types incorrectly identified as target type, False Negative (FN) represents the number of tooth types that originally belong to this category but are misjudged and True Negative (TN) represents the number of tooth types that not belong to this class and are not misjudged.

We select seven tooth types as the classification objects, which are the central incisors, lateral incisors, canines, first premolars, second premolars, first molars and second molars. The third molars are excluded from the tooth classification tasks because of the small number of samples. In tooth classification tasks, similarities among different tooth types are different and the samples are usually mis-classified as highly similar categories. To distinguish highly similar tooth categories, we utilize gradient descent method to minimize the loss function to fine-tune the network parameters, and improve the classification performance of the classifier through repeated iteration training. Level-1 network (4-class) is used to classify tooth types to get the general features of all tooth categories and the tooth categories with high similarity. Level-2 network (2-class) is constructed by increasing the number of fully connected layers of the pre-classification network, and the feature expression and feature self-learning ability of the network are improved. We use the model parameters in the pre-classification network to initialize the Level-2 network, and then obtain the specific features for the tooth datasets with higher similarity. Table 2 summarizes the evaluation indexes of the test results with the different datasets and levels.

From Table 2, we can see that the measurement indicator values of the Level-1 network are higher than the Level-2 network. Because the feature complexities among different categories are different, and the learning ability of the network is different, so the classification accuracy of Level-1 network is slightly better than the Level-2 network. We also find that the network trained with the DatasetIII can get the highest values in all three measurement indexes, which indicates the accuracy of feature classification of deep learning is

TABLE 2. The evaluation indexes of the test results with different datasets.

		Level-1 (k=4) (%)	Level-2 (k=2) (%)					
			Incisor		Premolar		Molar	
			central	lateral	1st	2st	1st	2st
DatasetI	MA	91.44	84.38		85.42		85.83	
	MR	92.57	88.73		85.35		83.80	
	MS	90.08	80.02		85.49		87.86	
DatasetII	MA	95.43	87.25		88.13		88.96	
	MR	96.59	84.32		89.42		89.12	
	MS	93.88	90.50		86.83		88.79	
DatasetIII	MA	95.96	85.63		89.17		89.37	
	MR	96.38	89.58		89.86		89.75	
	MS	95.42	81.67		88.46		89.02	

related to the number of training samples, and the accuracy of our network can be further improved by using the virtual sample generation technology. It is also interesting to see that DatasetIII has good accuracy, which is above 85.63%. Finally, we can see from the macro-recall and macro-specificity that the imbalanced tooth datasets have little influence on the learning ability of our network, and it shows that the network has higher robustness.

To further test whether the hierarchical classification network is sensitive to the depth of the network, we set block to different sizes, keep other parameters unchanged and plot the classification accuracy of the DatasetIII dataset in Fig. 4.

In Fig. 4, we show the performance of classification network gradually improves with the increase of the layer numbers. It reveals that Level-1 network achieves the highest accuracy of 95.96% and Level-2 network obtains an average accuracy of 86.01% when basic unit block size is 7. However, due to the small number of incisors in the imbalanced tooth datasets, the classification accuracy is relatively low. Finally, the results indicate that the deep network structure can acquire more abstract features in high-level expression.

2) EFFECTIVENESS OF TOOTH SEGMENTATION

A dental model consists of teeth and gingiva, however, around 50% of point clouds in dental model belong to gingiva, the

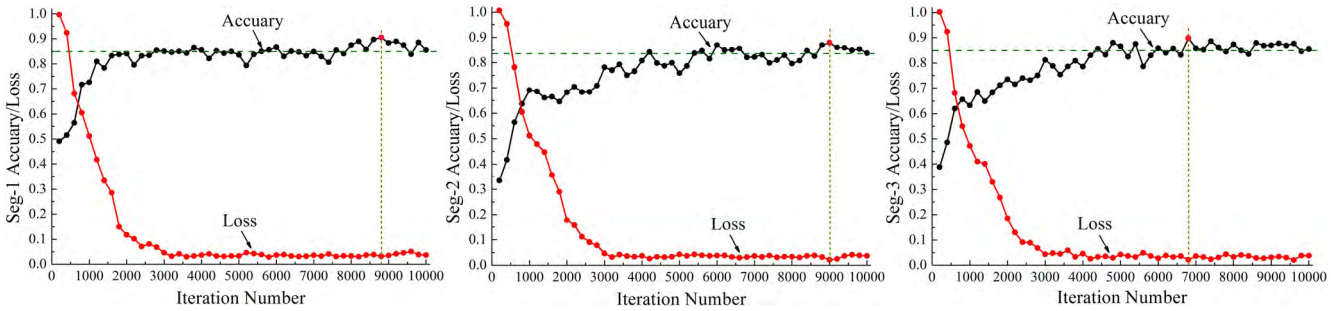


FIGURE 5. The segmentation accuracy and loss with different iteration number.

remaining part consists of different tooth types, yielding a severe imbalanced label distribution problem. To address this problem, we design a hierarchical tooth segmentation architecture. In order to verify the effectiveness and robustness of the proposed dental segmentation net-work, we adopt the patient's dental model with different degrees of tooth deformity to test the network performance. Considering the incomplete symmetry of dental model and avoiding increasing the burden of network training, we divide each tooth model into 4 parts for training (Seg-2 network), which are the left molars and premolars, the left incisors and canine, the right molars and premolars, the right incisors and canine, as shown in Fig. 1. In this paper, the trained segmentation network shown in Fig. 3 is used to segment the gingival margin and the inter-teeth fusion region. We divide the teeth and gingiva through the Seg-1 network (teeth-gingiva caffemodel), and then use the Seg-2 and Seg-3 networks (inter-teeth caffemodel) to complete the segmentation of the individual teeth. To quantitatively analyze the segmentation results of the dental model, we adopt accuracy to evaluate the performance of our network directly.

$$Accuracy = \sum_{p\in\mathbb{N}} a_p f_p\left(l_p\right) \Big/ \sum_{p\in\mathbb{N}} a_p$$

where a_p is the number of correctly predicted point p, l_p denotes the predicted label, $f_p(\cdot)$ denotes a parameter function of l_p, which equals to 1 if the prediction is correct, otherwise 0. The accuracy of a test is its ability to segment the target regions correctly, and which is the first criterion to evaluate our network to segment the gingival margin and the inter-teeth fusion region. Specifically, Fig. 5 has reported the training accuracy on the DatasetIII dataset with respect to every 200 iterations. We can see that the segmentation accuracy generally improves with the increase of the iteration number. Although small fluctuations exist with the iteration number due to the difference in the training samples, our segmentation network training gradually converges. It shows that the Seg-3 network achieves the highest accuracy of 89.81% when the iteration number is 6800. Therefore, once the network is trained, we used it to obtain the segmentation of new dental samples.

TABLE 3. The test accuracy of different methods.

	Wang et al. [23]	Qi et al. [32]	Ours
Seg-1(%)	82.27	80.91	91.52
Seg-2(%)	77.36	79.18	87.93
Seg-3(%)	78.59	76.85	89.81

To evaluate the effectiveness of our method and the reasoning ability of 3D CNN framework, we make a comparison with O-CNN(8) model in [23] and PointNet model in [32], which perform very well in arbitrary dental models. From Table 3, we can find that our method has significant advantage over other methods to segment all teeth and gingiva.

Fig. 6 shows the test results of tooth segmentation. It shows that CRF technique effectively reduces the inconsistencies between the boundary of the gingival margin and the inter-teeth fusion region, and the segmented boundary is closer to the ground truth.

IV. DISCUSSION

Computer-aided orthodontic technique is a combination of stomatology and computer science. Its clinical application has changed the traditional orthodontic method that mainly depends on doctor's experience and patient's subjective feeling to determine therapeutic scheme. As an important step of computer-aided orthodontic system, the main task of tooth segmentation is to accurately locate, identify and extract teeth from patient's digital dental model. The automatic segmentation of individual tooth is not a simple task since the teeth shapes are complex and teeth arrangement varies from person to person. However, the success application of deep learning in medical diagnosis have demonstrated its superiority in the reduction of labor costs and manual intervention. Additionally, the usage of the feature self-learning ability of CNN can improve the accuracy and efficiency of the network dramatically. And it is also of great significance for improving the intelligence level of the denture repair CAD system.

A. COMPARISON WITH STATE-OF-THE-ART METHODS

To data, several methods and network models have been proposed to solve the problem of automatic segmentation and classification of teeth. According to the 2D image obtained

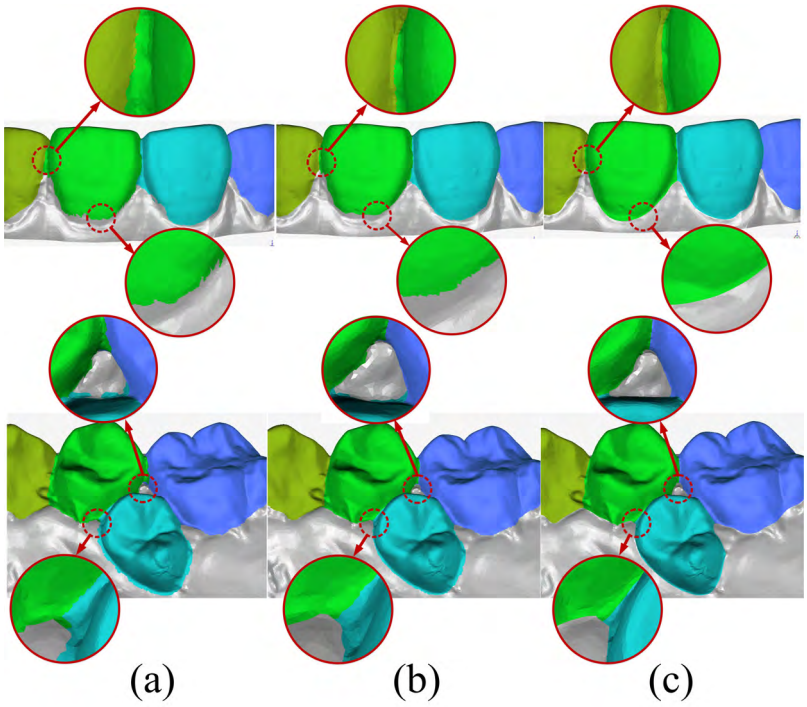


FIGURE 6. (a) The prediction result without CRF refinement. (b) The prediction result with CRF refinement. (c) Ground truth.

TABLE 4. Comparison between our method and some published papers on the tooth segmentation and classification.

	Method	Algorithm complexity	Function	Extraction method	Data	Accuracy	Model requirement
Wongwaen [3]	Image segmentation method based on Thresholding	High	Segmentation	Semi-automatic	2D	Low	Dental model with distinct characteristics
Kumar[8]	Shortest path search	Moderate	Segmentation	Semi-automatic	3D	Moderate	Arbitrary dental model
Miki[20]	Deep learning	Low	Classification	Automatic	2D	Moderate	--
Xu[21]	Deep learning	Moderate	Segmentation	Automatic	3D+2D	High	Simplified dental model
Our method	Deep learning	Low	Segmentation +Classification	Automatic	3D	High	Arbitrary dental model

by the 3D dental model projection, Wongwaen *et al.* [3] achieved the manual segmentation of teeth by combining occlusal plane and arch shape information. However, the image-based tooth segmentation methods often result in inaccurate segmentation results due to the loss of spatial information. Zou et al. [10] proposed a novel interactive tooth segmentation framework based on harmonic fields, which can successfully segment teeth from the complicated dental model with indistinctive tooth boundaries. However, due to the use of sufficient constraints for each target tooth, the method requires too many manual operations, which is tiring and time-consuming. Deep CNNs are widely used in computer-aided diagnosis because of their strong universality and robustness, but there are few reports on related research in the field of dentistry. Xu *et al.* [21] employed a boundary-aware tooth simplification method to preprocess dental models, and then extracted a 2D (20×30) geometric feature image for each mesh face of dental models and fed it into a 2D CNN classification network together with the face label. Finally, the fuzzy clustering algorithm was used to refine the boundary. However, it operates on hand-crafted geometric descriptors organized in a 2D feature matrix lacking spatially correlation structure for conventional convolution. When the boundary information of the tooth root region and the inter-teeth fusion region are corrupted by the simplification process, it will lead to an inaccurate prediction. And they think that the dental model is symmetrical, the classification results of the teeth are obtained by training the 8-label network, and then the left and right part is separated by geometric information. However, the patient's dental model with wisdom teeth or malformations teeth are actually asymmetrical. So the amount of wisdom teeth in the dental model will make the training step difficult. Miki *et al.* [20] used the pre-trained AlexNet network structure to classify tooth types on dental cone-beam CT images, but the results showed that the misclassifications were higher among the neighboring teeth, and the average classification accuracy using the augmented training data was 88.8%. Therefore,

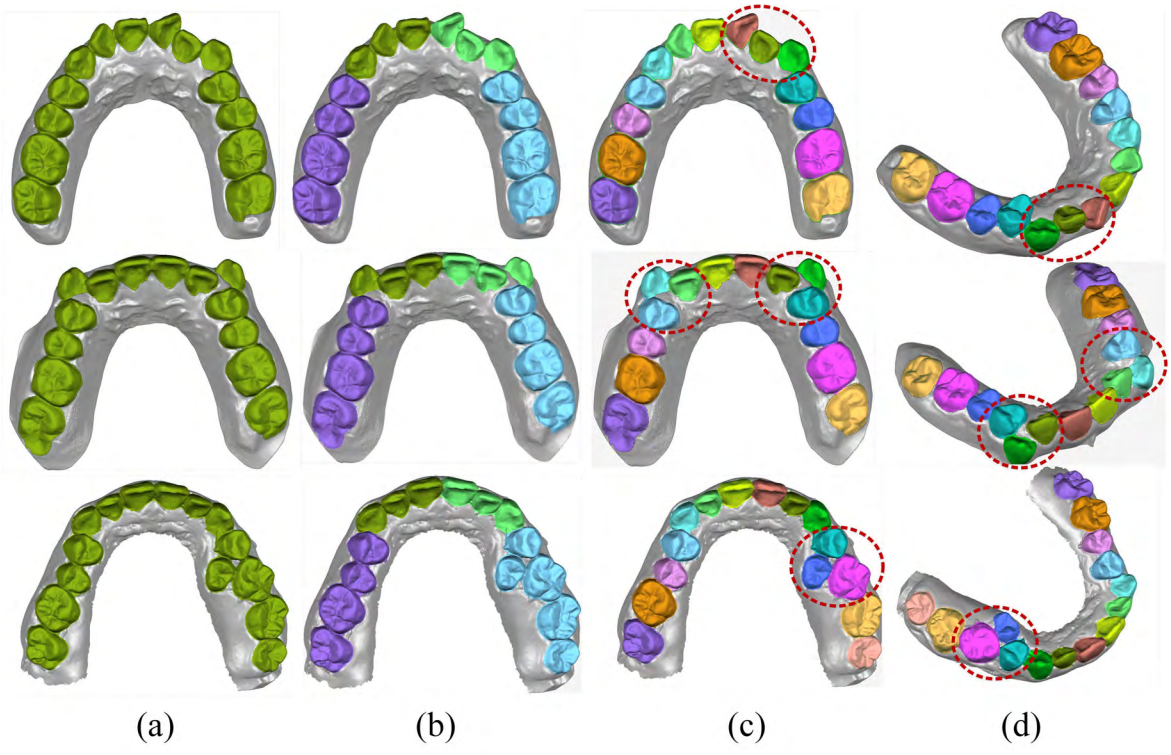


FIGURE 7. (a) Teeth-gingiva segmentation. (b) Inter-teeth segmentation. (c) Our results. (d) Another view of our results.

in view of the shortcomings of the existing tooth segmentation or classification methods, the two-level hierarchical classification method proposed in this paper can improve the classification performance and reduce the misclassification in highly similar categories. Meanwhile, the intelligent segmentation method based on CRF can be applied to various cases of dislocation or tooth missing, and the segmented boundary is closer to the ground truth. Since the simplified dental model cannot maintain the original maxillofacial morphology of the teeth, we use the original 3D dental model obtained from a laser scanner as the input to the 3D CNN for training. The segmentation result thus obtained is more realistic and more applicable to clinical practice. More importantly, we are the first attempt that makes use of hierarchical feature learning framework based on 3D CNN to segment and classify 3D teeth models. Through comparison with other state-of-the-art technologies, we can find that our method has advantages over the traditional geometry-based methods and image-based deep learning methods, and the comparison results are shown in Table 4.

B. FLEXIBILITY AND LIMITATION OF OUR METHOD

The layer of the deep learning network architecture is often very deep and requires to learn many parameters. We can effectively avoid over-fitting problem when training the network only on the condition that the training data is sufficient. In this study, despite the limited number of dental samples, the classification accuracy was relatively high (above 84.38%) even without data augmentation. By increasing the number of samples, the classification performance is further improved. From Table 2, we find that the performance of classification model gradually improves with the increase of the layer numbers, and the proposed classification method based on hierarchical feature learning can effectively improve the accuracy of teeth annotation. Especially, the hierarchical classification method can achieve better classification results than using the general feature extractor trained on all tooth categories. To quantify the quality of the segmentation results, Fig.7 shows the training results of the three representative dental models. We can see that the proposed method is robust to various complex malformations in patient teeth, and the segmentation boundary is reasonable and accurate. More-over, the segmentation method has important application value for virtual tooth arrangement in subsequent orthodontic treatment.

Although our method achieves considerable performance accuracy and efficiency, there are, nevertheless, several limitations to the current study. The first limitation is that the numbers of training data are too small to perform optimal deep learning. In order to overcome the limited sample size, we use only the teeth with less foreign matters on dental model surface as the training dataset, and randomly augmented 10 times. Although the number of virtual samples is relatively large, it is not enough to prove the generalization ability of the network. Another deficiency is the hierarchical segmentation network structure. The segmentation results of

Seg-1 structure will have a negative effect on the latter two layers. Taking a tooth-defect as an example, if the teeth-gingiva segmentation regards it as gingiva, it will no longer take part in the inter-teeth segmentation step. In the future work, we will expand our dental datasets to ensure that we can get more tooth data to train a more powerful network to segment or classify the tooth types, and reduce the inaccuracy caused by imbalanceed datasets.

V. CONCLUSION

The accurate tooth segmentation from 3D dental models is the most critical component in computer-assisted orthodontic treatment system for measuring the parameters of teeth and simulating the movement and rearrange of the teeth. This paper presents an automatic segmentation and classifycation method for 3D dental model via 3D CNN. To the best of our knowledge, it is the first attempt that makes use of hierarchical feature learning framework based on 3D CNN for automatically extracting high-dimensional featurees from 3D teeth models to segment and classify the tooth types. To reduce the misclassification in highly similar categories, the general features are extracted from all tooth categories by using Level-1 network and the specific features are extracted from highly similar tooth categories by using Level-2 network. Finally, an automatic boundary saliency detection of dental models based on conditional random field are used to solve the problem that the segmentation boundary is rough. Last but not least, the hierarchical segmentation network is robust to various complex malformations in patient teeth, and which has important application value for virtual tooth arrangement in subsequent orthodontic treatment.

REFERENCES

[1] T. Vos, A. A. Abajobir, and K. H. Abate, "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet*, vol. 390, no. 10100, pp. 1211–1259, 2017.

[2] Z. Li, X. Ning, and Z. Wang, "A fast segmentation method for STL teeth model," in *Proc. IEEE Int. Conf. Comp. Med. Eng.*, May 2007, pp. 163–166.

[3] N. Wongwaen and C. Sinthanayothin, "Computerized algorithm for 3D teeth segmentation," in *Proc. IEEE Int. Conf. Electron. Inf. Eng.*, Aug. 2010, pp. 277–280.

[4] 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.

[5] 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.*, 2010, pp. 521–530.

[6] A. Poonsri, N. Aimjirakul, T. Charoenpong, and C. Sukjamsri, "Teeth segmentation from dental X-ray image by template matching," in *Proc. IEEE 9th Biomed. Eng. Int. Conf.*, Dec. 2016, pp. 1–4.

[7] K. Wu, L. Chen, J. Li, and Y. Zhou, "Tooth segmentation on dental meshes using morphologic skeleton," *Comput. Graph.*, vol. 38, no. 1, pp. 199–211, 2014.

[8] T. Yuan, W. Liao, N. Dai, X. Cheng, and Q. Yu, "Single-tooth modeling for 3D dental model," *Int. J. Biomed. Imag.*, vol. 2010, no. 1, 2010, Art. no. 535329.

[9] Y. Kumar, R. Janardan, B. Larson, and J. Moon, "Improved segmentation of teeth in dental models," *Comput.-Aided Des. Appl.*, vol. 8, no. 2, pp. 211–224, 2011.

[10] 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.

[11] Z. Li and H. Wang, "Interactive tooth separation from dental model using segmentation field," *PLoS ONE*, vol. 11, no. 8, 2016, Art. no. e0161159.

[12] R. Fan, X. Jin, and C. C. L. Wang, "Multiregion segmentation based on compact shape prior," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 3, pp. 1047–1058, May 2015.

[13] Y. Zhou, J. Xu, Q. Liu, C. Li, Z. Liu, M. Wang, H. Zheng, and S. Wang, "A radiomics approach with CNN for shear-wave elastography breast tumor classification," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1935–1942, Sep. 2018.

[14] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Comput. Methods Program. Biomed.*, vol. 127, pp. 248–257, Apr. 2016.

[15] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, Feb. 2015.

[16] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, and M. Lillholm, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1322–1331, May 2016.

[17] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," *Inf. Process. Med. Imag.*, vol. 24, pp. 588–599, 2015.

[18] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[19] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *J. Dent*, vol. 77, pp. 106–111, Oct. 2018.

[20] Y. Miki, C. Muramatsu, and T. Hayashi, "Classification of teeth in cone-beam CT using deep convolutional neural network," *Comput. Biol. Med.*, vol. 80, pp. 24–29, Jan. 2017.

[21] 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.

[22] K. Guo, D. Zou, and X. Chen, "3D mesh labeling via deep convolutional neural networks," *ACM Trans. Graph.*, vol. 35, no. 1, pp. 1–12, 2015.

[23] P. S. Wang, Y. Liu, Y. X. Guo, C. Y. Sun, and X. Tong, "O-CNN: Octree-based convolutional neural networks for 3D shape analysis," *ACM Trans. Graph.*, vol. 36, no. 4, p. 72, 2017.

[24] C. M. Bishop, "Training with noise is equivalent to Tikhonov regularization," *Neural Comput.*, vol. 7, no. 1, pp. 108–116, 1995.

[25] X. J. Mao, C. Shen, and Y. B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*. [Online]. Available: https://arxiv.org/abs/1606.08921

[26] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.

[27] R. Mehrizi, X. Peng, X. Xu, S. Zhang, and K. Li, "A deep neural network-based method for estimation of 3D lifting motions," *J. Biomech.*, vol. 84, pp. 87–93, Feb. 2019.

[28] G.-H. Song, X.-G. Jin, G.-L. Chen, and Y. Nie, "Two-level hierarchical feature learning for image classification," *Frontiers Inf. Technol. Electron. Eng.*, vol. 17, no. 9, pp. 897–906, 2016.

[29] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern. Recognit.*, Jun. 2010, pp. 1–7.

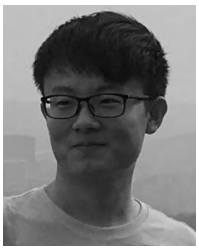
[30] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, pp. 109–117, 2011.

[31] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

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

SUKUN TIAN received the B.Sc. degree from Zaozhuang University, Zaozhuang, China, in 2011, and the M.Sc. degree from the University of South China, Hengyang, China, in 2013. He is currently pursuing the Ph.D. degree with the College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics. His research interests include biomedical engineering, deep learning, and computer-aided design.

NING DAI received the B.Sc. and Ph.D. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2000 and 2006, respectively, where he is currently an Associate Professor with the College of Mechanical and Electrical Engineering. His research interests include biomedical engineering, CAD/CAM, additive manufacturing, and deep learning.



BEI ZHANG received the B.Sc. degree from Yangzhou University, Yangzhou, China, in 2017. He is currently pursuing the M.Sc. degree with the College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics. His research interests include biomedical engineering and deep learning.

FULAI YUAN received the B.Sc. degree from the Harbin University of Science and Technology, Harbin, China, in 2017. He is currently pursuing the M.Sc. degree with the College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics. His research interests include biomedical engineering and deep learning.

QING YU received the Ph.D. degree from the School and Hospital of Stomatology, Peking University, Beijing, China, in 2009. He is currently a Professor and a Chief Physician with the Nanjing Stomatological Hospital, Medical School of Nanjing University. His research interests include prosthetic dentistry and digital dentistry.

XIAOSHENG CHENG received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2007, where he is currently a Professor and a Doctoral Tutor. His research interests include biomedical engineering and CAD/CAM.