TranSDFNet: Transformer-Based Truncated Signed Distance Fields for the Shape Design of Removable Partial Denture Clasps

Xinze Shen, Changdong Zhang, *Member, IEEE*, Xiuyi Jia, *Member, IEEE*, Dawei Li, Tingting Liu, Sukun Tian, Wei Wei, Yuchun Sun, *Member, IEEE*, and Wenhe Liao

***Abstract*—The ever-growing aging population has led to an increasing need for removable partial dentures (RPDs) since they are typically the least expensive treatment options for partial edentulism. However, the digital design of RPDs remains challenging for dental technicians due to the variety of partially edentulous scenarios and complex combinations of denture components. To accelerate the design of RPDs, we propose a U-shape network incorporated with Transformer blocks to automatically generate RPD clasps, one of the most frequently used RPD components. Unlike existing dental restoration design algorithms, we introduce the voxel-based truncated signed distance field (TSDF) as an intermediate representation, which reduces the sensitivity of the network to resolution and contributes to more smooth reconstruction. Besides, a selective insertion scheme is proposed for solving the memory issue caused by Transformer blocks and enables the algorithm to work well in scenarios with insufficient data. We further design two weighted loss functions to filter out the noisy signals generated from the zero-gradient areas in TSDF. Ablation and comparison studies demonstrate that our algorithm outperforms state-of-the-art reconstruction methods by a large margin and can serve as an intelligent auxiliary in denture design.**

***Index Terms*—Partial edentulism, RPD clasp design, transformer, deep learning, truncated signed distance field.**

Manuscript received 18 December 2022; revised 30 May 2023; accepted 9 July 2023. Date of publication 20 July 2023; date of current version 5 October 2023. This work was supported in part by the National Key R&D Program of China under Grant 2019YFB1706900, in part by the National Natural Science Foundation of China under Grants 52035001 and 52105265, and in part by the Beijing Training Project for the Leading Talents in S&T under Grant Z191100006119022. *(Xinze Shen and Changdong Zhang contributed equally to this work.) (Corresponding authors: Tingting Liu; Sukun Tian; Wenhe Liao.)*

Xinze Shen, Changdong Zhang, Dawei Li, Tingting Liu, and Wenhe Liao are with the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: sxz@njust.edu.cn; zcd@njust.edu.cn; ldw@njust.edu.cn; liutingting@mail.njust.edu.cn; cnwho@mail.njust.edu.cn).

Xiuyi Jia is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: jiaxy@njust.edu.cn).

Sukun Tian and Yuchun Sun are with the Center of Digital Dentistry, Faculty of Prosthodontics, Peking University School and Hospital of Stomatology, Beijing 100081, China, also with the National Center of Stomatology, Beijing 100081, China, also with the National Clinical Research Center for Oral Diseases, Beijing 100081, China, also with the National Engineering Research Center of Oral Biomaterials and Digital Medical Devices, Beijing 100081, China, also with the Beijing Key Laboratory of Digital Stomatology, Beijing 100081, China, and also with the Research Center of Engineering and Technology for Computerized Dentistry Ministry of Health, Beijing 100081, China (e-mail: sukhum169@hotmail.com; kqsyc@bjmu.edu.cn).

Wei Wei is with the Shandong Hoteam Software Company Ltd., Jinan 250101, China (e-mail: willwill@netease.com).

Digital Object Identifier 10.1109/JBHI.2023.3295387

I. Introduction

DENTAL diseases have become a global concern [1]. Partial edentulism [2], or partial tooth loss, is one of the most prevalent oral diseases and continuously imposes an enormous financial burden on society [3]. In 2015, the costs of dental diseases worldwide were estimated at $544.41 billion. Meanwhile, 67% of productivity losses were attributable to severe tooth loss [4]. With the growth of the economy, the global prevalence of severe tooth loss has declined for decades [5]. However, a simultaneous increase in the actual number of edentulous people has arisen as a result of population growth and aging [6]. Additionally, there is a clear socioeconomic gradient in dental resources available between and within countries, restricting the application of advanced but expensive therapies such as fixed partial denture (FPD) treatment [7]. Compared with FPDs, removable partial dentures (RPDs) provide a minimally invasive and cost-effective option for subjects with partial edentulism, and they are widely used in clinical practice [8].

Despite its popularity, the digital design of complicated RPD frameworks is difficult for most junior dental technicians, even with the help of advanced dental software [9]. For example, the clasp (see Fig. 1) is a part of RPD that serves as a retainer by engaging a portion of the abutment tooth surface [10]. In popular digital design workflows, technicians need to manually complete the following steps: 1) drawing skeleton lines of the clasp arms, 2) brushing the coverage of the rest and minor connectors, and 3) smoothing the junction between components. The shape, size, and pose of the clasp need to be carefully matched with the abutment to provide adequate retentive force and ensure the successful insertion of the denture. Due to the lack of perception and comprehension of the residual dentition, traditional computer graphics algorithms cannot accomplish the design of personalized RPDs without manual intervention. Meanwhile, the complex components of RPD also remain challenging for traditional computer-aided design methods. The design principles vary with different RPD components and partially edentulous scenarios,

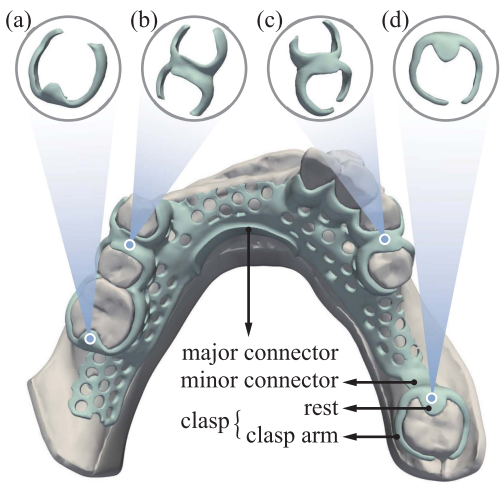

Fig. 1. Basic components of a removable partial denture with clasps highlighted. (a) and (d) are three-arm clasps. (b) and (c) are combined clasps.

which requires technicians to have high professional skills and software proficiency. The quality of RPD products thus strongly depends on the experience and level of technicians. It should be noted that several studies have reported automatic design methods for RPDs (Meada [11], MacRPD [12], RaPiD [13], and Chen [14]), but they are limited to generating 2D diagrams, which are far from end products. Therefore, a data-driven workflow of automatic RPD design, in which the experience of senior technicians is embedded, should be developed to reduce the burden on dental technicians.

Deep learning (DL), as an essential branch of machine learning, has been widely used in health informatics and contributed to more timely, effective, and objective treatment [15]. DL for dentistry is one of the most impressive advances. Abundant evidence shows that DL-based methods have tremendous advantages in dental image analysis, such as tooth segmentation [16], [17], [18], tooth detection [19], landmark localization [20], diagnosis of dental diseases [1], and so on. Besides, recent progress in DL-based dental restoration is also encouraging. Based on conditional generative adversarial networks (CGANs) [21], Hwang et al. [22] presented an approach for the auto-design of personalized dental crowns, which can generate the target occlusal surface directly from the crown-missing and the opposing depth scan. To achieve more anatomically realistic results, Yuan [23] and Tian [24] proposed new loss functions, enabling the generator to capture more functional features. A similar method has proven its success in dental inlay restoration [25]. Apart from depth map-based pipelines, competitive results can be found in point cloud-based research. Lessard [26] and Zhu [27] formulated the crown restoration as a point cloud completion issue, which extends the crown design to a three-dimensional analysis. However, high-quality RPDs designed by experienced technicians are hard to access. The intricate topology of RPDs poses a significant challenge for learning-based algorithms, especially GAN-based models, particularly when working with limited datasets.

Recently, implicit fields have gained popularity in 3D reconstruction because of their continuity and smoothness. Dai [28] represented partial scans as voxelized truncated signed distance fields (TSDFs) and proposed a CNN-based architecture to infer missing parts of objects. Chen [29] introduced an implicit field decoder for shape generation, which predicts the binary occupancy value based on the location of a query point and the latent code of global shape. Mescheder [30] designed a similar network combined with a multi-grid strategy to generate high-resolution meshes progressively. To reduce reliance on training data, Park [31] described an auto-decoder pipeline for learning signed distance fields (SDFs), named DeepSDF. Amiranashvili [32] introduced the implicit auto-decoder scheme into anatomical shape reconstruction and achieved excellent performance even with sparse measurements. However, due to the absence of an encoder, such networks can only be used for self-reconstruction, which is based on observations from the object itself, and cannot handle tasks where the input and output come from different objects (e.g., RPD design). Based on an encoder-decoder architecture [33], Ping [34] combined the implicit representation with the transformer for a more detailed reconstruction. However, their discussion did not include the memory overhead issue that transformers may cause. Furthermore, as a truncated version of SDF, the TSDF is not globally unit-gradient, and the zero-gradient areas contribute little to the shape learning. The information imbalance arising from gradients can even lead to a local optimum during training.

To tackle these issues, we develop a novel DL-based framework, called TranSDFNet, to provide component-level acceleration for digital RPD design. Given the importance of the clasps for the retention and successful insertion of the personalized RPDs, we evaluate our algorithm on a clinical RPD clasp dataset. Since it is widely accepted that the orderless nature of meshes and point clouds impedes the direct application of efficient convolution, we instead adopt voxel-based TSDFs as intermediate representations and formulate the design of RPD clasps as the regression of an implicit field. Given the advantage of the self-attention mechanism [35] in feature extraction, we incorporate Transformer blocks into the pipeline for more flexible information interaction, which is crucial for capturing the topology of the abutment. Moreover, to ease the information imbalance in TSDF, we propose two loss functions incorporated with weighting schemes, which encourage the network to focus on the near-surface areas.

The main contributions of our approach are summarized as follows:

1) We develop an end-to-end learning-based framework that can generate a realistic and smooth clasp directly from a selective abutment. Voxelized TSDFs are adopted as intermediate representations, and the mechanism of how TSDFs reduce their sensitivity to resolution is explored. To the best of our knowledge, this is the first work that shows how to automate the design of RPD components using 3D deep learning.
2) To facilitate the global feature capture and generation, we introduce the self-attention mechanism into the framework and achieve a balance between performance and

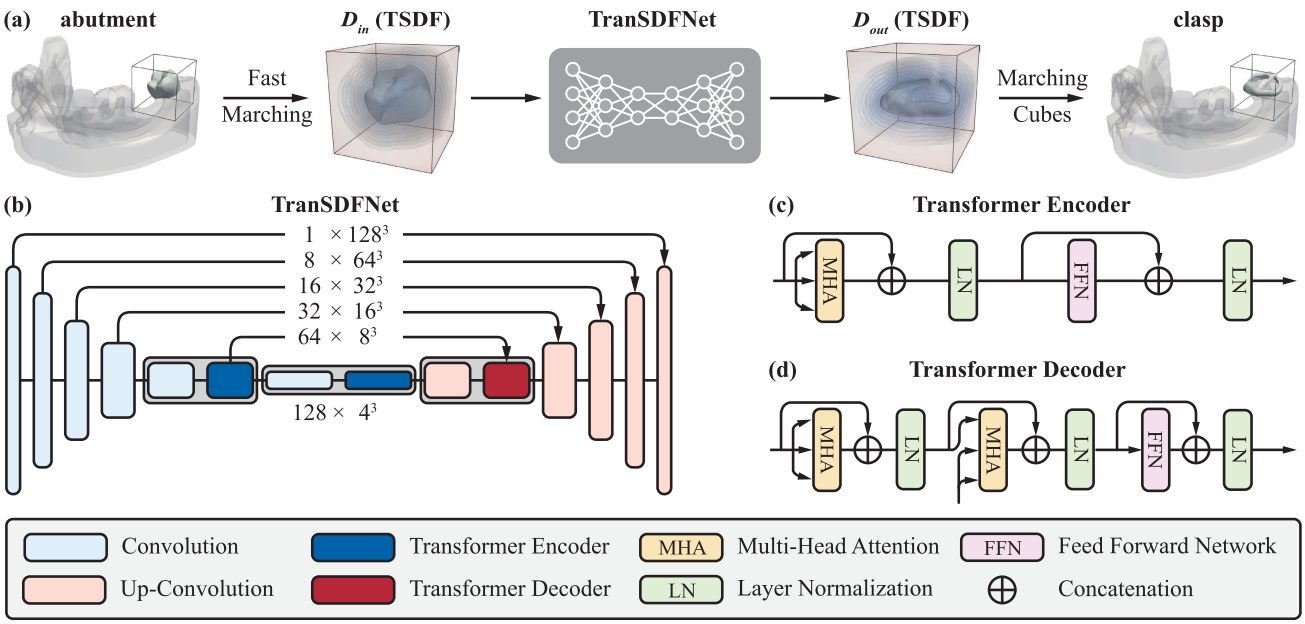

Fig. 2. Pipeline of RPD clasp design is illustrated in (a). The selected abutment is first represented as a TSDF ($\boldsymbol{D}_{\text{in}}$) with the Fast Marching algorithm and fed into the proposed TranSDFNet (b), which is composed of Convolution blocks (Convolution-InstanceNorm-LeakyReLU-MaxPool), Up-Convolution blocks (Upsample-Convolution-InstanceNorm-LeakyReLU), Transformer encoders (c) and Transformer decoders (d). Then, the TSDF of the target clasp ($\boldsymbol{D}_{\text{out}}$) is generated from the network and finally converted back to a surface using Marching Cubes.

memory overhead simply by adjusting the number of inserted Transformer blocks, without introducing complex Transformer structures or auxiliary operations. Besides, to mitigate the negative impact of zero-gradient regions on training and balance the information distribution, we design two weighted loss functions to filter out useless gradient information. Furthermore, we provide the first systematic and comprehensive comparison of loss functions for SDF/TSDF.

3) Ablation studies conducted on clinical data demonstrate the effectiveness of our algorithm even with limited training data. Qualitative and quantitative comparisons indicate that the design quality of TranSDFNet is significantly higher than state-of-the-art reconstruction methods.

The rest of this article is organized as follows. Section II describes the implicit pipeline, Transformer blocks, and loss function design. Section III presents the acquisition and preprocessing of RPD data, evaluation metrics, experimental setup, and result analysis. We discuss the benefits and limitations of our framework in Section IV.

II. Methods

By engaging with the abutment surface, RPD clasps are typically used to hold a removable dental prosthesis in place. Although the shape and size vary with the selected abutment, their topology remains statistically consistent for a specific type of clasps. Therefore, we propose to capture the morphological feature of the abutment and synthesize the clasp using deep learning. As illustrated in Fig. 2, with TSDF as an intermediate representation, our approach takes as input the selected abutment and directly generates the matching clasp. To facilitate the perception of global features, we introduce Transformer blocks into the implicit design pipeline. In addition, we design two weighted loss functions to narrow the region of interest, which fit well with TSDF.

A. Deep Implicit RPD Clasp Design

The pipeline of generation tasks highly depends on the representations of 3D objects. The point cloud is essentially the simplest form of a 3D model and has been widely used in 3D shape reconstruction [36]. However, its lack of topological adjacency leads to heavy time consumption in establishing the neighborhood. The triangle mesh is a well-defined alternative with edges providing adjacency information, yet it is hard to integrate it into the learning methods since it is orderless in nature. The depth map can only be used to represent a part of the object due to the existence of self-occlusion. Besides, a majority of learning-based approaches adopt voxelized representations, such as binary occupancy grids (BOGs), where each grid point stores an occupancy value, with zero indicating empty and one indicating occupied. The boundary between empty cells and occupied cells is an approximation of the object's surface. Despite its excellent compatibility with convolution operations, vanilla voxelization will result in irreversible loss of information and poor reconstruction quality, especially when the object contains small or thin features, such as the arms of clasps.

To achieve a balance between information integrity and structural regularity, we propose to use the voxel-based TSDF, with each grid point saving the signed distance from itself to the surface rather than a binary indicator. As defined in [37], given

a watertight surface $\mathcal{S} \subset \mathbb{R}^3$, which is the interface of domain Ω, the unsigned distance field $d(\boldsymbol{x})$ is:

$$d(\boldsymbol{x}) = \min_{\boldsymbol{x}_I \in \partial\Omega} (|\boldsymbol{x} - \boldsymbol{x}_I|), \tag{1}$$

where $\boldsymbol{x}$ is a point in Ω and $\partial\Omega$ denotes the boundary of Ω, i.e., surface $\mathcal{S}$. A signed distance field $sdf(\boldsymbol{x})$ can thus be represented as:

$$sdf(\boldsymbol{x}) = \begin{cases} d(\boldsymbol{x}), & \boldsymbol{x} \in \Omega^+ \\ 0, & \boldsymbol{x} \in \partial\Omega \\ -d(\boldsymbol{x}), & \boldsymbol{x} \in \Omega^- \end{cases}, \tag{2}$$

where Ω^+ and Ω^- denote the exterior and interior of Ω, respectively.

In practice, to reduce the redundancy of information, we often use the variant of SDF (i.e., TSDF [38]), which is truncated to a maximum distance δ. An excessively large truncation threshold can cause TSDF inside the bounding box to degrade into SDF. Conversely, an overly small truncation threshold will make most regions in TSDF zero-gradient, which may impede the training of the network (refer to Section II-C). Therefore, the truncation threshold δ is empirically set to 0.5. The TSDF is defined as:

$$tsdf(\boldsymbol{x}) = \begin{cases} 0.5, & sdf(\boldsymbol{x}) > 0.5 \\ sdf(\boldsymbol{x}), & |sdf(\boldsymbol{x})| \le 0.5 \\ -0.5, & sdf(\boldsymbol{x}) < -0.5 \end{cases}. \tag{3}$$

In this article, we use the voxelized version of TSDF, denoted as $\boldsymbol{D} \in \mathbb{R}^{N \times N \times N}$. As depicted in Fig. 2, both the input abutment and the output clasp are represented as voxelized TSDF, denoted as $\boldsymbol{D}_{\text{in}}$ and $\boldsymbol{D}_{\text{out}}$. Then, the design of RPD clasps is converted to learning a consistent mapping from $\boldsymbol{D}_{\text{in}}$ to $\boldsymbol{D}_{\text{out}}$, which is a regression problem.

Concretely, given the closed surface of an abutment, we have to represent it as a voxelized TSDF, which is a problem of solving Eikonal equation on a discrete domain. The Fast Marching algorithm [39] is an effective numerical method to solve this kind of equation and is adopted in this article. We first normalize the input surface to $[-1, 1]^3$ and convert it to a 128^3 TSDF using the Fast Marching algorithm. Then, we extract the geometric features of the abutment using successive 3D convolution blocks, which consist of Conv3D (kernel size $= 3$), InstanceNorm, LeakyReLU, and MaxPooling (kernel size $= 2$). The local-to-global encoder enables capturing both low-level information (e.g., texture) and high-level information (e.g., shape and size), which is essential for RPD clasp design. During the decoding, to reduce the memory footprint, we adopt the combination of trilinear interpolation and convolution as the core of up-convolution blocks, specifically including Upsampling (scale factor $= 2$), Conv3D (kernel size $= 3$), InstanceNorm, and LeakyReLU. The decoder takes as input the feature generated from the encoder, i.e., a 128-dimensional tensor with a resolution of 4^3. After several upsampling layers, the decoder outputs the TSDF of the desired clasp $\boldsymbol{D}_{\text{out}}$ with a resolution of 128^3. Besides, skip connections are used to address vanishing gradients. Finally, the mesh of the clasp can be extracted from the voxel-based TSDF using Marching Cubes [40]. More details about the network can be found in Section III.

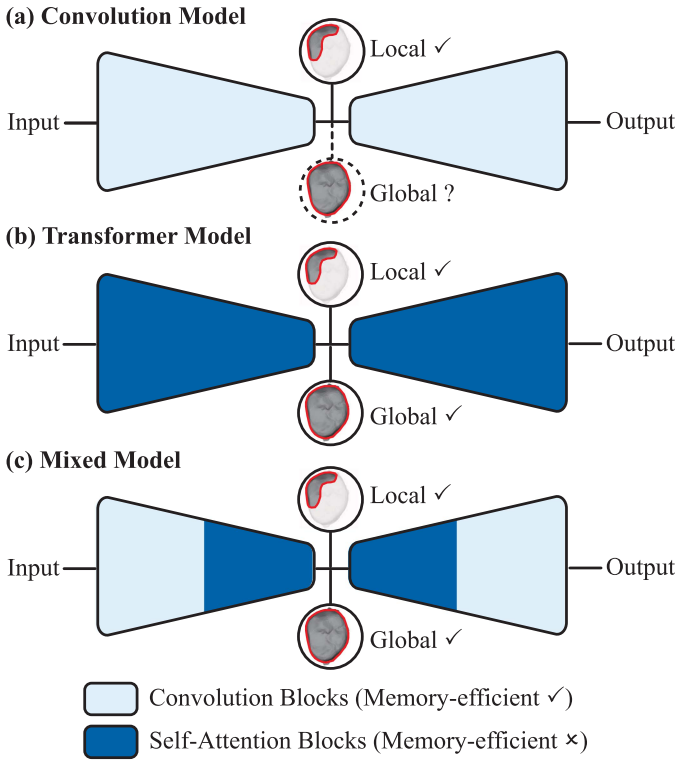

Fig. 3. Illustration of differences between convolution models, Transformer models and mixed models in RPD clasp design. (a) Convolution models capture local features through cascading convolution blocks, however, they do not guarantee the effective generation of global features. (b) Transformer models can capture both local and global features through self-attention operations. (c) Mixed models obtain a balance of feature extraction and memory overhead by replacing the self-attention blocks in high-resolution layers with convolutional blocks.

B. Self-Attention Encoder and Decoder

Despite all the advantages, implicit representations describe a family of iso-surfaces and commonly lead to more complex data distributions compared with BOGs, which only represent a single surface. It is hard for traditional convolution models (see Fig. 3(a)) to capture the underlying correspondences between implicit inputs and outputs due to their weakness in capturing global features. In comparison, Transformer models (see Fig. 3(b)) can directly establish global correspondences via self-attention operations [35]. More flexible information interaction enables Transformers to better extract high-level features crucial for the design of RPD clasps, such as the abutment shape and size. Therefore, we incorporate Transformer blocks into our baseline for more reasonable results.

However, there is a trade-off between the benefits of self-attention and the cost of memory footprint due to its quadratic complexity with respect to the input size. It is usually unacceptable to perform self-attention operations on high-resolution inputs, especially 3D volume data, such as a grid with a resolution of 128^3. Besides, the large number of learnable parameters leads to the need for large amounts of training data, hampering its clinical practice in prosthesis design, because high-quality RPDs designed by experienced technicians are often difficult to obtain.

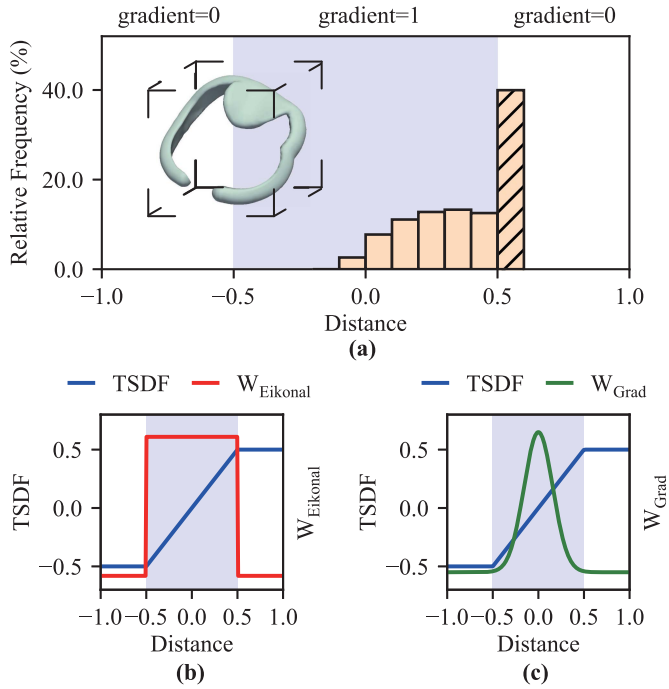

Fig. 4. Relative frequency histogram of clasp TSDF and illustration of the two proposed weighting schemes. (a) Most of the regions in clasp TSDF have positive distances, with approximately 40% of the values being greater than the truncation threshold (i.e., 0.5). (b) Truncated Eikonal Loss filters out noisy signals generated from zero-gradient areas in TSDF via a piecewise weighting function (i.e., the red line). (c) Weighted Gradient Loss serves the same purpose, but with a continuous function (i.e., the green line).

To tackle these issues, we recommend inserting self-attention modules after the middle layers of the network (i.e., layers with resolutions of 4^3 and 8^3), rather than all layers, as shown in Fig. 3(c). In this way, the performance of the network can be improved with the memory footprint remaining at a relatively low level. In addition, we directly adopt the implementation of *Transformer* encoders and decoders described in [35], which are mainly composed of Multi-Head Attention (MHA) blocks, Feed Forward Networks (FFN), and Layer Normalization (LN).

C. Loss Functions

For small-sample learning, prior constraints can help avoid overfitting and improve generalization. The unit-gradient property of SDF is a useful shape prior, which has proven its advantages in 3D reconstruction [41], [42]. However, TSDF is not globally unit-gradient. When the truncation threshold is set to 0.5, only the area where the signed distance is between -0.5 and 0.5 has a unit gradient, while the gradient in other areas is zero (refer to the interval with blue background in Fig. 4). For TSDF of thin structures like RPD clasps, most of the regions have positive distances, with up to 40% of them being zero-gradient (see Fig. 4(a)). The insufficient information inside the zero iso-surface and the redundant information from the zero-gradient regions result in a severe information imbalance, which undermines the role of shape prior to some extent. To tackle this issue, we use weighting strategies to filter out the gradient signals generated from outside the truncation boundary.

Concretely, we formulate the loss functions as follows:

$$L = L_{\text{base}} + WL_{TSDF}, \tag{4}$$

where L_{base} is a basic regression loss for the global error evaluation, and we adopt Smooth L1 Loss as the L_{base} in this article. L_{TSDF} is a Gradient regularization item. W is a weighting function tailored to fit the specific truncation threshold of TSDF. Building upon the existing gradient loss functions, we validate two weighting schemes, including a step form (Truncated Eikonal Loss, $L_{T-\text{Eikonal}}$) and a smooth form (Weighted Gradient Loss, $L_{W-\text{Grad}}$).

1) Truncated Eikonal Loss: The Eikonal regularization [41], [42] has achieved success in SDF prediction by introducing shape prior (i.e., the unit gradient of SDF) to the optimization. However, as aforementioned, it cannot be directly used to constrain the shape of TSDF because the gradient of most areas in TSDF is zero. To reduce the negative influence of the zero-gradient areas, we suggest using a piecewise weighting function W_{Eikonal} (see Fig. 4(b)) that activates the Eikonal regularization only when the gradient is non-zero. W_{Eikonal} is defined as:

$$W_{\text{Eikonal}} = \begin{cases} \lambda, & \text{if } |y| < \delta \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where λ is a scale factor. δ is used to adjust the region of interest. To fully utilize the gradient information while avoiding introducing noise from zero-gradient regions, δ is made to be consistent with the truncation threshold of TSDF (i.e., 0.5) Taking the Eikonal regularization as L_{TSDF}, the Truncated Eikonal Loss can be represented as:

$$L_{T-\text{Eikonal}} = \begin{cases} L_{\text{base}} + \lambda(\|\nabla\hat{y}\|_2 - 1)^2, & \text{if } |y| < \delta \\ L_{\text{base}}, & \text{otherwise} \end{cases}, \tag{6}$$

where y and $\hat{y}$ denote the ground truth and predicted distance value, respectively. $(\|\nabla\hat{y}\|_2 - 1)^2$ is the Eikonal item and $\|\cdot\|_2$ denotes L_2-norm. λ is used to balance the two parts of the loss function and is empirically set to 0.005.

2) Weighted Gradient Loss: Inspired by the Gaussian function, we also evaluate a continuous weighting scheme for a smoother convergence (see Fig. 4(c)). It should be noted that we adopt a bilateral form of Gaussian function, defined as:

$$W_{\text{Grad}} = \alpha\left(e^{-\beta y^2} + e^{-\beta\hat{y}^2}\right), \tag{7}$$

where α and β are respectively used to control the height and width of the Gaussian function. Based on additional tests, when α and β are set to 0.04 and 20, respectively, the loss function can effectively amplify information in unit-gradient regions and suppress noise in zero-gradient regions.

With higher weights assigned to the areas where $\hat{y}$ or y is close to zero, we encourage the network to focus on the near-surface areas both in the prediction and the ground truth, which can effectively facilitate shape learning. Taking the gradient loss introduced by [43] as L_{TSDF}, the Weighted Gradient Loss is

defined as:

$$L_{W-\mathrm{Grad}} = L_{\mathrm{base}} + \alpha \left(e^{-\beta y^2} + e^{-\beta \hat{y}^2} \right) \|\nabla y - \nabla \hat{y}\|_2, \quad (8)$$

where $\|\nabla y - \nabla \hat{y}\|_2$ is the L_2 distance between the gradient of the ground truth and prediction.

III. Experiments and Results

A. Dataset and Evaluation Metrics

We evaluate our algorithm on a realistic dataset, including scanned models of defective dentitions and RPDs designed by experienced technicians. The scanned dental data are acquired from [44], which has been approved by the Bioethics Committee of Peking University School and the Hospital of Stomatology, P. R. China. According to existing reports, missing molars are highly prevalent in clinical scenarios [45], and the adjacent molars are often selected as abutment teeth in RPD treatment. Therefore, we separate the molars (e.g., 2^{nd} molars in this article) and the corresponding clasps (e.g., three-arm clasps in this article) from the original models to build a sub-dataset for experiments. Our dataset contains 39 samples in total, some of which were segmented from different positions in the same dental arch. The database is further split into 30 samples for training and 9 for validation. Besides, considering the symmetry of a dentition model, we transform the teeth and clasps located at the mandible (i.e., #37 and #47) and the right side of the maxilla (i.e., #27) to the left side of the maxilla (i.e., #17) using the horizontal and vertical mirror flipping. In this way, the complexity of the problem can be significantly reduced.

To quantitatively evaluate our algorithm, we use the volumetric Intersection over Union (IoU) [39], Chamfer-L_2 Distance (CD-L_2) [46], Earth Mover's Distance (EMD) [46] and F-Score@1%[39] to measure the similarity between the predicted clasps and ground truth. IoU measures the overall similarity of topology, including functional features with clinical significance such as the engagement angle and end position of the clasp arm. CD-L_2 and EMD provide a more fine-grained evaluation of surface similarity on the point cloud, which can serve as important morphological and aesthetic indicators of clasp surface quality (such as smoothness). F-Score is calculated to reduce the bias of the above metrics, as suggested by [47].

B. Training Details

We use the PyTorch framework to implement TranSDFNet. All experiments are performed on a machine with two Intel(R) Xeon(R) Gold 6226R CPUs (2.90 GHz), four GeForce RTX 3090 GPUs (24 GB), and 251 GB RAM. Adam, with an initial learning rate of 0.001, is adopted to solve the optimization problem. With the batch size set as 16, we empirically train the network for 4000 epochs. The number of attention heads in Transformer blocks is set to 8. During the evaluation, point-based metrics (i.e., CD-L_2, EMD, and F-Score) are computed on 8192 points.

Besides, it should be noted that we do not use any data augmentation tricks in the network training. It is widely accepted that data augmentation can improve the generalization ability. However, the compatibility between data augmentation and specific tasks has received little attention, especially in industrial or clinical scenarios. Different from general 3D shape synthesis, the shape, size, pose, and location of RPD clasps have stronger prior constraints. For instance, before the design of clasps, the path of insertion has to be determined first, which can highly affect the pose of clasps. Since the poses of clasps in our dataset have been carefully predetermined, commonly-used transformations (e.g., translation, rotation, and scale), even with small amplitudes, may finally hinder the insertion of the denture. Hence, we do not use data augmentation despite the relatively small dataset.

TABLE I
Effect Analysis on TSDF and Sensitivity Analysis on Resolution

Resolution	Representation	IoU(↑)	CD-L_2(↓)	EMD(↓)	F-Score(↑)
32^3	BOG	6.85%	0.1592	0.2373	18.96%
	TSDF	**19.24%**	**0.1105**	**0.1947**	**19.96%**
64^3	BOG	**42.17%**	**0.0120**	**0.0707**	**83.52%**
	TSDF	35.61%	0.0222	0.0968	66.44%
128^3	BOG	**44.56%**	0.0123	0.0739	**84.98%**
	TSDF	41.75%	**0.0121**	**0.0725**	82.45%

The bold values denote the best results.

C. Ablation Study

Three ablation experiments are conducted to validate the effectiveness of our network and loss functions. To ensure that the proposed component causes any changes in the results, we take the architecture described in Section II as the standard model and replace only the component of interest in each experiment. Other variables that are not of interest but may affect the outcomes remain the same, such as the training epochs, batch size, optimizer parameters, et al. Details about the ablation experiments are described in the following sections.

1) Truncated Signed Distance Field: Voxelized TSDF is an augmented version of voxelized representation, which embeds the abutment/clasp into an implicit field. The signed distances saved in the grid points enable a more fine-grained shape representation. To validate its benefits, we compare TSDF against vanilla voxelized representation (i.e., BOG) with resolution kept the same. Besides, different from continuous TSDF [31], voxelized TSDF cannot avoid the issue of resolution. Although implicit representations have many advantages, low resolution means inadequate information input and output, which can seriously affect the quality of shape generation, especially for thin structures (e.g., clasp arms). To explore the sensitivity of voxelized TSDF to resolution, we evaluate the network at three different resolutions, i.e., 32^3, 64^3, and 128^3.

Table I proves the existence of a resolution issue. Both BOG and TSDF have poor performance when the resolution is 32^3, and the accuracy improves significantly as the resolution increases. Besides, as mentioned above, despite the information integrity of TSDF, its more complex data distribution makes training difficult and finally leads to the similar precision of BOG and TSDF in the high-resolution groups (i.e., 128^3). The benefits of TSDF can be more clearly observed in the visual results. As illustrated in Fig. 5, the result generated from the BOG-based

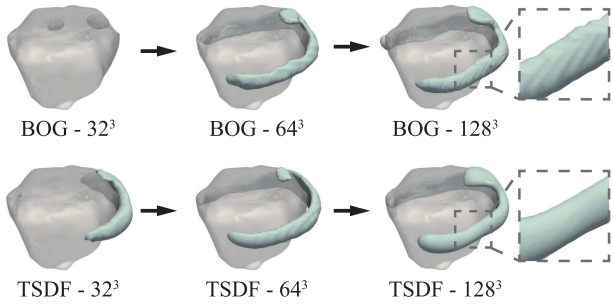

Fig. 5. Visual comparison between BOG and TSDF with grid resolutions of 32^3, 64^3, and 128^3. The local appearances of clasps are highlighted by dashed boxes.

model has obvious stepwise textures, which arise from the information loss during voxelization and cannot be effectively suppressed by further subdivision. In contrast, voxelized TSDF eradicates these pleats when the resolution is increased to 128^3, indicating its relatively low sensitivity to resolution. The high-quality surface generated by voxelized TSDF can be attributed to its richer information embedded in grid points. Specifically, vanilla voxelization directly pushes/pulls surface points to their nearest grid points which leads to irreversible deformation, while voxelized TSDF saves the distances between points which serve as more accurate localization information. The network can thus capture more fine-grained geometric features and finally generate a realistic RPD clasp with a more natural appearance.

2) Self-Attention Block: Since the development of Transformer models, it has been a common and effective trick to apply self-attention blocks to various tasks, such as dental image analysis [48], [49], [50]. Section II-B theoretically analyzes the mechanism of self-attention blocks for the PRD clasp design and introduces a selective insertion strategy to achieve a balance between precision and memory costs without modifying the attention structure. In this section, we provide detailed experiments to support the design of TranSDFNet.

Concretely, we insert self-attention encoders and decoders into different stages of the pipeline and measure the performance differences caused by the added blocks. As shown in the left part of Fig. 6, four structures are compared, including zero (T0), one (T1), three (T3), and five (T5) Transformer blocks, respectively. Limited by available computing resources, the evaluation of larger models, such as T7, is not discussed here. It should be noted that we do not display all skip connections in Fig. 6, most of which are built up symmetrically. Instead, only the skip connections between Transformer blocks are highlighted for clarity.

The performance differences between the four structures are depicted in the right part of Fig. 6. There is a clear trend of increasing IoU (T0: 37.08% $\rightarrow$ T1: 41.10% $\rightarrow$ T3: 41.75%) as we use more Transformer blocks, which proves the benefits of self-attention mechanisms. However, the design accuracy declines significantly when the number of self-attention layers exceeds three (T5: 36.62%). This special case arises because the available data cannot support the training of a network with a progressively increasing number of parameters. Compared with shape accuracy, the growing trend of memory overhead is more apparent. When the number of self-attention layers increases from zero to five, the GPU memory usage keeps rising from 420.49 MB to 5,677.28 MB (more than 10$\times$), indicating the relatively high complexity of the algorithm. Among the four structures, T3 achieves the highest shape precision with the memory cost remaining acceptable and is adopted as the basic architecture for RPD clasp design.

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT LOSS FUNCTIONS

Loss	IoU (↑)	CD-L_2 (↓)	EMD (↓)	F-Score (↑)
WeightLoss1 [46]	**40.93%**	**0.0139**	**0.0777**	**81.56%**
WeightLoss2 [41]	2.16%	0.0709	0.2084	40.53%
EikonalLoss [41], [42]	38.51%	0.0133	0.0750	82.02%
T-EikonalLoss (ours)	**40.27%**	**0.0115**	**0.0725**	**84.05%**
GradLoss [43]	38.79%	0.0132	0.0726	81.82%
W-GradLoss (ours)	**41.75%**	**0.0121**	**0.0725**	**82.45%**

The bold values denote the best results.

3) T-Eikonalloss and W-Gradloss: In Section II-C, we introduce two weighting strategies to filter out the noisy signals generated from zero-gradient areas in TSDF. To validate the efficacy of the proposed weighting schemes in RPD clasp design, we combine them with two existing loss functions (denoted as EikonalLoss [42] and GradLoss [43]) and derive two new loss functions (named as T-EikonalLoss and W-GradLoss). Their performances are measured on the same dataset. Furthermore, two other weighting loss functions (denoted as WeightLoss1 [46] and WeightLoss2 [41]) are included for a more comprehensive comparison. The main difference between our loss functions and WeightLoss1/WeightLoss2 is that we consider both value errors and weighted gradient errors of TSDF, while they only measure the weighted value errors of SDF without introducing shape prior. The definitions of all the above functions are listed in Appendix A "Details of loss functions for comparison".

Quantitative results are illustrated in Table II. Compared to WeightLoss1 and WeightLoss2, our methods outperform them on most metrics, indicating the effectiveness of introducing gradient-based prior constraints into small-sample learning. It should be noted that WeightLoss2 [41] has relatively lower scores because it is designed to penalize points far from the surface for creating SDF values close to zero, which is only a biased approximation to SDF and can even lead to a local optimum when used alone. The success of [41] is actually attributed to the combination of WeightLoss2 and global shape constraints (e.g., Eikonal regularization). Besides, T-EikonalLoss and W-GradLoss achieved better performance than their baselines, indicating that filtering out noise signals from zero-gradient areas allows the network to focus on learning the gradient shape near the iso-surface, which contributes to a more compact shape prior. Considering the similar performance between T-EikonalLoss and W-GradLoss, we only adopt W-GradLoss for training in the rest of the article.

D. Comparison With State-of-the-art Methods

In this section, we compare the proposed method against state-of-the-art reconstruction algorithms, including three point-based methods (i.e., PCN [51], GRNet [52], and PoinTr [53]), a

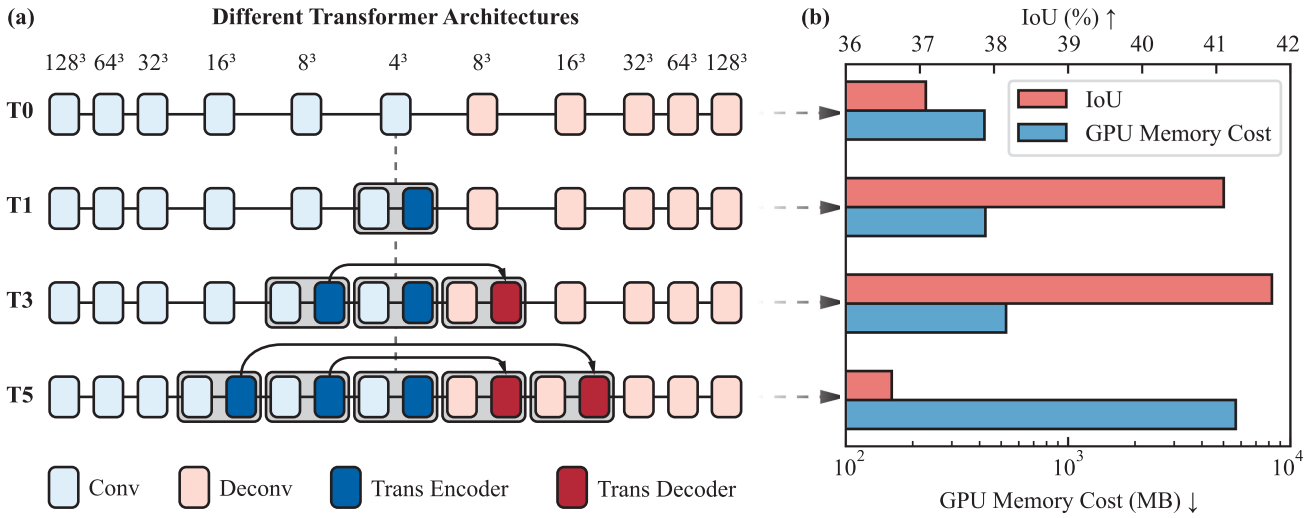

Fig. 6. Comparisons between different architectures incorporated with self-attention blocks. (a) presents four networks with self-attention blocks inserted in different stages. Skip connections between Transformer blocks are displayed and the others are hidden for clarity. (b) is the comparison in terms of design accuracy and GPU memory cost.

GAN-based method (i.e., 3D-RecGAN [54]) and an implicit function-based method (i.e., IF-Net [33]). Based on the encoder-decoder pipeline, PCN [51] proposes to extract geometric features using PointNet [55] and generate the complete point cloud in a coarse-to-fine fashion. GRNet [52] adopts 3D grids as intermediate representations and introduces a differentiable Gridding operation to convert unordered points to regular 3D grids. In this way, efficient 3D CNN can be directly used to process the features. PoinTr [53] aggregates local point features using a lightweight graph CNN and predicts the missing point cloud based on a transformer encoder-decoder architecture. 3D-RecGAN [54] uses adversarial learning to reconstruct the complete object from an arbitrary depth view. In addition, the proposed algorithm is compared with IF-Net [33], which predicts the occupancy field based on multi-scale features queried at continuous positions in 3D feature maps. For a fair comparison, we train all the above networks on our dataset with the same basic settings (i.e., batch size and training epochs). The other hyperparameters remain consistent with the optimal values mentioned in their papers. Besides, considering the small size of our dataset, the transfer learning technique is utilized to accelerate the training. However, to our best knowledge, there are no large datasets built for mesh-to-mesh generation tasks so far. Thus, we only pretrain PCN, GRNet, and PoinTr on a point cloud reconstruction dataset (i.e., the ShapeNet55 dataset) for 150 epochs and finetune them on our RPD clasp design dataset. The pretrained models are denoted as PCN*, GRNet*, and PoinTr*, respectively.

1) Quantitative Results: TranSDFNet is compared with other methods in terms of CD-L_2, EMD, and F-Score. (We do not consider IoU because most of the results generated from the point-based methods are so blurred that cannot be reconstructed as meshes or voxels.) To avoid potential data bias caused by limited data, we employ a four-fold cross-validation as suggested by [56] and report the average performance in Table III. Furthermore, we also perform statistical testing on the cross-validation results. In view of the limited size of our dataset, the Kruskal-Wallis test is adopted, and the results ($p < 0.05$) indicate a significant difference between our algorithm and others. Compared with PoinTr, which achieves the best scores among point-based algorithms, TranSDFNet leads to a 22.57% improvement of F-Score and has significantly lower CD-L_2 and EMD. The accuracies of 3D-RecGAN and IF-Net are relatively close to that of TranSDFNet due to the adversarial learning mechanism and multi-scale feature fusion, respectively, which greatly refine the reconstruction results. Lastly, it is observed that the performances of some pretrained models (i.e., PCN* and PoinTr*) are similar to or worse than their unpretrained versions, implying that the transfer learning scheme does not always seem

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT ALGORITHMS

Methods	CD-L_2 (↓)	EMD (↓)	F-Score (↑)
PCN [51]	0.0609 ± 0.0096	0.2127 ± 0.0202	23.89% ± 8.31%
PCN* [51]	0.0837 ± 0.0154	0.2258 ± 0.0197	20.98% ± 9.92%
GRNet [52]	0.0730 ± 0.0047	0.2477 ± 0.0159	29.50% ± 4.54%
GRNet* [52]	0.0408 ± 0.0041	0.1667 ± 0.0114	51.98% ± 5.12%
PoinTr [53]	0.0258 ± 0.0041	0.1263 ± 0.0119	60.31% ± 7.68%
PoinTr* [53]	0.0499 ± 0.0198	0.1596 ± 0.0284	42.01% ± 14.00%
3D-RecGAN [54]	0.0252 ± 0.0088	0.1193 ± 0.0169	69.74% ± 6.06%
IF-Net [33]	0.0195 ± 0.0026	0.0935 ± 0.0055	73.91% ± 4.50%
TranSDFNet	**0.0122 ± 0.0021**	**0.0713 ± 0.0059**	**82.88% ± 3.43%**

Values are denoted as mean ± standard deviation.
The bold values denote the best results.

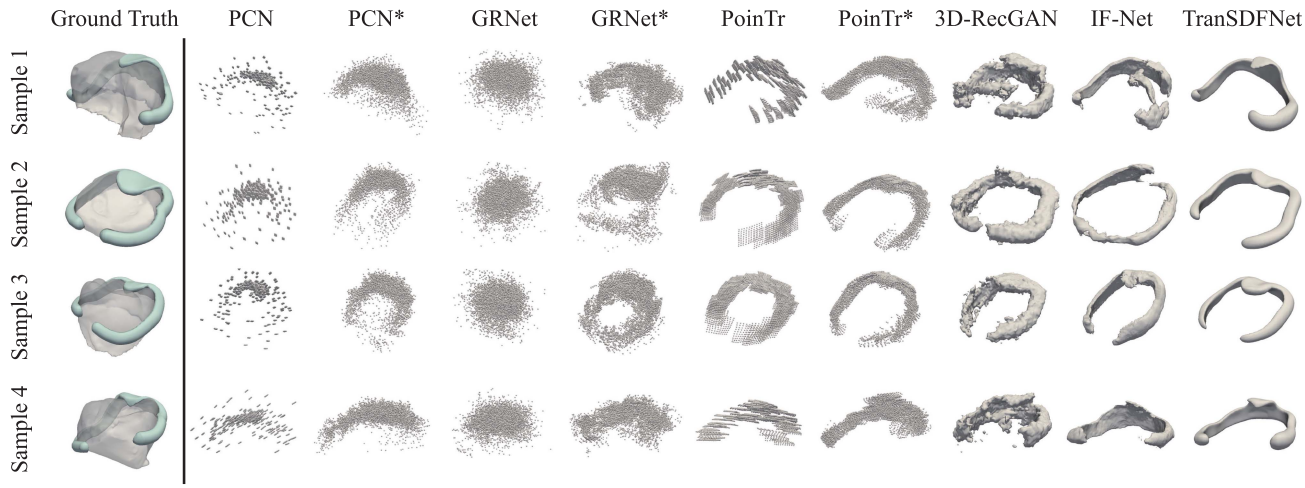

Fig. 7. Qualitative comparison of different algorithms. The first column presents four typical samples with abutments and clasps painted translucent white and light green, respectively. The other columns are claps generated by eight algorithms, where * means the pretrained model.

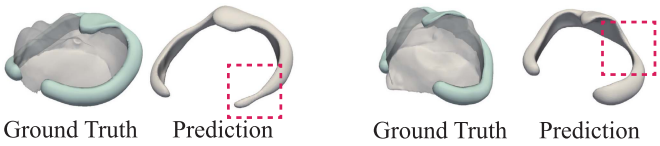

Fig. 8. Bad cases with atrophic distortions in thin structures.

to work. The unintuitive result arises from the huge gap between the source domain (general object completion) and the target domain (RPD clasp design), which hampers the generalization of domain knowledge.

2) Qualitative Results: Fig. 7 provides four examples for the visual comparison. The first column is the ground truth, including abutments (painted translucent white) and corresponding clasps (painted light green). The other columns are results designed by six algorithms. As mentioned above, we do not reconstruct surfaces for the results generated from the point-based methods due to their blurred shapes, which can be observed in Fig. 7. Although the pretraining strategy has alleviated the training problem and improved the fidelity of the shapes to some extent, there still exists a lot of noise signals. The underfitting results generated by 3D-RecGAN also arise from insufficient data. As an implicit shape generation algorithm, IF-Net produces results with smoother surfaces compared to other comparative algorithms. However, observable artifacts are present in the generated thin structures. Among all the six networks, TranSDFNet produces the most realistic RPD clasps, indicating that our method can be well trained even with a limited dataset. The smooth surfaces and accurate topology demonstrate the advantages of implicit representations and the proposed loss functions. It is also worth noting that the poses of synthesized clasps are consistent with the ground truth, which is crucial for the insertion of the denture.

To comprehensively evaluate the TranSDFNet, we perform an analysis of bad cases. As shown in Fig. 8, there are some atrophic distortions in thin structures, e.g., clasp arms (highlighted by red dashed boxes). Although the voxelized TSDF reduces the information loss during the data transformation, it is still a discrete approximation depending on query points. The limited grid points occupied by the thin structures cannot provide enough information to represent and learn shapes effectively. Recent continuous implicit representations [31] may be helpful in solving this question.

IV. Conclusion and Discussion

In this work, we propose a novel end-to-end method for the auto-design of RPD clasps and provide an in-depth theoretical analysis of the critical issues in 3D reconstruction. With implicit fields as intermediate representations, the clasps design is formulated as a regression problem. Concretely, the algorithm takes as input the selective abutment and embeds it into a voxel-based TSDF. Then, a U-shaped architecture incorporated with self-attention mechanisms is built to capture the input features and predicts the TSDF of the target clasp. The Marching Cubes algorithm is finally used to extract the surface from the implicit field. We have evaluated our method on a realistic dataset both quantitatively and qualitatively. The ablation studies and comparisons with the state-of-the-art reconstruction algorithms demonstrate the superiority of our overall architecture and the proposed loss functions. Our study validates the utility and efficacy of deep learning techniques in dentistry. The model can be embedded in AI-driven laboratory design software and contribute to more intelligent and personalized dental treatment.

Despite the outstanding performance of our framework, the ablation studies and the bad case analysis have exposed some limitations, which are summarized as follows:

A. Sensitivity to Resolution

As mentioned in Section III, voxel-based TSDF cannot avoid resolution issues due to its voxelized nature. The global representation leads to a data imbalance, which makes the training and

inference more difficult, especially in those cases containing thin structures. Simply increasing the resolution will not solve this issue completely and will significantly increase time consumption. The additional tests show that the mean time for calculating a TSDF with a resolution of 128^3 can be up to more than 10 seconds, while the network inference only takes 0.4 seconds. To tackle these issues, continuous implicit learning methods that are resolution-independent need to be developed and validated.

1) Limitations of the Experimental Subjects: Due to the lack of available data, we only evaluate our algorithm on the three-arm clasp. Despite its importance for the retention of dentures, the clasp is only one of a variety of RPD components. The extremely complex topology of RPDs and interference constraints in the oral cavity remain major challenges for deep learning. In the future, we would combine the learning-based and rule-based algorithms to develop a more intelligent and complete system for auto-designing RPDs.

Appendix A
Details of Loss Functions for Comparison

With Smooth L1 Loss as the basic item L_{base}, we compare our T-EikonalLoss and W-GradLoss with four other loss functions, which are defined as follows. The hyperparameters are consistent with the optimal values in their papers.

1) WeightLoss1: By assigning higher weights for points whose signed distance is below a certain threshold δ, WeightLoss1 [46] forces the network to concentrate on recovering the details near the zero iso-surface.

$$L_{\text{Weight1}} = \begin{cases} m_1 L_{\text{base}}, & \text{if } |y| < \delta \\ m_2 L_{\text{base}}, & \text{otherwise} \end{cases}, \tag{9}$$

where $m_1 = 4$, $m_2 = 1$, $\delta = 0.01$.

2) WeightLoss2: WeightLoss2 [41] is designed to penalize points outside the surface for creating SDF values close to zero.

$$L_{\text{Weight2}} = L_{\text{base}} + \lambda e^{-\alpha|\hat{y}|}, \tag{10}$$

where $\lambda = 3000$, $\alpha = 100$.

3) EikonalLoss: According to the unit-gradient nature of SDF, EikonalLoss [41], [42] constrains the norm of field gradients to be one.

$$L_{\text{Eikonal}} = L_{\text{base}} + \lambda(\|\nabla\hat{y}\| - 1)^2, \tag{11}$$

where $\lambda = 0.005$.

4) GradLoss: Different from EikonalLoss, GradLoss [43] directly minimize the L_2 distance between the predicted SDF gradients and ground truth gradients. In this way, the magnitude and orientation of the gradients can be simultaneously constrained.

$$L_{\text{Grad}} = L_{\text{base}} + \lambda\|\nabla y - \nabla\hat{y}\|, \tag{12}$$

where $\lambda = 0.005$.

References

[1] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, "A smart dental Health-IoT platform based on intelligent hardware, deep learning, and mobile terminal," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 898–906, Mar. 2020.

[2] J. Zhang, J. J. Xia, J. Li, and X. Zhou, "Reconstruction-based digital dental occlusion of the partially edentulous dentition," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 201–210, Jan. 2017.

[3] M. A. Peres et al., "Oral diseases: A global public health challenge," *Lancet*, vol. 394, no. 10194, pp. 249–260, Jul. 2019.

[4] A. J. Righolt, M. Jevdjevic, W. Marcenes, and S. Listl, "Global-, regional-, and country-level economic impacts of dental diseases in 2015," *J. Dent. Res.*, vol. 97, no. 5, pp. 501–507, May 2018.

[5] N. Kassebaum, E. Bernabé, M. Dahiya, B. Bhandari, C. Murray, and W. Marcenes, "Global burden of severe tooth loss: A systematic review and meta-analysis," *J. Dent. Res.*, vol. 93, no. 7, pp. 20S–28S, Jul. 2014.

[6] N. Kassebaum et al., and GBD 2015 Oral Health Collaborators, "Global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 countries, 1990: A systematic analysis for the global burden of diseases, injuries, and risk factors," *J. Dent. Res.*, vol. 96, no. 4, pp. 380–387, Apr. 2017.

[7] R. G. Watt et al., "Ending the neglect of global oral health: Time for radical action," *Lancet*, vol. 394, no. 10194, pp. 261–272, Jul. 2019.

[8] S. D. Campbell et al., "Removable partial dentures: The clinical need for innovation," *J. Prosthetic Dent.*, vol. 118, no. 3, pp. 273–280, Sep. 2017.

[9] H. Ye et al., "Preliminary clinical application of removable partial denture frameworks fabricated using computer-aided design and rapid prototyping techniques," *Int. J. Prosthodontics*, vol. 30, no. 4, pp. 348–353, Jul. 2017.

[10] C. Bural and O. Geckili, "Direct and indirect retainers," in *Removable Partial Dentures*. Cham, Switzerland: Springer, 2016, pp. 101–122.

[11] Y. Maeda, S. Tsutumi, M. Minoura, M. Okada, T. Nokubi, and Y. Okuno, "An expert system for designing removable partial dentures. preliminary report," *J. Osaka Univ. Dent. Sch.*, vol. 25, pp. 79–84, Dec. 1985.

[12] A. J. Beaumont and H. J. Bianco, "Microcomputer-aided removable partial denture design," *J. Prosthetic Dent.*, vol. 62, no. 4, pp. 417–421, Oct. 1989.

[13] S. Modgil, T. J. Hutton, P. Hammond, and J. C. Davenport, "Combining biometric and symbolic models for customised, automated prosthesis design," *Artif. Intell. Med.*, vol. 25, no. 3, pp. 227–245, Jul. 2002.

[14] Q. Chen, S. Lin, J. Wu, P. Lyu, and Y. Zhou, "Automatic drawing of customized removable partial denture diagrams based on textual design for the clinical decision support system," *J. Oral Sci.*, vol. 62, no. 2, pp. 236–238, Mar. 2020.

[15] D. Ravi et al., "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

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

[17] Z. Cui et al., "TSegNet: An efficient and accurate tooth segmentation network on 3D dental model," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101949.

[18] Z. Cui et al., "A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images," *Nature Commun.*, vol. 13, no. 1, Apr. 2022, Art. no. 2096.

[19] M. Chung et al., "Individual tooth detection and identification from dental panoramic X-ray images via point-wise localization and distance regularization," *Artif. Intell. Med.*, vol. 111, Jan. 2021, Art. no. 101996.

[20] Y. Lang et al., "DLLNet: An attention-based deep learning method for dental landmark localization on high-resolution 3D digital dental models," in *Proc. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2021, pp. 478–487.

[21] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

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

[23] F. Yuan, N. Dai, S. Tian, Z. Bei, and H. Liu, "Personalized design technique for the dental occlusal surface based on conditional generative adversarial networks," *Int. J. Numer. Methods Biomed. Eng.*, vol. 36, no. 5, Feb. 2020, Art. no. e3321.

[24] S. Tian et al., "DCPR-GAN: Dental crown prosthesis restoration using two-stage generative adversarial networks," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 151–160, Jan. 2022.

[25] S. Tian et al., "Efficient computer-aided design of dental inlay restoration: A deep adversarial framework," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2415–2427, Sep. 2021.

[26] O. Lessard, F. Guibault, J. Keren, and F. Cheriet, "Dental restoration using a multi-resolution deep learning approach," in *Proc. IEEE 19th Int. Symp. Biomed. Imag.*, 2022, pp. 1–4.

[27] H. Zhu, X. Jia, C. Zhang, and T. Liu, "ToothCR: A two-stage completion and reconstruction approach on 3D dental model," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2022, pp. 161–172.

[28] A. Dai, C. R. Qi, and M. Nießner, "Shape completion using 3D-Encoder-Predictor CNNs and shape synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6545–6554.
[29] Z. Chen and H. Zhang, "Learning implicit fields for generative shape modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5932–5941.
[30] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3D reconstruction in function space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4455–4465.
[31] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning continuous signed distance functions for shape representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 165–174.
[32] T. Amiranashvili, D. Lüdke, H. Li, B. Menze, and S. Zachow, "Learning shape reconstruction from sparse measurements with neural implicit functions," in *Proc. Conf. Med. Imag. Deep Learn.*, 2022, pp. 1–13.
[33] J. Chibane, T. Alldieck, and G. Pons-Moll, "Implicit functions in feature space for 3D shape reconstruction and completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6968–6979.
[34] Y. Ping, G. Wei, L. Yang, Z. Cui, and W. Wang, "Self-attention implicit function networks for 3D dental data completion," *Comput. Aided Geometric Des.*, vol. 90, Oct. 2021, Art. no. 102026.
[35] A. Vaswani et al., "Attention is all you need," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
[36] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 605–613.
[37] S. Osher and R. Fedkiw, "Signed distance functions," in *Level Set Methods and Dynamic Implicit Surfaces*. New York, NY, USA: Springer, 2003, pp. 17–22.
[38] R. A. Newcombe et al., "KinectFusion: Real-time dense surface mapping and tracking," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2011, pp. 127–136.
[39] P.-S. Wang, Y. Liu, and X. Tong, "Dual octree graph networks for learning adaptive volumetric shape representations," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 103:1–103:15, Jul. 2022.
[40] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, Aug. 1987.
[41] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2020, pp. 7462–7473.
[42] M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotlagh, and A. Eriksson, "Implicit surface representations as layers in neural networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4743–4752.
[43] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman, "Implicit geometric regularization for learning shapes," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 3789–3799.
[44] S. Tian, N. Dai, X. Cheng, L. Li, Y. Sun, and H. Cui, "Relative trajectory-driven virtual dynamic occlusal adjustment for dental restorations," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 59–70, Jan. 2019.
[45] T. Hirotomi, A. Yoshihara, H. Ogawa, and H. Miyazaki, "Tooth-related risk factors for tooth loss in community-dwelling elderly people," *Community Dent. Oral Epidemiol.*, vol. 40, no. 2, pp. 154–163, Apr. 2012.
[46] Q. Xu, W. Wang, D. Ceylan, R. Mech, and U. Neumann, "DISN: Deep implicit surface network for high-quality single-view 3D reconstruction," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2019, pp. 492–502.
[47] M. Tatarchenko, S. R. Richter, R. Ranftl, Z. Li, V. Koltun, and T. Brox, "What do single-view 3D reconstruction networks learn?," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3405–3414.
[48] S. Li et al., "Automatic dental plaque segmentation based on local-to-global features fused self-attention network," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 5, pp. 2240–2251, May 2022.
[49] Y. Li et al., "AGMB-Transformer: Anatomy-guided multi-branch transformer network for automated evaluation of root canal therapy," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 4, pp. 1684–1695, Apr. 2022.
[50] Y. Zhao et al., "Two-stream graph convolutional network for intra-oral scanner image segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 4, pp. 826–835, Apr. 2022.
[51] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "PCN: Point completion network," in *Proc. IEEE Int. Conf. 3D Vis.*, 2018, pp. 728–737.
[52] H. Xie, H. Yao, S. Zhou, J. Mao, S. Zhang, and W. Sun, "GRNet: Gridding residual network for dense point cloud completion," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 365–381.
[53] X. Yu, Y. Rao, Z. Wang, Z. Liu, J. Lu, and J. Zhou, "PoinTr: Diverse point cloud completion with geometry-aware transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 12478–12487.
[54] B. Yang, H. Wen, S. Wang, R. Clark, A. Markham, and N. Trigoni, "3D object reconstruction from a single depth view with adversarial learning," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 679–688.
[55] R. Q. Charles, 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.*, 2017, pp. 652–660.
[56] Z. Su, T. E. Tavolara, G. Carreno-Galeano, S. J. Lee, M. N. Gurcan, and M. K. K. Niazi, "Attention2majority: Weak multiple instance learning for regenerative kidney grading on whole slide images," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102462.