Assignment Theory-Augmented Neural Network for Dental Arch Labeling

Tudor Dascalu(✉) and Bulat Ibragimov

Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
{tld,bulat}@di.ku.dk

Abstract. Identifying and detecting a set of objects that conform to a structured pattern, but may also have misaligned, missing, or duplicated elements is a difficult task. Dental structures serve as a real-world example of such objects, with high variability in their shape, alignment, and number across different individuals. This study introduces an assignment theory-based approach for recognizing objects based on their positional inter-dependencies. We developed a distance-based anatomical model of teeth consisting of pair-wise displacement vectors and relative positional scores. The dental model was transformed into a cost function for a bipartite graph using a convolutional neural network (CNN). The graph connected candidate tooth labels to the correct tooth labels. We reframed the problem of determining the optimal tooth labels for a set of candidate labels into the problem of assigning jobs to workers. This approach established a theoretical connection between our task and the field of assignment theory. To optimize the learning process for specific output requirements, we incorporated a loss term based on assignment theory into the objective function. We used the Hungarian method to assign greater importance to the costs returned on the optimal assignment path. The database used in this study consisted of 1200 dental meshes, which included separate upper and lower jaw meshes, collected from 600 patients. The testing set was generated by an indirect segmentation pipeline based on the 3D U-net architecture. To evaluate the ability of the proposed approach to handle anatomical anomalies, we introduced artificial tooth swaps, missing and double teeth. The identification accuracies of the candidate labels were 0.887 for the upper jaw and 0.888 for the lower jaw. The optimal labels predicted by our method improved the identification accuracies to 0.991 for the upper jaw and 0.992 for the lower jaw.

Keywords: Multi-object recognition · Assignment theory · Dental instance classification

1 Introduction

Object detection and identification are key steps in many biomedical imaging applications [13]. Digital imaging is essential in dentistry, providing practitioners

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 295–304, 2023.
https://doi.org/10.1007/978-3-031-43898-1_29

with detailed internal and surface-level information for the accurate diagnosis and effective treatment planning of endodontic and orthodontic procedures [18]. Intra-oral scans (IOS) are a specific type of digital dental imagery that produce 3D impressions of the dental arches, commonly referred to as dental casts [12]. Surface level visualizations can be utilized for the automatic design and manufacturing of aligners and dental appliances [18].

A precursor to fully automated workflows consists of accurate detection and recognition of dental structures [4,9,11]. The difficulty of the tasks stems from inherent anatomical variability among individuals, as well as the presence of confounding factors such as treatment-related artifacts and noise [4,8]. To date, a limited number of studies have been conducted to experiment with algorithms performing instance segmentation in dental casts [3,15,17,19–21]. Tian et al. [17] proposed a multi-level instance segmentation framework based on CNNs, investigating the impact of incorporating a broad classification stage that classified structures into incisors, canines, premolars, and molars. Xu et al. [19] adopted a similar broad-to-narrow classification strategy, developing two CNNs for labeling mesh faces based on handcrafted features. Sun et al. [16] applied the FeaStNet graph CNN algorithm for dental vertex labeling. Cui et al. [3] developed a tooth detection pipeline that included tooth centroid localization followed by instance segmentation applied on cropped sub-point clouds surrounding the centroids.

The identification of dental instances is a complex task due to the presence of anatomical variations such as crowding, missing teeth, and "shark teeth" (or double teeth) [7]. Xu et al. [19] employed PCA analysis to correct mislabeled pairs of teeth caused by missing or decayed teeth. Sun et al. [16] analyzed crown shapes and the convexity of the border region to address ambiguous labeling of neighboring dental instances. However, previous studies have not effectively addressed the issue of tooth labeling in the presence of dental abnormalities such as misaligned and double teeth.

The labeling of dental casts presents a non-trivial challenge of identifying objects of similar shapes that are geometrically connected and may have duplicated or missing elements. This study introduces an assignment theory-based approach for recognizing objects based on their positional inter-dependencies. We developed a distance-based dental model of jaw anatomy. The model was transformed into a cost function for a bipartite graph using a convolutional neural network. To compute the optimal labeling path in the graph, we introduced a novel loss term based on assignment theory into the objective function. The assignment theory-based framework was tested on a large database of dental casts and achieved almost perfect labeling of the teeth.

2 Method

2.1 Generation of Candidate Labels

The database utilized in the present study comprised meshes that represented dental casts, with the lower and upper jaws being depicted as separate entities.

Each mesh vertex was associated with a label following the World Dental Federation (FDI) tooth numbering system. A large proportion of the samples in the dataset was associated with individuals who had healthy dentition. A subset of patients presented dental conditions, including misaligned, missing, and duplicated teeth. Furthermore, the dataset consisted of patients with both permanent and temporary dentition.

The dental cast labeling task was divided into two stages: the detection of candidate teeth (1), which involved identifying vertices forming instances of teeth, and the assignment of labels to the candidate teeth, with geometric and anatomical considerations (2). The process of detecting candidate teeth consisted of indirect instance segmentation. The dental casts were converted to binary volumetric images with a voxel resolution of 1mm; voxels containing vertices were assigned a value of 1, while those without vertices were assigned a value of 0 [5]. The binary images were then segmented using two separate 3D U-net models, each specifically trained for either the upper or lower dental cast types. The models were trained to segment 17 different structures. When applied to a new volumetric dental cast, the models generated 17 probability maps: one for each of the 16 tooth types, corresponding to the full set of normal adult human teeth present in each jaw, plus an additional one for non-dental structures like gums. The outputs were converted to vertex labels over the input dental cast, by finding spatial correspondences between voxels and vertices.

2.2 Dental Anatomical Model

The difficulty of segmenting dental structures stems from the high inter-personal shape and position variability, artifacts (e.g. fillings, implants, braces), embedded, and missing teeth [1,4,6]. These challenges combined with the tendency for neighboring instances to have similar shapes affect the performance and accuracy of the U-Net models. As a result, the output produced by the segmentors may contain missed or incorrectly assigned labels.

To address these challenges and build an accurate instance segmentation pipeline, we first generated a dental anatomical model that provided a framework for understanding the expected positions of the teeth within the jaw. The dental model relied on the relative distances between teeth, instead of their actual positions, for robustness against translation and rotation transformations. The initial step in modeling the jaws was calculating the centroids of the dental instances by averaging the coordinates of their vertices. The spatial relationship between two teeth centroids, $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$, was described by the displacement vector, $\boldsymbol{d} = \boldsymbol{c_1} - \boldsymbol{c_2}$. To evaluate the relative position of two instances of types t_1 and t_2, we calculated the means (μ_x, μ_y, μ_z) and standard deviations $(\sigma_x, \sigma_y, \sigma_z)$ for each dimension (x, y, z) of the displacement vectors corresponding to instances of types t_1 and t_2 in patients assigned to the training set. The displacement rating r for the two instances was computed as the average of the univariate Gaussian probability density function evaluated at each dimension (d_x, d_y, d_z) of the displacement vector $\boldsymbol{d}$:

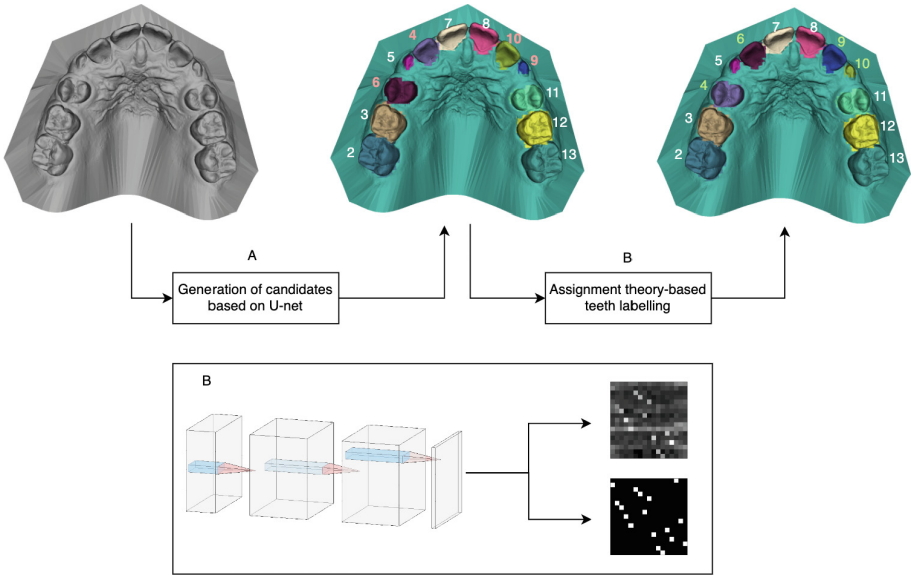

Fig. 1. The assignment theory-based object recognition pipeline. The initial phase (A) depicts the process of producing candidate instances and the results, with two pairs of swapped teeth (6–4, 10–9). The subsequent phase (B) presents the cost map and optimal assignment path produced by DentAssignNet.

$$r = \frac{1}{3} \sum_{i \in \{d_x, d_y, d_z\}} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(d_i - \mu_i)^2}{2\sigma_i^2}} \tag{1}$$

The dental model of a patient consisted of a 3D matrix $\mathbf{A}$ of shape $m \times m \times 4$, where m corresponded to the total number of tooth types. For each pair of dental instances i and j, the elements a_{ij0}, a_{ij1}, a_{ij2} correspond to the x, y, and z coordinates of the displacement vector, respectively, while a_{ij3} represents the relative position score. The presence of anatomical anomalies, such as missing teeth and double teeth, did not impact the information embedded in the dental model. If tooth i was absent, the values for the displacement vectors and position ratings in both row i and column i of the dental model $\mathbf{A}$ were set to zero.

2.3 The Assignment Problem

The next step was evaluating and correcting the candidate tooth labels generated by the U-net models using the dental anatomical model. We reformulated the task as finding the optimal label assignment. Assignment theory aims to solve the similar task of assigning m jobs to m workers in a way that minimizes expenses or maximizes productivity [14]. For the purpose of dental cast labeling, the number of jobs m is defined as t teeth types present in a typical healthy individual and d

double teeth observed in patients with the "shark teeth" condition, $m = t+d$. To ensure that the number of jobs matched the number of workers, k "dummy" teeth without specific locations were added to the n teeth candidates generated by the U-net. The dental model was transformed into a cost matrix $\mathbf{B}$ of dimensions $m \times m$, where each element b_{ij} represented the cost associated with assigning label j to candidate instance i. We solved the assignment problem using the Hungarian method [10]. The optimal assignment solution $\mathcal{C}^*$ for n candidate teeth was not affected by the presence of k "dummy" teeth, provided that all elements b_{ij} in the cost matrix $\mathbf{B}$ where either i or j corresponds to a "dummy" instances were assigned the maximum cost value of q.

Proposition 1. *Let the set of candidate teeth be $\mathcal{C}$ and the number of possible candidate teeth $m = t + s$, such that $|\mathcal{C}| < m$. The optimal label assignment to the candidate instances is $\mathcal{C}^* = f^*(\mathcal{C})$, where $f^* \in \mathcal{F}$ is the optimal assignment function. Let us assume that there exists only one optimal assignment function f^*. The sum of costs associated with assigning candidate teeth to their optimal labels according to f^* is less than the sum of costs associated with any other assignment function $p \in \mathcal{F}\backslash\{f^*\}$, $\sum_{x\in\mathcal{C}} \mathbf{B}_{x,f^*(x)} < \sum_{x\in\mathcal{C}} \mathbf{B}_{x,p(x)}$. The inclusion of r dummy teeth, with $r = m - |\mathcal{C}|$, and maximum assignment cost q cannot alter the optimal assignment $\mathcal{C}^*$.*

Proof. Let's assume that the addition of one dummy object θ with maximum assignment cost q changes the optimal assignment to $g \in \mathcal{F}\backslash\{f^*\}$ on the set $\mathcal{C} \cup \{\theta\}$. The assignment cost of the new candidate set $\mathcal{C} \cup \{\theta\}$ is

$$\sum_{x\in\mathcal{C}\cup\{\theta\}} \mathbf{B}_{x,g(x)} = \sum_{x\in\mathcal{C}} \mathbf{B}_{x,g(x)} + q \tag{2}$$

considering that $\mathbf{B}_{\theta,g(\theta)} = q$. The definition of the optimal label assignment function f^* states that its cumulative assignment cost is smaller than the cumulative assignment cost of any $g \in \mathcal{F}\backslash\{f^*\}$ on the set $\mathcal{C} \cup \{\theta\}$, which indicates that

$$\sum_{x\in\mathcal{C}} \mathbf{B}_{x,g(x)} + q > \sum_{x\in\mathcal{C}} \mathbf{B}_{x,f^*(x)} + q = \sum_{x\in\mathcal{C}\cup\{\theta\}} \mathbf{B}_{x,f^*(x)} \tag{3}$$

This contradicts the assumption that g is the optimal assignment for $\mathcal{C} \cup \{\theta\}$. This proof can be generalized to the case where multiple dummy teeth are added to the candidate set.

In other words, Proposition 1 states that adding "dummy" instances to the candidate teeth sets does not affect the optimal assignment of the non-dummy objects. The dummy teeth played a dual role in our analysis. On one hand, they could be used to account for the presence of double teeth in patients with the "shark teeth" condition. On the other hand, they could be assigned to missing teeth in patients with missing dentition.

2.4 DentAssignNet

The optimal assignment solution f^* ensured that each candidate tooth would be assigned a unique label. We integrated the assignment solver into a convolutional neural network for labeling candidate teeth, entitled DentAssignNet (Fig. 1).

The input to the convolutional neural network consisted of a matrix $\mathbf{A}$ of shape $m \times m \times 4$, which represented a dental model as introduced in Sect. 2.2. For each pair of dental instances i and j, the element a_{ij} included the coordinates of the displacement vector and the relative position score. The architecture of the network was formed of 3 convolutional blocks. Each convolutional block in the model consisted of the following components: a convolutional layer, a rectified linear unit (ReLU) activation function, a max pooling layer, and batch normalization. The convolutional and pooling operations were applied exclusively along the rows of the input matrices because the neighboring elements in each row were positionally dependent. The output of the convolutional neural network was a cost matrix $\mathbf{B}$ of shape $m \times m$, connecting candidate instances to potential labels. The assignment solver transformed the matrix $\mathbf{B}$ into the optimal label assignment $\mathcal{C}^*$. The loss function utilized during the training phase was a weighted sum of two binary cross entropy losses: between the convolutional layer's output $\hat{\mathbf{Y}}$ and the ground truth $\mathbf{Y}$, and between $\hat{\mathbf{Y}}$ multiplied by the optimal assignment solution $\hat{\mathbf{Y}}'$ and the ground truth $\mathbf{Y}$.

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_{BCE(\hat{\mathbf{Y}},\mathbf{Y})} + \lambda\mathcal{L}_{BCE(\hat{\mathbf{Y}}',\mathbf{Y})} \tag{4}$$

The direct application of the assignment solver on a dental model $\mathbf{A}$ with dimensions $m \times m \times 4$ was not possible, as the solver required the weights associated with the edges of the bipartite graph connecting candidate instances to labels. By integrating the optimal assignment solution in the loss function, DentAssignNet enhanced the signal corresponding to the most informative convolutional layer output cells in the task of labeling candidate teeth.

3 Experiment and Results

3.1 Database

The database employed in this study was introduced as part of the 3D Teeth Scan Segmentation and Labeling Challenge held at MICCAI 2022 [2]. It featured 1200 dental casts, depicting lower and upper jaws separately. The dental structures were acquired using intra-oral scanners (IOS) and modeled as meshes. The average number of vertices per mesh was 117377. The cohort consisted of 600 patients, with an equal distribution of male and female individuals. Approximately 70% of the patients were under 16 years of age, while around 27% were between 16 and 59 years of age, and the remaining 3% were over 60 years old.

3.2 Experiment Design

To obtain the candidate tooth labels, we employed the U-net architecture on the volumetric equivalent of the dental mesh. Considering that the majority of the samples in the database featured individuals with healthy dentition, a series of transformations were applied to the U-net results to emulate the analysis of cases with abnormal dentition.

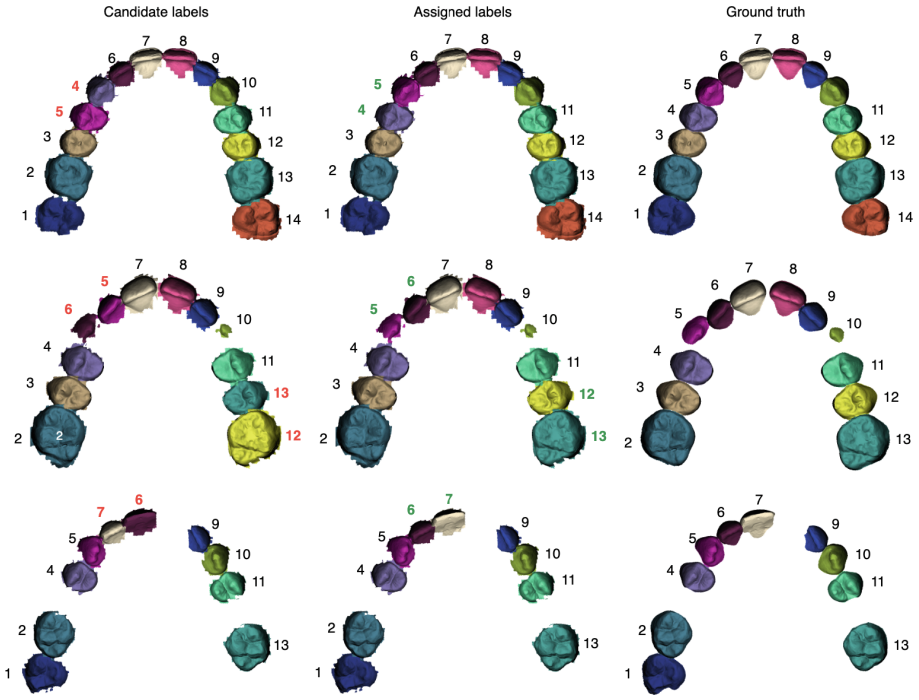

Fig. 2. The axial view of the input dental candidates assigned by the U-net (column 1), the corresponding output produced by DentAssignNet (column 2), and the ground truth (column 3). Each row represents a new patient. Each tooth type is represented by a unique combination of color and label. The color red indicates instances that were mislabeled, while the color green indicates instances that were previously misclassified but were correctly labeled by the framework. (Color figure online)

To simulate the simultaneous occurrence of both permanent and temporary dental instances of the same type, we introduced artificial centroids with labels copied from existing teeth. They were placed at random displacements, ranging from d_{min} to d_{max} millimeters, from their corresponding true teeth, with an angulation that agreed with the shape of the patient's dental arch. The protocol for constructing the database utilized by our model involved the following steps. Firstly, we calculated the centroids of each tooth using the vertex labels

Table 1. The performance of the U-net (row 1), ablated candidate teeth (row 2), and DentAssignNet (row 3), for the lower and upper jaws. Column 1 denotes the identification accuracy. Columns 3 and 4 include the number of mislabeled teeth followed by the total number of teeth in parenthesis.

Model	Accuracy		Swapped teeth errors		Double teeth errors	
	Lower jaw	Upper jaw	Lower jaw	Upper jaw	Lower jaw	Upper jaw
U-net	0.972	0.971	22 (1306)	25 (1276)	0 (2)	0 (2)
U-net ablated	0.888	0.887	1368 (12239)	1345 (11944)	40 (503)	34 (485)
DentAssignNet	**0.992**	**0.991**	**40 (12239)**	**43 (11944)**	**11 (503)**	**6 (485)**

generated by the U-Net model. Subsequently, we augmented the U-net results by duplicating, removing, and swapping teeth. The duplication procedure was performed with a probability of $p_d = 0.5$, with the duplicate positioned at a random distance between $d_{min} = 5$ millimeters and $d_{max} = 15$ millimeters from the original. The teeth removal was executed with a probability of $p_e = 0.5$, with a maximum of $e = 4$ teeth being removed. Lastly, the swapping transformation was applied with a probability of $p_w = 0.5$, involving a maximum of $w = 2$ tooth pairs being swapped. Given that neighboring dental instances tend to share more similarities than those that are further apart, we restricted the label-swapping process to instances that were located within two positions of each other.

3.3 Results

Each sample in the database underwent 10 augmentations, resulting in 9000 training samples, 1000 validation samples, and 2000 testing samples for both jaws. The total number of possible tooth labels was set to $m = 17$, consisting of $t = 16$ distinct tooth types and $d = 1$ double teeth. The training process involved 100 epochs, and the models were optimized using the RMSprop algorithm with a learning rate of 10^{-4} and a weight decay of 10^{-8}. The weighting coefficient in the assignment-based loss was $\lambda = 0.8$.

The metrics reported in this section correspond to the detection and identification of teeth instances. The U-net models achieved detection accuracies of 0.989 and 0.99 for the lower and upper jaws, respectively. The metrics calculated in the ablation study take into account only the dental instances that were successfully detected by the candidate teeth proposing framework. Table 1 presents identification rates for the candidate dental instances (prior to and following ablation) and the performance of DentAssignNet. Our framework achieved identification accuracies of 0.992 and 0.991 for the lower and upper jaws, respectively. There was a significant improvement in performance compared to the U-net results (0.972 and 0.971) and the artificially ablated input teeth (0.888 and 0.887). Figure 2 depicts the ability of DentAssignNet to handle patients with healthy dentition (row 1), erupting teeth (row 2), and missing teeth (row 3). For comparison purposes, we refer to the results of the 3D Teeth Scan Segmentation and Labeling Challenge at MICCAI 2022 [2]. The challenge evaluated the

algorithms based on teeth detection, labeling, and segmentation metrics, on a private dataset that only the challenge organizers could access. Hoyeon Lim et al. adapted the Point Group method with a Point Transformer backbone and achieved a labeling accuracy of 0.910. Mathieu Leclercq et al. used a modified 2D Residual U-Net and achieved a labeling accuracy of 0.922. Shaojie Zhuang et al. utilized PointNet++ with cast patch segmentation and achieved a labeling accuracy of 0.924. Our identification accuracies of 0.992 and 0.991 for the lower and upper jaw, respectively, compare favorably to the results from the challenge. However, it must be noted that this is not a direct comparison as the results were achieved on different segments of the dental cast challenge database. Additionally, the metrics used in the challenge were specifically designed to accommodate the dual task of detection and identification, calculating labeling accuracy relative to all dental instances, including those that were not detected.

4 Conclusion

We proposed a novel framework utilizing principles of assignment theory for the recognition of objects within structured, multi-object environments with missing or duplicate instances. The multi-step pipeline consisted of detecting and assigning candidate labels to the objects using U-net (1), modeling the environment considering the positional inter-dependencies of the objects (2), and finding the optimal label assignment using DentAssignNet (3). Our model was able to effectively recover most teeth misclassifications, resulting in identification accuracies of 0.992 and 0.991 for the lower and upper jaws, respectively.

Acknowledgments. This work was supported by Data+ grant DIKU, the University of Copenhagen, and the Novo Nordisk Foundation grant NNF20OC0062056.

References

1. Amer, Y.Y., Aqel, M.J.: An efficient segmentation algorithm for panoramic dental images. Procedia Comput. Sci. **65**, 718–725 (2015). https://doi.org/10.1016/j.procs.2015.09.016
2. Ben-Hamadou, A., et al.: Teeth3DS: a benchmark for teeth segmentation and labeling from intra-oral 3D scans (2022). https://doi.org/10.48550/arXiv.2210.06094
3. Cui, Z., et al.: TSegNet: an efficient and accurate tooth segmentation network on 3D dental model. Med. Image Anal. **69**, 101949 (2021). https://doi.org/10.1016/j.media.2020.101949
4. Dascalu, T.L., Kuznetsov, A., Ibragimov, B.: Benefits of auxiliary information in deep learning-based teeth segmentation. In: Medical Imaging 2022: Image Processing. vol. 12032, pp. 805–813. SPIE (2022). https://doi.org/10.1117/12.2610765
5. Dawson-Haggerty et al.: trimesh, https://trimsh.org/
6. Ehsani Rad, A., Rahim, M., Rehman, A., Altameem, A., Saba, T.: Evaluation of current dental radiographs segmentation approaches in computer-aided applications. IETE Tech. Rev. **30**, 210–222 (2013). https://doi.org/10.4103/0256-4602.113498

7. Ip, O., Azodo, C.C.: "Shark Teeth" Like Appearance among Paediatric Dental
8. Jin, C., et al.: Object recognition in medical images via anatomy-guided deep learning. Med. Image Anal. **81**, 102527 (2022). https://doi.org/10.1016/j.media.2022.102527
9. Kondo, T., Ong, S., Foong, K.: Tooth segmentation of dental study models using range images. IEEE Trans. Med. Imaging **23**, 350–62 (2004). https://doi.org/10.1109/TMI.2004.824235
10. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Res. Logistics Quart. **2**(1–2), 83–97 (1955). https://doi.org/10.1002/nav.3800020109
11. Lian, C., et al.: MeshSNet: Deep Multi-scale Mesh Feature Learning for End-to-End Tooth Labeling on 3D Dental Surfaces. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 837–845. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_93
12. Mangano, F., Gandolfi, A., Luongo, G., Logozzo, S.: Intraoral scanners in dentistry: a review of the current literature. BMC Oral Health **17**, 149 (2017). https://doi.org/10.1186/s12903-017-0442-x
13. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annu. Rev. Biomed. Eng. **2**(1), 315–337 (2000). https://doi.org/10.1146/annurev.bioeng.2.1.315
14. Singh, S.: A comparative analysis of assignment problem. IOSR J. Eng. **02**(08), 01–15 (2012). https://doi.org/10.9790/3021-02810115
15. Sun, D., et al.: Automatic Tooth Segmentation and Dense Correspondence of 3D Dental Model. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 703–712. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_68
16. Sun, D., Pei, Y., Song, G., Guo, Y., Ma, G., Xu, T., Zha, H.: Tooth segmentation and labeling from digital dental casts. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 669–673 (2020). https://doi.org/10.1109/ISBI45749.2020.9098397, iSSN: 1945-8452
17. Tian, S., Dai, N., Zhang, B., Yuan, F., Yu, Q., Cheng, X.: automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. IEEE Access **7**, 84817–84828 (2019). https://doi.org/10.1109/ACCESS.2019.2924262
18. Vandenberghe, B.: The crucial role of imaging in digital dentistry. Dental Mater. **36**(5), 581–591 (2020). https://doi.org/10.1016/j.dental.2020.03.001
19. Xu, X., Liu, C., Zheng, Y.: 3D tooth segmentation and labeling using deep convolutional neural networks. IEEE Trans. Vis.Comput. Graph. **25**(7), 2336–2348 (2019). https://doi.org/10.1109/TVCG.2018.2839685
20. Zhao, Y., et al.: Two-stream graph convolutional network for intra-oral scanner image segmentation. IEEE Trans. Med. Imaging **41**(4), 826–835 (2022). https://doi.org/10.1109/TMI.2021.3124217
21. Zhao, Y., et al.: 3D Dental model segmentation with graph attentional convolution network. Pattern Recogn. Lett. **152**, 79–85 (2021). https://doi.org/10.1016/j.patrec.2021.09.005