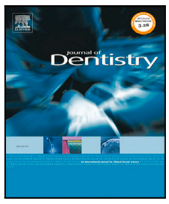
Journal of Dentistry 159 (2025) 105767

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent



Morphological effects of input data quantity in AI-powered dental crown design

Alexander Broll [a], Markus Goldhacker [b], Sebastian Hahnel [a], Martin Rosentritt [a,*]

[a] Department of Prosthetic Dentistry, University Hospital Regensburg, Regensburg, Germany
[b] Faculty of Mechanical Engineering, OTH Regensburg, Regensburg, Germany

ARTICLE INFO

Keywords:
Tooth reconstruction
Dental prosthesis design
Deep learning
Input data quantity
Digital dentistry

ABSTRACT

Objectives: This retrospective in vitro study evaluated the impact of input data quantity on the morphology of dental crowns generated by AI-based software. The hypothesis suggests that increased input data quantity improves the quality of generated occlusal surfaces.

Methods: A dataset comprising n=30 patients (11 males, 19 females; age: 22–31 years) was analyzed. Input data was categorized into full dentition (full), quadrant data (quad), and adjacent teeth (adj). AI-based software (Dentbird Crown, Imageworks Inc.) generated crowns for a single lower first molar (36/46). Metrics were proposed to assess the morphology and occlusal relationships of the crowns, with the original tooth as reference. Statistics: Friedman Chi-Square tests, Wilcoxon signed rank tests, Kendall correlation and Fligner–Killeen tests ($\alpha = 0.05$).

Results: Full and quad groups provided consistent reconstruction quality with no significant differences in morphology and occlusal relationships. The adj group showed significant ($p < 0.05$) morphological deviations and higher reconstruction failure rates compared to the full and quad groups. Correlations (median: 0.19; min–max range: 0.01–0.54) indicate that the proposed metrics capture distinct morphological and functional crown aspects.

Conclusion: The software reliably reconstructed crowns with at least quadrant-level input data. Performance declined with reduced input. Full-jaw scans did not enhance accuracy compared to quadrant data.

Clinical Significance: Increased input data quantity can improve the accuracy of AI-based restorations. As a result, prosthodontists benefit from predictable, accurate restoration proposals that reduce the need for digital chairside adjustments as well as manual modifications after fabrication. This streamlines clinical workflows and enhances the quality of restorations. Quadrant-level data has proven sufficient to generate high-quality reconstructions. Further input data did not significantly improve the accuracy of the reconstructions. The proposed metrics enable quantitative assessments of morphological and functional restoration quality, supporting reliable AI-driven workflows.

1. Introduction

Advances in digital technologies have transformed dental restoration processes, with computer aided design and computer aided manufacturing systems now widely integrated into clinical workflows [1,2]. These systems offer reduced labor intensity for both dental technicians and clinicians, along with good reconstruction quality. Fully integrated chairside systems have streamlined workflows for smaller restorations by enabling an in-house, end-to-end process. This includes intraoral scanning of the patient's dentition, designing the fixed dental prosthesis, and fabricating the final restoration. To this point, manual adjustments are often required to refine the restoration proposal either during the design phase or in the clinical setting. This is particularly important for ensuring proper masticatory function, encompassing both dynamic contacts during mandibular movements and static contacts in terminal occlusion. Inaccurate adjustments can lead to occlusal discomfort, temporomandibular disorders, or even restoration failure [3,4].

While most of the current reconstruction approaches are based on conventional digital methods, emerging deep learning (DL)-based software offers a novel way to generate accurate dental restorations [5]. These systems leverage deep neural networks to infer complex relationships between the existing dentition and the desired restoration based on the knowledge acquired from extensive training on large datasets. Due to the data-driven nature of DL, the quality and accuracy of the

* Correspondence to: Department of Prosthetic Dentistry, University Hospital Regensburg, Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany.
E-mail address: martin.rosentritt@ukr.de (M. Rosentritt).

https://doi.org/10.1016/j.jdent.2025.105767
Received 10 February 2025; Received in revised form 10 April 2025; Accepted 16 April 2025
Available online 7 May 2025

generated restorations is expected to improve with richer and more comprehensive input data [6].

The aim of this study is to evaluate the morphology and occlusal relationships of the occlusal surfaces generated by a commercial AI-based dental restoration software (Dentbird Crown, Imageworks Inc.) [7] with respect to input data quantity. While several studies have compared different digital dental restoration systems [8–11], there is a lack of research addressing the influence of input data quantity on the accuracy of AI-based systems. By investigating the relationship between input data quantity and restoration quality, this study seeks to provide insights for optimizing data acquisition during clinical practice to improve the restoration outcomes. Enhanced restoration accuracy can reduce the need for manual adjustments and ensure better clinical outcomes with the potential to lower costs, streamline treatment processes, and provide patients with more reliable and consistent restorations.

To quantitatively evaluate the quality of the generated occlusal surfaces, a set of metrics is required to comprehensively assess the overall morphology and functional occlusal relationships. While there are widely adopted metrics for evaluating the similarity between two point clouds (PCDs) [12,13] (i.e. overall morphology), there is a lack of common metrics specifically designed to evaluate occlusal surfaces with respect to the patient's contact points [5].

We propose a set of occlusion-specific metrics designed to evaluate the occlusal surfaces to quantify the quality of the contact situation between a target tooth and its antagonist teeth. We aim to translate qualitative and visually assessed information, typically evaluated by a dentist during clinical examinations, into a quantitative framework. While manual measurements are commonly used in dental research, the customized metrics are developed to provide an automated, differentiable and quantitative evaluation framework. This mitigates the subjectivity and variability of manual measurements and streamlines an objective, reproducible evaluation process that is applicable to large amounts of data without user intervention. The metrics can then be integrated for the evaluation of existing methods or as part of customized loss functions during the training and validation of new methods.

We hypothesize that increasing the input data quantity significantly improves the morphological accuracy and occlusal functionality of AI-designed dental crowns.

2. Methods

2.1. Study design

This retrospective in vitro study was conducted using intraoral scan data from $n = 30$ patients, including 11 males and 19 females, aged between 22 and 31 years. The patients were selected based on the total quantity ($n_{\text{teeth,patient}} \geq 28$) and condition of their teeth. The data was acquired using an intraoral scanner (CEREC PrimeScan, Dentsply Sirona) and exported as mesh files with the highest resolution setting. Each patient was scanned once to provide a single full jaw scan file that served as the baseline for the subsequent digital derivation of the different data quantities. There were no further patient interventions. The intraoral scanning technique used was according to the manufacturer's instructions with occlusal, buccal, lingual and proximal scans in a sequence. Soft tissue and unnecessary gingival data were cropped during the acquisition process to avoid unwanted artifacts. Ethical approval for data usage was obtained from the Ethics Committee of the University Regensburg (25-4027-104).

The reconstruction target was the lower first molar (36/46), selected based on the condition of the tooth, ensuring it shows no signs of decay or structural damage and is free of restorations. The original tooth served as the ground truth (GT) for the evaluation.

The data quantity was varied among three categories: full dentition (full), the respective quadrant (quad) or the adjacent teeth (adj) of the reconstruction target (Fig. 1). The antagonist teeth were included for each category to the extents of the lower jaw data quantities.

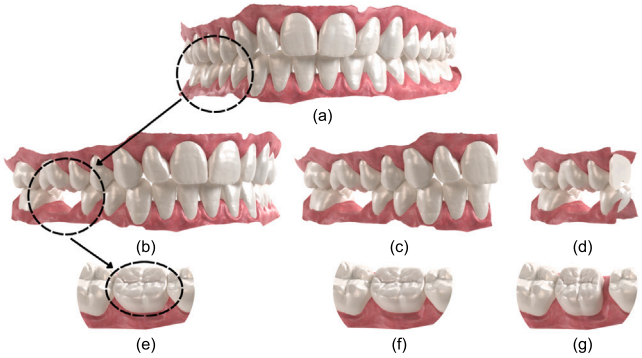


Fig. 1. Examples of different data quantities (b–d) including their reconstructions (e–g) for an exemplary patient and a reconstruction target 46. (a) uncorrupted full jaw data (b) full: full jaw data (c) quad: quadrant data (d) adj: adjacent teeth data (e) full reconstruction (f) quad reconstruction (g) adj reconstruction.

For each patient, the full dentition was imported into the software as a virtual crown case and the reconstruction target tooth was removed. The resulting corrupted data $P_{\text{full,corr}}$ was then exported as the baseline for the derivation of the additional data quantities $P_{\text{quad,corr}}$ and $P_{\text{adj,corr}}$. To generate these datasets, the full jaw data was manually cropped to isolate the respective quadrant or adjacent teeth.

In the following, a PCD is formally defined as a set of points $P = \left\{ \boldsymbol{p}_i \mid \boldsymbol{p}_i \in \mathbb{R}^3 \ , \ i \in \{1, \dots |P|\} \right\}$ with the cardinality $|P|$ as the number of points $\boldsymbol{p}_i$. The GT for the reconstruction target P_{gt} can be extracted from the complete full jaw data P_{full} by calculating the minimal distances between the corrupted and full points

$$D_{\text{full}} = \left\{ \min_j \|\boldsymbol{p}_{\text{full,corr},j} - \boldsymbol{p}_{\text{full},i}\|_2 \ \ \forall \ i \in \{1, \dots |P_{\text{full}}|\} \right\}.$$

The GT is then defined as the points of the full jaw data that lie within an empirical threshold $\varepsilon = 1 \times 10^{-6}$ mm from the corrupted data

$$P_{\text{gt}} = \left\{ \boldsymbol{p}_{\text{full},i} \mid d_{\text{full},i} < \varepsilon \right\}.$$

The insertion direction of a crown in a practical application can alter the morphology of the occlusal surface during the design phase. Therefore, a fixed direction was established for each patient based on the orientation $\boldsymbol{R}_{B,gt} = \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{r}_3 \end{bmatrix}$ of the bounding box of the GT with its three axes $\boldsymbol{r}_i \in \mathbb{R}^3$ (Fig. 2). The negative insertion direction is defined as the vector $\boldsymbol{r}_i$ that minimizes the angle between itself and the subset of normal vectors $\boldsymbol{N}_{\text{prep,gt}}$ of the preparation surface. This is formally defined as

$$-\boldsymbol{r}_{\text{ins}} = \arg\max_{\boldsymbol{r}_i} \|\boldsymbol{N}_{\text{prep,gt}}^T \boldsymbol{r}_i\|_2,$$

where $\|\cdot\|_2$ represents the L_2 norm of a vector $\boldsymbol{p} \in \mathbb{R}^n$

$$\|\boldsymbol{p}\|_2 = \sqrt{\sum_{i=1}^{n} p_i^2}. \tag{1}$$

The insertion direction was subsequently used to generate the reconstructions for the different data quantities. The occlusal surfaces of the reconstructions were extracted using boolean operations with the respective corrupted input data for all groups full, quad, adj and downsampled to a fixed number of points $|P_{\text{rec}}| = |P_{\text{gt}}| = 8192$ using furthest point sampling [14].

2.1.1. Sample size calculation

The sample size was calculated based on a pilot study with $n = 5$ patients for all data quantities. $\mathcal{L}_{\text{chamfer,L2}}$ was used as the reference metric. The preliminary evaluation yielded mean values of 0.416 mm (full), 0.392 mm (quad), 0.673 mm (adj) and a mean within group standard deviation of 0.228 mm. The a priori sample size calculation for a repeated measures ANOVA with a target power of 0.8, a significance level of 0.05 and an effect size of 0.56 determined $n = 33$ per group. The final sample size was set to $n = 30$ for an estimated power of 0.77.

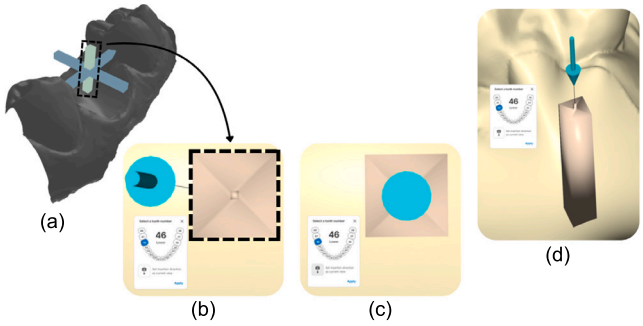


Fig. 2. Determination of the insertion direction for a crown reconstruction target based on the orientation of the bounding box of the GT. (a) The three axes of the bounding box. The insertion direction is defined as the negative vector of the axis that minimizes the angle between itself and the normal vectors of the preparation surface. (b) The axis with the smallest angle to the normal vectors is loaded into the software as an additional cuboid. (c) The insertion direction is aligned with the axis of the cuboid. (d) The cuboid is removed, and the insertion direction is set for the reconstruction.

2.2. Metrics

2.2.1. Morphological metrics

The morphological metrics assess the overall morphology and spatial characteristics of the generated 3D object in comparison to the GT. These metrics are widely recognized in the domain of 3D shape reconstruction and are well-established for quantifying the similarity between two PCDs.

Chamfer Distance (CD): The chamfer distance (CD) [13] is a commonly used metric for evaluating the similarity between two PCDs [12]. It is defined by the sums of the distances between each point in one PCD to its nearest neighbor in the other PCD. Formally, the two-sided CD between two point sets P and Q with L_2 distances according to (1) is calculated as

$$\mathrm{CD}_{\mathrm{L2}}(P,Q) = \frac{1}{|P|}\sum_{i=1}^{|P|}\min_j \|\boldsymbol{p}_i - \boldsymbol{q}_j\|_2 + \frac{1}{|Q|}\sum_{j=1}^{|Q|}\min_i \|\boldsymbol{q}_j - \boldsymbol{p}_i\|_2.$$

The CD loss is defined as the mean of the two-sided CD with

$$\mathcal{L}_{\mathrm{chamfer,L2}}(P,Q) = \frac{\mathrm{CD}_{\mathrm{L2}}(P,Q)}{2}. \tag{2}$$

Similar to the Root Mean Square Error, the domain agnostic CD evaluates all aspects of the 3D object, including the overall shape, spatial distribution, and surface details. A low CD indicates that a digital dental reconstruction accurately replicates the intricate details of the original tooth surface.

Intersection over Union (IoU): In the dental reconstruction context, the intersection over union (IoU) evaluates the pose and spatial characteristics of the generated tooth in a broader context, ensuring that it fits correctly within the patient's existing dentition [15]. The spatial alignment of the generated crown with the adjacent and antagonist teeth is crucial for the overall functionality and fit of the restoration. In contrast to the CD, which measures local discrepancies between corresponding points, the IoU evaluates the crown as a whole. The IoU between two PCDs P and Q is formally defined as

$$\mathrm{IoU}(P,Q) = \frac{|P \cap Q|}{|P \cup Q|},$$

where $P \cap Q$ is the intersection of the two PCDs, representing the points common to both sets. The union $P \cup Q$ of the two PCDs includes all points belonging to either P, Q, or both. In this context, the IoU is computed based on the bounding boxes $B_P \in \mathbb{R}^{8\times3}$ and $B_Q \in \mathbb{R}^{8\times3}$, each defined by its eight corner points $\boldsymbol{b}_i$. These bounding boxes represent the smallest oriented cuboids that enclose all points of the respective 3D objects. Their orientation is determined by the principal axes of the object, obtained through principal component analysis of the point set [16].

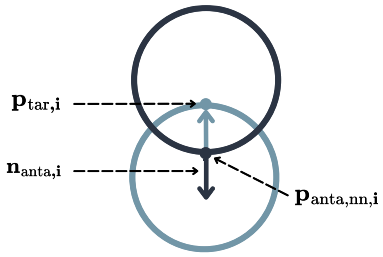


Fig. 3. Determination of the occluded points P_{occl} between the target tooth (tar) and the antagonist teeth (anta) based on the normal vectors of the nearest neighbors $\boldsymbol{n}_{\mathrm{anta},i}$. A point is considered occluded if the dot product of the normal vector and the vector between the target $\boldsymbol{p}_{\mathrm{tar},i}$ and the nearest neighbor from the antagonist tooth $\boldsymbol{p}_{\mathrm{anta,nn},i}$ is negative. I.e. the normal vector points in the opposite direction.

The IoU is calculated as the intersection of the two bounding boxes divided by their union. This approach is commonly used to calculate the IoU for PCDs, as it is computationally more efficient than directly calculating the IoU based on the individual points [17]. To align with the direction of the other metrics (lower is better), the IoU is complemented, and the loss is given by

$$\mathcal{L}_{\mathrm{cIoU}} = 1 - \mathrm{IoU}(B_P, B_Q).$$

2.3. Occlusion specific metrics

2.3.1. Definition of the contact points

All metrics are calculated based on the occluded points P_{occl} between the reconstruction P_{rec} or the GT P_{gt} and the antagonist teeth P_{anta}. For simplicity, the reconstruction or GT data is referred to as the target P_{tar} in the following.

Occluded points: To identify the occluded points, the nearest neighbor of each point in P_{tar} is calculated in P_{anta}. This process yields the points $P_{\mathrm{anta,nn}}$ with its distances $D_{\mathrm{tar,anta}}$ to P_{tar}. The normal vectors of the nearest neighbors are utilized to determine whether the points are occluded (Fig. 3).

The occlusion distances D_{occl} are defined as the set

$$D_{\mathrm{occl}} = \left\{ \begin{cases} -d_{\mathrm{tar,anta},i}\,, & \text{if } (\boldsymbol{p}_{\mathrm{tar},i} - \boldsymbol{p}_{\mathrm{anta,nn},i})^{\mathrm{T}}\boldsymbol{n}_{\mathrm{anta},i} < 0 \\ d_{\mathrm{tar,anta},i}\,, & \text{otherwise} \end{cases} \quad \forall\, i \in \{1,\dots,|P_{\mathrm{tar}}|\} \right\}.$$

The threshold $d_{\mathrm{thresh}} = 0\,\mathrm{mm}$ defines whether a point is considered occluded. Based on this threshold, the set of occluded points is given by

$$P_{\mathrm{occl}} = \left\{ \boldsymbol{p}_{\mathrm{tar},i} \mid d_{\mathrm{occl},i} < d_{\mathrm{thresh}} \right\}.$$

This formulation identifies points from the target tooth where the distance to the antagonist teeth is below the threshold.

Contact points: These points can now be used to calculate the size, position, and number of contact points on the occlusal surface, i.e. the clusters of occluded points. The mean shift algorithm [18] is used to identify the clusters of occluded points with their respective cluster centers (Fig. 4). The non-parametric clustering algorithm is used to find the modes of a set of points P. The algorithm is based on the concept of kernel density estimation and iteratively shifts the points towards the modes of the data distribution.

Given the set of points P, the mean shift algorithm aims to find the updated positions $\boldsymbol{m}_i$ for each point $\boldsymbol{p}_i$ in the set. For each point $\boldsymbol{p}_i$, the pairwise distance d_{ij} to all other points $\boldsymbol{p}_j$ is calculated as

$$d_{ij} = \|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2.$$

The distances are used to determine the weights w_{ij} between the points using the Gaussian kernel $\mathrm{K}(d_{ij}, h_{\mathrm{G}})$. It is given by

$$w_{ij} = \mathrm{K}(d_{ij}, h_{\mathrm{G}}) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_{\mathrm{G}}}\right)^2\right),$$

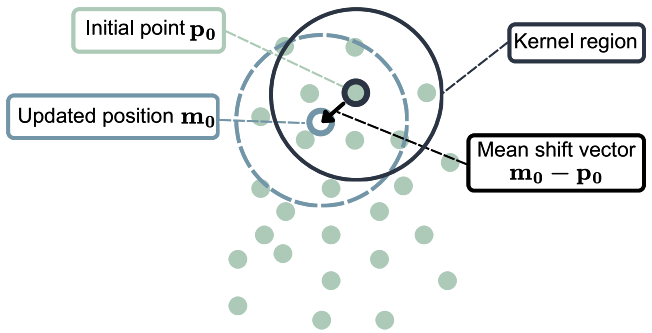


Fig. 4. Visualization of the mean shift algorithm used for clustering the occluded points of a reconstruction target. The initial point p_0 is adjusted to its new position m_0 by calculating the weighted average of distances to other points within the kernel region. The update is performed by adding the mean shift vector, $m_0 - p_0$, to the initial point.

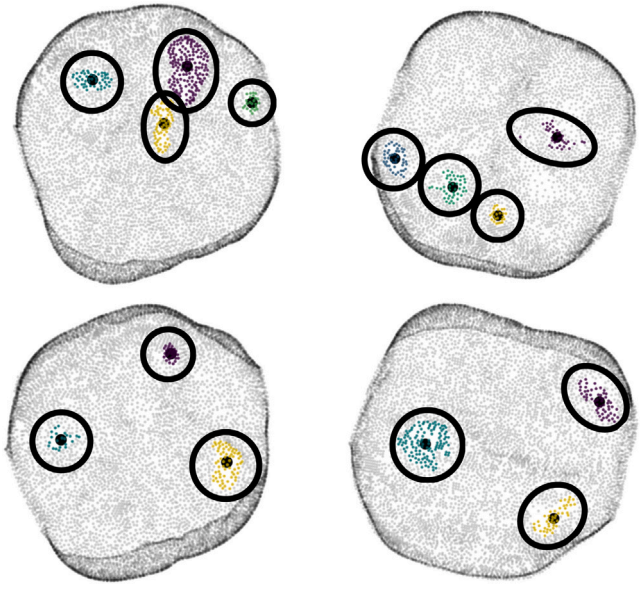


Fig. 5. The clustering results for four exemplary teeth with each cluster represented by a distinct color. The cluster centers, denoted as m_i, are indicated by black dots, and the associated occluded points are enclosed within black ellipses.

where h_G is the bandwidth of the Gaussian kernel and $\exp(\cdot)$ is the natural exponential function. The new position m_i for each point p_i is calculated as

$$m_i = \frac{\sum_{j=1}^{|P|} w_{ij} p_j}{\sum_{j=1}^{|P|} w_{ij}},$$

where m_i is the updated position of p_i. The algorithm iterates until convergence is reached with

$$\max_i \|m_i - p_i\|_2 < \varepsilon_{\text{conv}}.$$

That is, the maximum distance between the updated position and the original position of each point is below a threshold $\varepsilon_{\text{conv}}$. The threshold $\varepsilon_{\text{conv}} = 0.45\,\text{mm}$ is a predefined empirical tolerance level that was determined for the given data by comparison of the clustering results to the visual representation of the occluded points in a 3D visualization environment (Fig. 5). If the condition is satisfied, the algorithm stops and each point p_i is assigned to the nearest cluster center $m_i \in M$. The assignment is based on minimizing the distance $\|p_i - m_j\|_2$ for all cluster centers.

2.3.2. Definition of the metrics

Based on the occluded points P_{occl} and the cluster centers M of the occluded points (contact points), the occlusion specific metrics are defined as follows:

Penetration loss $\mathcal{L}_{\text{pen}}$: The penetration is the sum over all penetration distances

$$D_{\text{pen}} = \{d_{\text{occl},i} \mid d_{\text{occl},i} < d_{\text{thresh}}\}$$

between the generated occlusal surface and the antagonist teeth. The penetration loss is used to compare the mean penetration depth of the GT and the generated occlusal surfaces. It is given by

$$\mathcal{L}_{\text{pen}} = \sqrt{\left(\frac{1}{|P_{\text{pen,gt}}|}\sum_{i=1}^{|P_{\text{pen,gt}}|} d_{\text{pen,gt},i} - \frac{1}{|P_{\text{pen,rec}}|}\sum_{i=1}^{|P_{\text{pen,rec}}|} d_{\text{pen,rec},i}\right)^2}.$$

Contact point distance loss $\mathcal{L}_{\text{cp,dist}}$: The contact point distance loss is used to compare the overall spatial spread of the contact points for the GT and the generated occlusal surfaces. The mean contact point distance is defined as the mean of all off-diagonal elements of the pairwise distance matrix $\boldsymbol{D}$ of the contact points $m_i, m_j \,\forall\, i, j \in \{1, \ldots |M|\}$ according to

$$d_{ij} = \|m_i - m_j\|_2.$$

This denotes the element in the i–th row and j–th column of the matrix $\boldsymbol{D}$. Fig. 6 (a) illustrates the intra contact point distances for two sets of contact points.

With

$$\tilde{\mu}_{\boldsymbol{D}} = \frac{1}{|M|(|M|-1)}\sum_{\substack{i,j=1 \\ i\neq j}}^{|M|} d_{ij}$$

the loss can be calculated as

$$\mathcal{L}_{\text{cp,dist}} = \sqrt{\left(\tilde{\mu}_{\boldsymbol{D},\text{gt}} - \tilde{\mu}_{\boldsymbol{D},\text{rec}}\right)^2}.$$

Contact point position loss $\mathcal{L}_{\text{cp,pos}}$: The contact point position loss measures the positional deviation of the contact points to compare the overall shape of the contact point pattern. Following (2), it is defined as the CD between the contact points of the GT and the generated occlusal surfaces with

$$\mathcal{L}_{\text{cp,pos}} = \mathcal{L}_{\text{chamfer,L2}}(M_{\text{gt}}, M_{\text{rec}}).$$

Fig. 6 (b) shows the inter contact point position distances for two sets of contact points.

Contact point number loss $\mathcal{L}_{\text{cp,num}}$: The contact point number loss determines the deviation in the number of contact points between the GT and the generated occlusal surfaces. It is defined as

$$\mathcal{L}_{\text{cp,num}} = \sqrt{\left(|M_{\text{gt}}| - |M_{\text{rec}}|\right)^2}$$

with $|M_{\text{gt}}|$ and $|M_{\text{rec}}|$ as the number of the contact points for the GT and the generated occlusal surfaces, respectively.

Table 1 presents an overview of the evaluation metrics, outlining their technical characteristics and the specific aspects of the occlusal surface they assess.

2.3.3. Evaluation

The generated reconstructions were evaluated based on the previously proposed metrics. Since each patient contributed data across all three data quantity groups and all metrics failed the normality assumption according to the Shapiro–Wilk test ($p < 0.05$), the statistical analysis employed Friedman Chi-Square tests to detect significant differences among the groups. One-sided Wilcoxon signed rank tests were then used for the post hoc inter-group comparisons. The comparison groups were defined as G_1: quad–full, G_2: adj–quad, and G_3: adj–full. The alternative hypothesis for the test was defined as H_1: median$(g_i - g_j) > 0$ where median$(\cdot)$ refers to the median of the respective group

Table 1
Summary of the evaluation metrics, detailing their function and purpose. The function describes the technical characteristics of each metric, while the purpose explains the specific aspect of the generated occlusal surface being assessed. ↓↑ show the direction of the metric, indicating whether higher or lower values are preferable.

Metric	Function	Purpose
Morphological metrics		
Chamfer Distance $\mathcal{L}_{\text{chamfer,L2}}$ ↓	Nearest average distances between points in the reconstruction and GT.	General morphological accuracy.
Complemented Intersection over Union $\mathcal{L}_{\text{cIoU}}$ ↓	Overlap of two objects with respect to their union.	Spatial alignment and fit in the tooth gap.
Occlusion specific metrics		
Penetration loss $\mathcal{L}_{\text{pen}}$ ↓	Deviation of mean penetration depths for all occluded points.	Contact strength.
Contact point distance loss $\mathcal{L}_{\text{cp,dist}}$ ↓	Deviation of mean contact point distances.	Spatial spread of the contact points.
Contact point position loss $\mathcal{L}_{\text{cp,pos}}$ ↓	Nearest average distances between the contact point centers of the GT and the reconstruction.	Shape of the contact point pattern.
Contact point number loss $\mathcal{L}_{\text{cp,num}}$ ↓	Deviation in the number of contact points.	Number of contact points.

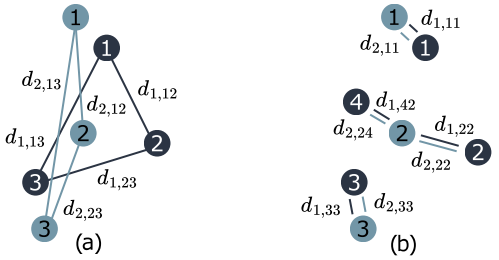


Fig. 6. Visualization of $\mathcal{L}_{\text{cp,dist}}$ and $\mathcal{L}_{\text{cp,pos}}$: (a) $\mathcal{L}_{\text{cp,dist}}$ measures the error in the mean pairwise distances between the contact points in each set. (b) $\mathcal{L}_{\text{cp,pos}}$ quantifies the discrepancy using the CD, defined as the distance between each contact point and its nearest neighbor in the opposing set.

$g_i \in G$. A post hoc power analysis was conducted using simulation with $n_{\text{sim}} = 1000$ bootstrap iterations to estimate the power of the statistically significant tests. There was no p-correction for multiple comparisons due to the hypothesis-driven nature of the study with its limited number of comparison groups.

The results of all metrics were normalized to $[0, 1]$ according to the maximum and minimum values of the respective metric across all groups. This ensures comparability between the metrics and allows for a comprehensive evaluation of the quality of the generated occlusal surfaces across different data quantities. All metrics were implemented such that lower values indicate better performance.

To account for cases where the process failed to generate a morphologically sound reconstruction for specific data quantities or patients, resulting in the inability to calculate any metrics, all metrics were assigned a value of 1.1. Similarly, for cases where the process failed to establish occlusion between the target and antagonist teeth, the occlusion-specific metrics were set to 1.1. This approach ensures that failed reconstructions and failed occlusion establishment are included in the evaluation, emphasizing potential limitations of the software for certain data quantities.

The set of metrics was evaluated using Kendall correlation to examine inter-metric relationships and identify potential redundancies or dependencies. Kendall correlation was chosen due to the non-normality of the metric distributions and its robustness against outliers.

The insertion direction dependent deviation of the generated occlusal surfaces was evaluated with the same metrics. For each insertion direction type, $n = 15$ new cases were created in the software and crowns were generated for a fixed single patient. This allowed for a direct comparison of the variance in the occlusal surfaces based on the insertion direction. The variances of the metrics were compared using Fligner–Killeen tests with the alternative hypothesis H_1: $\text{var}(g_{\text{auto}}) \neq \text{var}(g_{\text{fix}})$.

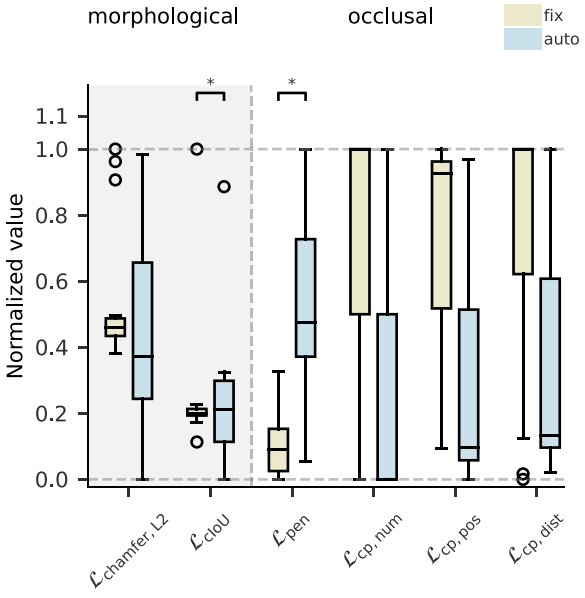


Fig. 7. Normalized results of the evaluation for $n = 15$ fixed and auto insertion direction cases. The labels on the significance bars refer to statistically significant deviations in variances of the metrics. The significance level is set to $\alpha = 0.05$.

3. Results

Fig. 7 displays the normalized metrics for fixed and auto insertion directions. Table 2 provides the ranges of real values for the normalized metrics that correspond to 0 and 1. The results show significant differences in variance ($p < 0.05$) for $\mathcal{L}_{\text{cIOU}}$ and $\mathcal{L}_{\text{pen}}$. The post hoc power analysis revealed a median power of 0.55 with a min–max range of $0.39 - 0.61$ for the Fligner–Killeen tests. The fixed insertion direction demonstrates lower variances in these metrics compared to the auto-determined insertion direction, indicating greater consistency in generating occlusal surfaces. As a result, the fixed insertion direction was adopted for subsequent evaluations.

Fig. 8 displays the results for all normalized metrics for the main part of the study. The data is arranged in descending order of input data quantity from left to right. Table 2 provides the ranges of real values for the normalized metrics that correspond to 0 and 1. Additionally, Table 3 provides a comprehensive summary of the statistical differences between the three comparison groups for the evaluation metrics. Significant differences (p; power) between groups were detected for $\mathcal{L}_{\text{chamfer,L2}}$ (0.007; 0.80) and $\mathcal{L}_{\text{pen}}$ (0.018; 0.75) based on Friedman Chi-Square tests. The post hoc Wilcoxon tests revealed

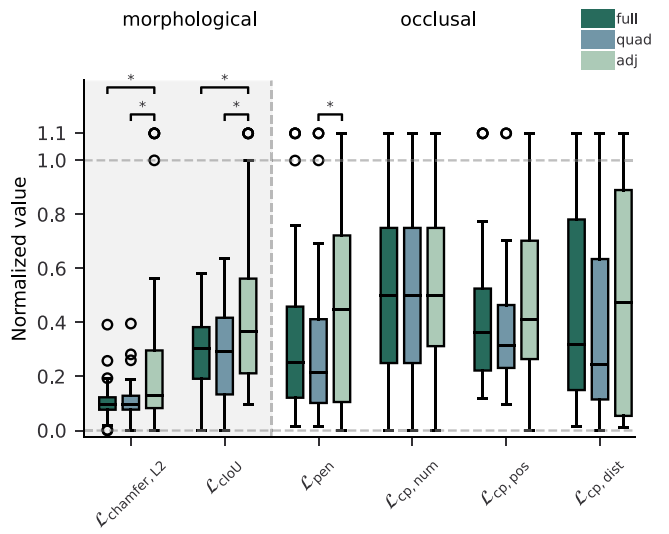


Fig. 8. Normalized results of the evaluation for the different comparison groups with $n = 30$ patients each. The labels on the significance bars refer to statistically significant deviations of the metrics according to Wilcoxon signed-rank tests. The significance level is set to $\alpha = 0.05$. Lower values indicate better performance.

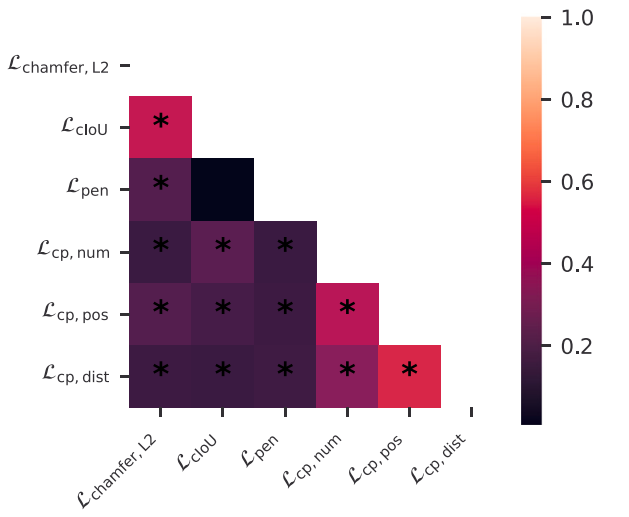


Fig. 9. Kendall correlations between the evaluation metrics. The correlation coefficients are color-coded according to the strength of the correlation with values ranging from −1 (strong negative correlation) to 1 (strong positive correlation). Statistical significances ($p < 0.05$) are marked with asterisks..

Table 2
Ranges of real values for the evaluation metrics in Figs. 7 and 8. The min and max values correspond to 0 and 1 in the normalized results, respectively. Lower values indicate better performance.

Metric	$\text{min}_{\text{Fig. 7}}$	$\text{max}_{\text{Fig. 7}}$	$\text{min}_{\text{Fig. 8}}$	$\text{max}_{\text{Fig. 8}}$
$\mathcal{L}_{\text{chamfer,L2}}$	0.379	0.448	0.247	1.616
$\mathcal{L}_{\text{cIoU}}$	0.290	0.474	0.198	0.648
$\mathcal{L}_{\text{pen}}$	0.001	0.034	0.001	0.153
$\mathcal{L}_{\text{cp,num}}$	1.000	2.000	0.001	4.000
$\mathcal{L}_{\text{cp,pos}}$	2.314	3.721	0.511	4.755
$\mathcal{L}_{\text{cp,dist}}$	1.152	5.813	0.022	5.975

Table 3
Statistically significant differences ($p < 0.05$) of the evaluation metrics across the different comparison groups based on Wilcoxon tests with H_1: median$(g_i - g_j) > 0$. Significant differences for the respective comparison group (column) and the metric (row) are shown as (p; power).

Metric	G_1: quad–full	G_2: adj–quad	G_3: adj–full
$\mathcal{L}_{\text{chamfer,L2}}$	–	0.008; 0.80	0.001; 0.97
$\mathcal{L}_{\text{cIoU}}$	–	0.041; 0.55	0.037; 0.57
$\mathcal{L}_{\text{pen}}$	–	0.016; 0.70	–
$\mathcal{L}_{\text{cp,num}}$	–	–	–
$\mathcal{L}_{\text{cp,pos}}$	–	–	–
$\mathcal{L}_{\text{cp,dist}}$	–	–	–

significant differences in the medians of the metrics across the comparison groups with additional pairwise significances in $\mathcal{L}_{\text{cIOU}}$. The power analysis showed a median power of 0.86 (range: 0.66–0.99).

The software failed to generate a morphologically sound reconstruction in four instances, all of which occurred in the adj group. Occlusion establishment failed across all groups for one specific patient, with the full group exhibiting one additional occlusion failure for a different patient. These failures are separate from cases where the software was unable to generate a morphologically sound reconstruction. All of these cases were included in the evaluation and assigned the respective metric values of 1.1. Additionally, the software failed to correctly identify the reconstruction target tooth in 23 cases within the adj group, requiring manual user intervention to select the correct tooth. The quad and full groups each exhibited one instance of failed target tooth identification, both involving the same patient.

The results reveal that the overall morphological details and spatial characteristics of the occlusal surfaces compared to the GT ($\mathcal{L}_{\text{chamfer,L2}}$, $\mathcal{L}_{\text{cIoU}}$) are significantly affected ($p < 0.05$) when the data quantity drops below the respective quadrant of the reconstruction target. Statistically significant differences in medians for these metrics are observed for comparison groups G_2 and G_3 in favor of higher data quantities.

Among the occlusion-specific metrics, the penetration loss ($\mathcal{L}_{\text{pen}}$) is the only metric to show significant differences. Medians differ significantly for comparison group G_2. The remaining metrics ($\mathcal{L}_{\text{cp,num}}$, $\mathcal{L}_{\text{cp,pos}}$, $\mathcal{L}_{\text{cp,dist}}$) exhibit no statistically significant differences between groups, indicating that the number and distribution of contact points are not influenced by data quantity.

The correlation analysis between the metrics in Fig. 9 reveals mostly weak to moderate correlations with a median of 0.19 and a min–max range of $0.01 - 0.54$. $\mathcal{L}_{\text{cIoU}}$ shows moderate correlations with $\mathcal{L}_{\text{chamfer,L2}}$. Similarly, $\mathcal{L}_{\text{cp,pos}}$ shows moderate correlations with $\mathcal{L}_{\text{cp,num}}$ and $\mathcal{L}_{\text{cp,dist}}$.

4. Discussion

The results indicate that an increase in input data quantity significantly improves the morphological accuracy of AI-designed dental crowns. However, since only one occlusion-specific metric showed a significant difference, the hypothesis is only partially supported. This suggests that occlusal relationships are largely independent of input data quantity and mainly determined by the presence and morphology of the antagonist teeth. The penetration loss ($\mathcal{L}_{\text{pen}}$) is the only occlusion specific metric to show significant differences between groups quad and adj. Notably, there was no significant difference between the full and adj group. While the results suggest that the software can benefit from increased input data quantity to achieve a more accurate strength of the established contact points, the sample size might not be sufficient to detect smaller effects. Future studies with larger sample sizes could provide further insights into the impact of data quantity specifically on the occlusion-specific metrics.

If the input data solely consists of the adjacent teeth (adj), the software's potential accuracy is significantly impaired. This leads to statistically significant deviations of the reconstructions general morphology ($\mathcal{L}_{\text{chamfer,L2}}$) and spatial characteristics ($\mathcal{L}_{\text{cIoU}}$) compared to the GT. It is to be noted that while no significant difference between groups was detected with Friedman tests for $\mathcal{L}_{\text{cIoU}}$, the post hoc Wilcoxon tests revealed significant differences between the adj and the remaining groups. Since the power of the pairwise tests for $\mathcal{L}_{\text{cIoU}}$ was below

the conventional threshold of 0.80, these results should be considered exploratory and require further validation in future research.

Although the adj data appears to include all the necessary information to establish the correct relationship between the target tooth and its surrounding teeth, the software benefits from supplementary data to produce more accurate outcomes. Apart from overall decreased accuracy as shown by the metrics, the software further struggles to identify the correct target tooth, establish occlusion or generate a morphologically sound reconstruction for the adj group. While quad and full groups did not exhibit any failed reconstructions, approx. 13% of the adj group reconstructions were not successful with output that severely deviated from the morphology of a human tooth. This indicates that the overall performance of the software is dependent on the presence of the quadrant data.

Data quantities above the respective quadrant of the reconstruction target do not lead to significant improvements in the occlusal surface quality. This suggests that the specific software does not effectively utilize the additional data present in full jaw scans compared to the quadrant scans. Given that the software is based on a 2D deep learning approach [9], the inherent data loss during the 2D–3D conversion process likely explains the observed outcomes. That is, even though full jaw scans contain more data, the conversion process primarily preserves the most salient features that are already captured in the quadrant scans. Furthermore, training data may not have included sufficient information to introduce necessary connections between the morphological features of the target tooth and teeth from other quadrants. As a result, the model was unable to effectively leverage the additional inter-quadrant context provided by full-jaw scans.

The exclusive use of the single software solution may therefore constrain the generalizability of the findings. Given the proprietary nature of the methodologies employed in commercial software, performance outcomes can differ across platforms. Future research should evaluate multiple AI-based software solutions on identical datasets to provide a more comprehensive understanding of the impact of data quantity on occlusal surface reconstruction.

Finally, there is the software independent possibility of anatomical irrelevance in the additional data, which would not contribute to the reconstruction process even it was utilized by the software. Based on the presented data, future studies can be designed with larger sample sizes to further investigate the effect of data quantity and detect potential smaller effects between full and quad groups for multiple software solutions.

Other approaches for occlusal surface reconstruction using 2D deep learning methods [15,19–22] have shown good results even with limited input data that is similar to the adj group. As commonly observed in DL applications, the data used for training and testing are often subsets of a shared dataset that may not be representative of an international patient population. The used dataset for this study is based on a European patient population, while the software is developed by a Korean company and therefore may be trained on a different ethnical dataset. This discrepancy in demographic and anatomical characteristics could partly explain the observed differences in performance compared to other studies using self-validated methods. To thoroughly evaluate the generalizability of AI-based software, future studies should consider testing the models on diverse, multi-ethnic datasets that encompass a broader range of patient demographics. Cross-validation across different geographic regions and ethnic groups would not only clarify the impact of demographic variation on performance but also show necessary adjustments for new methods to be reliably applied in various clinical settings worldwide.

The correlation analysis revealed that the metrics are mostly uncorrelated, indicating that they capture different aspects of the occlusal surfaces. The moderate correlation between the $\mathcal{L}_{\text{cIoU}}$ and $\mathcal{L}_{\text{chamfer,L2}}$ can be attributed to the shared spatial characteristics these metrics evaluate, albeit at different scales. The moderate correlations between the $\mathcal{L}_{\text{cp,pos}}$ and the $\mathcal{L}_{\text{cp,num}}$, $\mathcal{L}_{\text{cp,dist}}$ metrics, respectively, suggest that the spatial distribution of the contact points is related to the number and distance of the contact points. That is, with a higher deviation in the number of contact points and a differing spatial spread, the positional deviation of the contact points also increases with respect to the GT. However, due to the weak to moderate correlations, the metrics are not considered redundant and provide complementary information about the quality of the generated occlusal surfaces.

5. Conclusions

This study evaluated the impact of input data quantity on the morphology of crowns generated by AI-based dental restoration software. The results indicate that while the software achieves consistent results with full and quadrant input data, its performance significantly declines when the input is reduced to adjacent teeth only. Specifically, reductions in input data led to statistically significant impairments in the overall morphology and spatial characteristics of the reconstructed occlusal surfaces, as evidenced by higher values in CD ($\mathcal{L}_{\text{chamfer,L2}}$) and complemented IoU ($\mathcal{L}_{\text{cIoU}}$). Additionally, the frequency of reconstruction failures and target tooth misidentifications increased. The results for $\mathcal{L}_{\text{cIoU}}$ suggest that an increase in data quantity contributes to greater accuracy in the fit of the crown with respect to its surrounding teeth, i.e. the antagonist and adjacent teeth.

The proposed occlusion metrics represent clinically relevant parameters, including the number of contact points, the distances between them, their spatial distribution, and the contact strength. By adhering to established dental concepts, these metrics offer a comprehensive and clinically meaningful evaluation of the generated occlusion.

The analysis of the occlusion metrics indicates that the quality of occlusal relationships (number, position, and distribution of contact points) shows minimal dependency on the quantity of input data. Among these metrics, only the penetration loss ($\mathcal{L}_{\text{pen}}$) demonstrated a statistically significant deviation, implying that a larger quantity of input data enhances the accuracy of contact point strength.

The insertion direction was found to significantly impact the variance of the occlusal surfaces generated by the software for an identical clinical situation. That is, suboptimal insertion directions due to specific preparation designs may lead to increased variance in the generated occlusal surfaces.

The study also identified limitations in the software's ability to utilize full jaw data effectively. This may stem from the 2D–3D conversion process inherent in the software's methodological approach. The results underscore the importance of comprehensive data acquisition in clinical workflows in order to achieve reliable and accurate dental restorations with AI-based software and reduce the need for manual intervention. Based on the presented results, clinical practitioners should be aware of the potential limitations of AI-based dental restoration software when using reduced input data quantities. The findings suggest that the software's performance is significantly impacted by the quantity of input data (scanned area), with adjacent teeth data alone potentially yielding suboptimal or no results and necessitating a rescanning process to obtain adequate information.

CRediT authorship contribution statement

Alexander Broll: Writing – original draft, Methodology, Investigation, Conceptualization. **Markus Goldhacker:** Writing – review & editing. **Sebastian Hahnel:** Writing – review & editing. **Martin Rosentritt:** Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI's ChatGPT in order to improve the readability and language of the

manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Yuan, Y. Sun, Y. Wang, P. Lv, Computer-aided design of tooth preparations for automated development of fixed prosthodontics, Comput. Biol. Med. 44 (2014) 10–14, http://dx.doi.org/10.1016/j.compbiomed.2013.10.019.

[2] H. Watanabe, C. Fellows, H. An, Digital technologies for restorative dentistry, Dent. Clin. North Am. 66 (4) (2022) 567–590, http://dx.doi.org/10.1016/j.cden.2022.05.006.

[3] V. Preis, T. Dowerk, M. Behr, C. Kolbeck, M. Rosentritt, Influence of cusp inclination and curvature on the in vitro failure and fracture resistance of veneered zirconia crowns, Clin. Oral Investig. 18 (3) (2014) 891–900, http://dx.doi.org/10.1007/s00784-013-1029-9.

[4] K. Schnitzhofer, A. Rauch, M. Schmidt, M. Rosentritt, Impact of the occlusal contact pattern and occlusal adjustment on the wear and stability of crowns, J. Dent. 128 (2023) 104364, http://dx.doi.org/10.1016/j.jdent.2022.104364.

[5] A. Broll, M. Goldhacker, S. Hahnel, M. Rosentritt, Generative deep learning approaches for the design of dental restorations: A narrative review, J. Dent. 145 (2024) 104988, http://dx.doi.org/10.1016/j.jdent.2024.104988.

[6] A. Broll, M. Rosentritt, T. Schlegl, M. Goldhacker, A data-driven approach for the partial reconstruction of individual human molar teeth using generative deep learning, Front. Artif. Intell. 7 (2024) 1339193, http://dx.doi.org/10.3389/frai.2024.1339193.

[7] Imagoworks Inc., Dentbird crown, v.3.2.3, 2024, URL https://dentbird.com/.

[8] F. Wang, T. Zhang, Q. Zhou, Y. Lu, Comparison of the morphological accuracy of automatic crowns designed by multiple computer-aided design software programs with different levels of dentition information acquisition, J. Prosthet. Dent. 132 (2) (2024) 441–452, http://dx.doi.org/10.1016/j.prosdent.2023.01.024.

[9] J.-H. Cho, Y. Yi, J. Choi, J. Ahn, H.-I. Yoon, B. Yilmaz, Time efficiency, occlusal morphology, and internal fit of anatomic contour crowns designed by dental software powered by generative adversarial network: A comparative study, J. Dent. 138 (2023) 104739, http://dx.doi.org/10.1016/j.jdent.2023.104739.

[10] Z. Wu, C. Zhang, X. Ye, Y. Dai, J. Zhao, W. Zhao, Y. Zheng, Comparison of the efficacy of artificial intelligence-powered software in crown design: An in vitro study, Int. Dent. J. (2024) http://dx.doi.org/10.1016/j.identj.2024.06.023.

[11] J.-H. Cho, G. Çakmak, Y. Yi, H.-I. Yoon, B. Yilmaz, M. Schimmel, Tooth morphology, internal fit, occlusion and proximal contacts of dental crowns designed by deep learning-based dental software: A comparative study, J. Dent. 141 (2024) 104830, http://dx.doi.org/10.1016/j.jdent.2023.104830.

[12] H. Zhu, X. Jia, C. Zhang, T. Liu, ToothCR: A two-stage completion and reconstruction approach on 3D dental model, in: 26th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Springer, [S.l.], 2022, pp. 161–172, http://dx.doi.org/10.1007/978-3-031-05981-0_13.

[13] H. Fan, H. Su, L. Guibas, A point set generation network for 3D object reconstruction from a single image, in: 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Piscataway, NJ, 2017, pp. 2463–2471, http://dx.doi.org/10.1109/CVPR.2017.264.

[14] R.Q. Charles, H. Su, M. Kaichun, L.J. Guibas, PointNet: Deep learning on point sets for 3D classification and segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, http://dx.doi.org/10.1109/cvpr.2017.16.

[15] J.-J. Hwang, S. Azernikov, A.A. Efros, S.X. Yu, Learning beyond human expertise with generative models for dental restorations, 2018, arXiv:1804.00064v1.

[16] D. Dimitrov, C. Knauer, K. Kriegel, G. Rote, Bounds on the quality of the PCA bounding boxes, Comput. Geom. 42 (8) (2009) 772–789, http://dx.doi.org/10.1016/j.comgeo.2008.02.007.

[17] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, L. Wan-Yen, J. Johnson, G. Gkioxari, Accelerating 3D deep learning with PyTorch3D, 2020, arXiv:2007.08501v1.

[18] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619, http://dx.doi.org/10.1109/34.1000236.

[19] F. Yuan, N. Dai, S. Tian, B. Zhang, Y. Sun, Q. Yu, H. Liu, Personalized design technique for the dental occlusal surface based on conditional generative adversarial networks, Int. J. Numer. Methods Biomed. Eng. 36 (5) (2020) e3321, http://dx.doi.org/10.1002/cnm.3321.

[20] S. Tian, M. Wang, F. Yuan, N. Dai, Y. Sun, W. Xie, J. Qin, Efficient computer-aided design of dental inlay restoration: A deep adversarial framework, IEEE Trans. Med. Imaging 40 (9) (2021) 2415–2427, http://dx.doi.org/10.1109/tmi.2021.3077334.

[21] S. Tian, M. Wang, N. Dai, H. Ma, L. Li, L. Fiorenza, Y. Sun, Y. Li, DCPR-GAN: Dental crown prosthesis restoration using two-stage generative adversarial networks, IEEE J. Biomed. Heal. Informatics 26 (1) (2022) 151–160, http://dx.doi.org/10.1109/jbhi.2021.3119394.

[22] S. Tian, R. Huang, Z. Li, L. Fiorenza, N. Dai, Y. Sun, H. Ma, A dual discriminator adversarial learning approach for dental occlusal surface reconstruction, J. Heal. Eng. 2022 (2022) 1–14, http://dx.doi.org/10.1155/2022/1933617.