AI4TRT: Automatic Simulation of Teeth Restoration Treatment

Feihong Shen[1] , **Yuer Ye**[2]
[1]Chohotech
[2]Zhejiang University
gregoryfeihong@gmail.com, yye1101@zju.edu.cn

Abstract

Visualizing restoration treatments is a crucial task in dentistry. Traditionally, dentists drag the standard template tooth line onto the inner image from the front view to simulate the outcome of the restoration. This process lacks the precision needed for patient presentation. We find that calculating the camera pose and the relative positions of the upper and lower jaws can enhance visualization accuracy and efficiency while assisting dentists in treatment design. In this work, we leverage the optical flow model and a customized point renderer to help dentists show the treatment outcome to the patient. Specifically, we take the 3D scan model and the intraoral image pair as input. Our framework automatically outputs the camera pose and the relative position of the upper and lower jaws. With these parameters, dentists can directly design the restoration treatment on the 3D scan model without caring about the 2D visualization. Then the designed tooth line and other simulation modalities can be rendered on the intraoral image with our customized renderer. Our framework relieves the labor of dentists and shows the case precisely.

1 Introduction

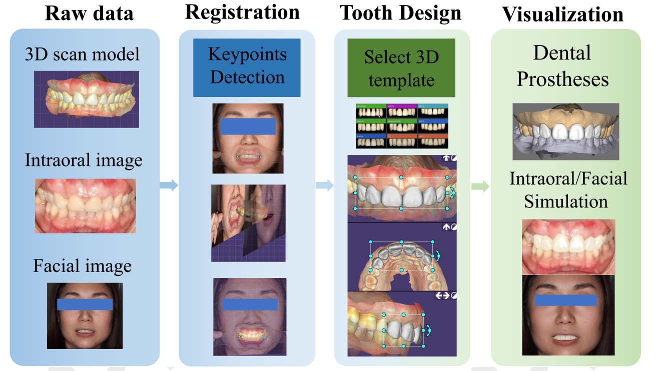

Figure 1: The ideal process of teeth restoration simulation and design with the automatic software product.

A harmonious smile, which is closely linked to dental aesthetics, not only boosts a patient's self-confidence but also leaves a lasting positive impression in social interactions [Gavic *et al.*, 2024]. Orthodontic treatment effectively addresses aesthetic issues caused by crowded or misaligned anterior teeth. However, the treatment period typically spans several months to years, and the long-term success of the treatment largely depends on patient compliance. Restorative treatments address a broader range of issues, including tooth loss and abnormalities in tooth shape or color caused by caries, wear, or crown fractures[Zaborowicz *et al.*, 2024]. Dentists can customize treatments based on individual facial features, preferences, and functional needs. Because of its shorter duration and straightforward process, restorative treatment is ideal for patients with high aesthetic demands who wish to promptly enhance the appearance of their anterior teeth[Li *et al.*, 2022].

The rapid advancement of digital dental technology is driving significant progress in aesthetic restorative treatments. Digital Smile Design (DSD) technology is a key method that employs software to analyze, design, and visualize cases before treatment. This technology has enhanced the predictability and quantification of treatment, improving communication efficiency between clinicians and patients[Jafri *et al.*, 2020]. Its workflow includes several steps: acquisition of facial and oral images or three-dimensional (3D) data, aesthetic analysis, and design[Omar and Duarte, 2018]. However, despite advancements in 3D dental design software, data registration, reference line drawing, and restoration design remain largely manual. During data registration, the operator must align the data multiple times. While the software can only perform an initial rough alignment based on an operator-selected anatomic feature point, usually using the intraoral image as a reference, the dentist must still manually adjust the 3D dentition and facial photograph to achieve visual alignment. The operational intelligence of the software affects both the efficiency of clinical work and the broader adoption of its clinical use. Furthermore, manual errors can impact the accuracy of both registration and design data[Alharkan, 2024].

We propose our ideal process in Fig. 1. The alignment of the facial image and the intraoral image can be achieved through keypoints. So once the software calculates the alignment parameters of the 3D scan model and the intraoral image, the dentists can design the restoration treatment on the

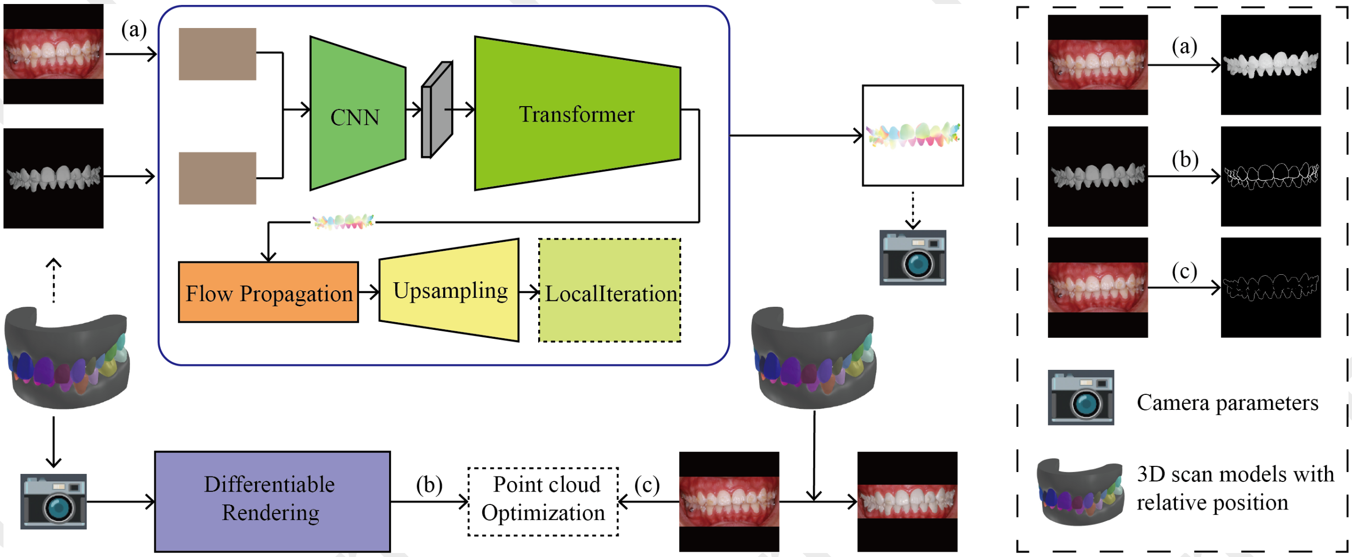

Figure 2: **Overview of our framework. (a)** The segmentation and gray-scale of teeth in the intraoral photograph. **(b)** The detection of teeth edge in the rendered teeth image. **(c)** The detection of teeth edge in the intraoral photograph. The detection is based on the result in **(a)**. The dash line and dash line box in the framework denotes that the process contains multiple iterations. In Step 1, the optical flow model takes the gray-scale image and the rendered image as input and outputs the optical flow image. The 3D scan models with relative position are rendered with the camera parameters to update the rendered image every iteration. In Step 2, the Point Renderer takes the coarse result in the Step 1 and outputs the NDC coordinate of points on the edge of 3D scan model. In the end, the restoration plan model is rendered on intraoral photograph based on the optimized camera pose.

3D scan model. With the rendering speed-up devices like GPU, the dentists are able to visualize the real-time change on the facial image along the design process. While the traditional software programs require the dentists to pull the 3D scan model on the intraoral image manually.

Automatic registration is the key component for achieving this ideal process. Most of the teeth are obscured due to the limited perspectives of intraoral images. And we didn't know the camera focal length from different data source. To overcome these problems, in this paper, we propose a pipeline to alignment the 3D scan teeth model and 2D intraoral image. Our pipeline relies on two essential models: the optical flow model and the point renderer. We train our optical flow model on virtual data rendered by the 3D scan model with augmentation. After the training, the optical flow model is applied to get the coarse result of the registration. To refine the parameters, we designed a point renderer that optimizes the coarse results using Chamfer distance in the Normalized Device Coordinate (NDC) system.

The main contributions of our method can be summarized as follows: **1)** We propose a novel framework that eliminates the laborious task of 3D and 2D data registration for dentists and circumvents errors inherent in manual processes. Our framework further enables dentists to visualize outcomes directly on facial images during the design phase. **2)** The framework is extensible to other dental applications, such as occlusion reconstruction and teeth alignment visualization. **3)** Experiments on real-world data demonstrate that our framework successfully aligns intraoral images with 3D scan models, even when the latter lack visual texture features.

2 Related Work

2.1 Optical Flow

Optical flow has traditionally relied on traditional methods[Sun *et al.*, 2010], such as the Horn-Schunck algorithm[Horn and Schunck, 1981], to solve the energy minimization problem. Until 2015, FlowNet[Dosovitskiy *et al.*, 2015] introduced neural network to this field and enabled the optical flow motion estimation by the data-driven approaches. FlowNet directly take two images as input and output an optical flow image. Then, FlowNet 2.0[Ilg *et al.*, 2017] proposed new network architecture and advanced training strategies to improve accuracy. After that, some following researchers import transformer-based approaches[Huang *et al.*, 2022] and refinement inference[Cheng *et al.*, 2024; Xu *et al.*, 2023] into the optical flow model to solve the problem that the optical flow model is insensitive to the small movement[Jung *et al.*, 2023]. But there still exist some gaps between the synthetic training data and real data in the application. In this work, we apply the optical flow model on the intraoral photographs. Given a large amount of synthetic training data, our model shows robust performance on edge cases. But we still face the problem that the rendered images are not completely align with the intraoral images. Unlike previous work, we design a customized point renderer to overcome the backwards.

2.2 Differentiable Rendering

Differentiable rendering bridge the domains of computer vision and computer graphic[Kato *et al.*, 2020]. Recent work about differentiable rendering[Durvasula *et al.*, 2023;

Gao and Qi, 2024] focus on improving efficiency and applicability to real-world tasks like our framework. In this section, we mainly focus on the optimization tasks figured out by differentiable rendering. The optimization algorithm based on differentiable rendering is widely used in the field of machine learning. For example, neural rendering[Mildenhall *et al.*, 2021] and implicit representations[Vicini *et al.*, 2022] of 3D scene leverage the differentiable rendering to optimize the neural network and Signed Distance Function (SDF). Besides, 3D Gaussian Splatting (3DGS)[Chen and Wang, 2024] also apply differentiable rendering to optimize the 3D Gaussian-based representation. In the field of face reconstruction[Deng *et al.*, 2019], differentiable rendering is utilized to optimize the express, identity and texture parameters to fit the human face in the images. Recently, differentiable rendering is used to achieve tasks like exoplanet detection[Feng *et al.*, 2025], digital human reconstruction[Wang and Li, 2023], path-guiding[Fan *et al.*, 2024] and text-to-3D sketch[Zhang *et al.*, 2024]. Most of them use gradient-based optimization within the rendering pipeline. In this paper, we utilize differentiable rendering to optimize the pose of the camera and the relative position of the lower jaw to the upper jaw.

3 Methodology

We outline the methodology of our framework as follows. First, we formalize the problem definition, then detail the framework's core components. Our approach comprises two stages: a coarse stage, where parameters are derived via an optical flow model, and a refinement stage, which optimizes the initial output using a customized point renderer. Finally, we explore potential downstream applications. The full pipeline is illustrated in Fig. 2.

3.1 Problem Formulation

Our simulation framework for tooth restoration treatment focuses on aligning a 2D intraoral image $\mathcal{I}$ with a 3D scan model $\mathcal{T}$ to enable virtual replacement of the initial scan with a clinician-designed treatment plan. To mitigate the noise bringing by the gum, we preprocess the 3D scan using a mesh segmentation network to isolate individual tooth crowns $\{t_i\}_{i=1}^N$, where N is the tooth count. While a pretrained 2D network segments tooth regions S of teeth in intraoral image I. The core contribution lies in robustly estimating two spatial parameters: the camera pose P_{camera} and the relative jaw position P_{rela}, which jointly enable accurate 3D-2D alignment for projecting the treatment plan into the intraoral view. In this paper, we abstract away medical rules and segmentation network architectures, prioritizing the optimization of P_{camera} and P_{rela} to establish a modular pipeline that integrates AI-driven anatomical filtering with physics-based alignment for clinical restoration simulation.

3.2 Optical Flow Matching

We first elucidate the training of our optical flow model. We can observe that the difference between the rendering result of the tooth crown set $\{t_i\}_{i=1}^N$ and the grayscale image of the teeth part $S \cdot I$ is trivial. As shown in Fig. 3, this similarity gives us the probability to train the optical flow network using the virtual data. After training the network using virtual data, we can directly use the trained network to predict the optical flow of the real data. The virtual training data is rendered by the preprocessed tooth crowns $\{t_i\}_{i=1}^N$. One pose matrix M_{init} and one transformation matrix M_{tran} are randomly initiated to render two images of $\{t_i\}_{i=1}^N$ from different viewpoints. In the rendering process, we can get the projection from the pixels to the faces of $\{t_i\}_{i=1}^N$ and the projection from faces to the pixels. Based on these projections from the initial and transformed poses, we can calculate the ground truth of the flow image.

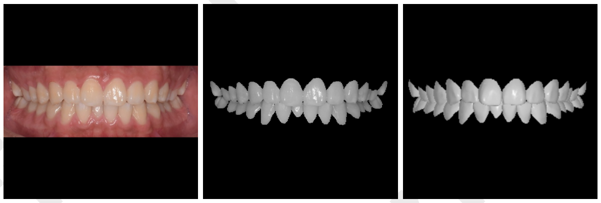

Figure 3: Similarity between the normalized gray scale teeth image and our rendered image. We leverage the 2D segmentation network to preprocess the intraoral photograph.

$$\mathcal{I}_0, pix_0, face_0 = R\left(M_{init}, \{t_i\}_{i=1}^N\right) \tag{1a}$$

$$\mathcal{I}_1, pix_1, face_1 = R\left(M_{tran} \cdot M_{init}, \{t_i\}_{i=1}^N\right) \tag{1b}$$

$$flow = Sample\left(face_0 - face_1, pix_0\right), \tag{1c}$$

where $\mathcal{I}_i, pix_i, face_i, R$ denotes the rendered image, the projection from pixel to face, the projection from face to pixel and our renderer. Here, $face_0 - face_1$ quantifies the change between corresponding faces at identical pixel locations, while the $Sample$ function selectively retains differences based on non-negative pix_0 values. Thus, we get the one label flow image and two input rendered images. In the training stage, the convolutional network in the model extracts the feature from two input images, while the transformer in the model helps to matching feature extracted from these two images. After propagation by self-attention and post-process of matching result, we get the final predicted flow image which makes L1 Loss $= |y - \hat{y}|$ with the target flow image.

During the inference phase, the model processes two inputs: 1) a processed grayscale image $S \cdot I$, which isolates the teeth region, and 2) an initial rendered image I_0. Initial estimates for both the camera pose and relative position are suboptimal. In one iteration, the camera pose is updated using a transformation matrix derived from the model's output flow map. Once the camera pose is refined, it is fixed alongside the upper jaw's position. Subsequently, the optical flow model generates a new flow map based on the updated image rendered by the refined camera pose. This subsequent flow map then guides adjustments to the lower jaw pose, thereby completing one full optimization cycle. The transformation of the lower jaw is achieved by separating a virtual camera. The transformation of this virtual camera is transferred into the movement of the lower jaw object in the scene. Through successive iterations, the framework converges to a final prediction of the camera pose $\hat{P}_{\text{camera}}$ and the relative jaw position $\hat{P}_{\text{rela}}$, achieving alignment by leveraging both geometric and image-based cues.

As shown in Fig. 2, our optical flow model includes one convolutional network (CNN), one transformer, one module for flow propagation, one module for upsampling and one module for refinement. There are self-attention and cross-attention mechanisms in transformer to refine the feature extracted from the CNN and build the long-range dependencies between the two input images. The module of flow propagation refines the entire feature map to the flow map by self-attention mechanism. The upsampling module upsamples the flow map to the original image resolution by a convolutional layer. After upsampling, we apply refinement on the flow map to improve the estimates. This refinement is achieved by iteratively passing the flow map through the LocalIteration module, which utilizes a U-Net architecture to improve the final output. The U-Net and CNN structures follow the model in the previous work [Xu *et al.*, 2023].

3.3 Point Renderer

In the optical flow matching stage, we obtain the coarse estimates for camera poses $\hat{P}_{\text{camera}}$ and the relative position of the upper/lower jaw $\hat{P}_{\text{rela}}$. However, a misalignment persists between the rendering result and the real intraoral image. To refine accuracy, we optimize these coarse estimates using differentiable rendering. Specifically, we design a point renderer R_{pt} that generates an edge map of input tooth crowns and the projection from pixels to faces. In the experiment, we find that the visual difference between rendered teeth and target intraoral photograph makes the pixel loss unsuitable for the optimization. Besides, the feature exhibited by the rendered edge map is too sparse to find the correct optimize path and is likely to fall into the suboptimal result. Thus, our R_{pt} can sample the coordinates of points rendered on the edge of teeth from its outputs.

In the optimization process, Parameters P_{camera} and P_{rela} are initialized using the coarse results $\hat{P}_{\text{camera}}$ and $\hat{P}_{\text{rela}}$. The rasterizer in R_{pt} transforms the mesh using these parameters and stores the transformed mesh's faces/vertices as intermediate values. Vertices are partitioned into per-tooth batches to preserve the occlusion relationships during parallel rasterization. The rasterized result is a multi-dimensional array with size $N \times H \times W$, while N is the number of teeth and W, H is the width and height of the intraoral image. Every $H \times W$ slice represents one rasterized tooth. Non-zero elements in one slice denote the rasterized face indices in that corresponding pixel. The values of the pixels without face will be set as -1. We set non-zero elements with the value 1 and get contours of rasterized teeth by applying convolution with our customized kernel. The customized kernel is a 3×3 matrix where the center value is 1, and the surrounding eight values are all -1/8. We obtain the face indexes of the tooth crowns which local on the edge of rendering teeth. Then we get the NDC positions of the center of these faces by sampling the face indexes in mediate value. A pre-trained segmentation network extracts tooth contours from the intraoral image. The pixels of predicted tooth contours are projected to the NDC space to become the optimization target. Since the whole rendering pipeline is differentiable, we can iteratively update P_{camera} and P_{rela} through the optimization algorithm like stochastic gradient descent, guided by the Chamfer distance between target points and sampled edge points. The optimization pipeline can be formalized as:

$$\text{E}_{\text{pred}}, \text{pix2face} = R_{\text{pt}} \left(P_{\text{camera}}, P_{\text{rela}}, \{t_i\}_{i=1}^{N} \right) \tag{2a}$$

$$Point_{\text{pred}} = \text{Sample} \left(\text{E}_{\text{pred}}, \text{pix2face}, \{t_i\}_{i=1}^{N} \right) \tag{2b}$$

$$\mathcal{L} = \text{Chamfer} \left(Point_{\text{pred}}, Point_{\text{gt}} \right), \tag{2c}$$

where $Point_{\text{gt}}$ is the NDC-transformed point cloud of intraoral image edges. We choose Chamfer distance as the loss function because the number of sampled edge points change every iteration, while Chamfer distance can measure the distance between different points set and keep regardless about exact number of points in the set.

3.4 Discussion

In Fig. 4, we list the visualization of teeth restoration treatment. In this paper, we will not discuss how the dentists design treatments for patients. The *de facto* core of our framework resides in the registration of the 3D scan model on the intraoral image. And this part make our method not limited to the application of teeth restoration visualization. In dentistry, some dentists are lack of 3D scanner to directly scan the patients' teeth. Plaster is the material they commonly use for taking dental impressions. However, plaster cannot capture the relative position of the upper and lower jaws, which is really essential information for dentists to design the treatment. In this way, our framework can help to get the correct relative position with the support from the intraoral photograph. What's more, in Fig. 4 we can observe that the 3D model fit the image in the pixel level. Since our point renderer can generate the projection from faces to pixels in the image, we can colorize the 3D scan model with the intraoral photograph. The colorized 3D scan model can be utilized on application like the visualization of alignment treatment. We give a simple example in Fig. 5

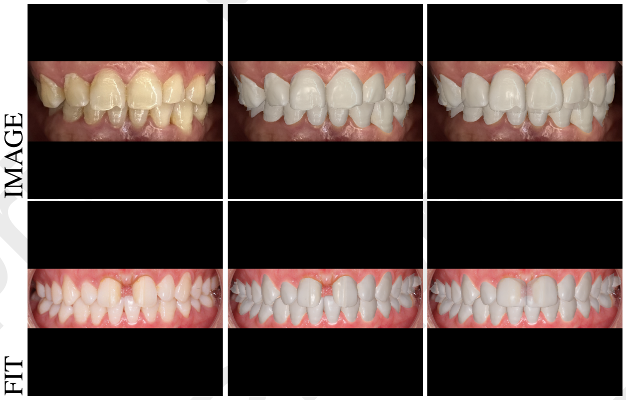

Figure 4: Visualization of restoration treatment using our automatic technology.

Method	Rotation			Translation			
	α	β	γ	$x - axis$	$y - axis$	$z - axis$	Accurate
SVD	2.809	0.911	1.478	0.687	1.926	1.111	80
DGCNN	4.452	3.264	4.212	0.493	0.822	0.705	26
VI-Net	2.742	0.918	2.121	0.717	1.304	1.085	30
Megapose	3.877	1.842	3.765	0.987	1.858	1.908	20
Ours	**0.887**	**0.334**	**0.984**	**0.348**	**0.711**	**0.655**	**295**

Table 1: Quantitative results on our upper and lower jaw relative position estimation experiment

Iteration	Point Renderer	Rotation			Translation		
		α	β	γ	$x - axis$	$y - axis$	$z - axis$
0	✓	1.848	0.638	2.294	0.823	1.345	0.901
1	×	1.175	0.512	1.308	0.622	0.861	0.682
2	×	1.006	0.506	1.291	0.629	0.851	0.661
3	×	0.972	0.481	1.148	0.587	0.827	0.639

Table 2: Ablation studies on the influence of the optical flow and differentiable rendering module.

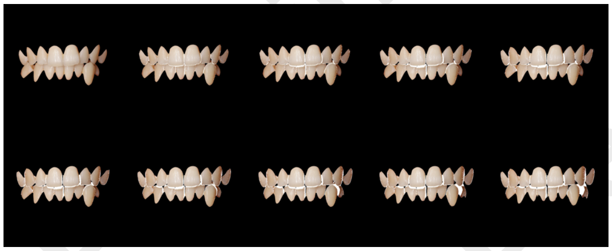

Figure 5: The visualization of alignment treatment using the registration result generated by our framework. The white part indicates the movement of the teeth in alignment.

4 Experiment

4.1 Experimental Setup

Datasets. We get two modal data from our collaborating medical institutions after obtaining the patient's informed consent: digital 3D scan models and intraoral photographs. These data provide the source information for our train and test dataset. As for the dataset, 1600 cases were collected to train our network. The test dataset contains restoration treatment plans from 400 cases. The restoration treatment plan is represented by another mesh file. Each case have one intraoral image from the front view and the image is resized to a resolution of 1000 × 1600 in preprocessing. As we mentioned before, the 3D scan models are segmented by a pretrained network to the teeth instance. After the segmentation, we create a virtual gum to rebuild its occlusion to the tooth roots. This is because the shape of the gum change in the intersection region after the restoration treatment and there is a large amount of redundant faces in the initial gum. The face of virtual gum is set to invisible after rasterizing.

Baselines. To validate the effectiveness of our proposed framework, we compared with three types of baselines: 1) *Traditional algorithm* We set a pair of template jaws to align the input mesh with the singular value decomposition (SVD)[Wall *et al.*, 2003]. 2) *Pure 3D Neural Network* We trained a DGCNN model [Phan *et al.*, 2018] with our dataset to predict the matrix. 3) *Multimedia Neural Network* We trained a VI-Net [Lin *et al.*, 2023] combined the vision feature from the image. 3) *Zero-shot Neural Network* We apply pose estimate models pretrained on large scale dataset like MegaPose[Labbé *et al.*, 2022] on our teeth dataset.

Evaluation Metrics. Since the ground truth in dataset does not contain the camera pose information. We test the accuracy of models base on the deviation of upper and lower relative position. We divide the position matrix to the deviation of Euler angles α, β and γ and translation on x, y, z axis. The relative position is considered accurate when the sum of translation errors in three axes are smaller than $2mm$ and the sum of rotation errors are smaller than 3°.

Experimental Settings. The optical flow model in our framework, DGCNN, and VI-Net are not trained on the tested categories. The Megapose model loads the pretrained checkpoints with the RGB dimension. In our experiment, our framework directly output the relative transformation matrix. Every compared model generates two matrixes which denotes the transformation of the upper jaw and lower jaw. The relative transformation matrix is calculated by multiplying the transformation matrix of lower jaw with the inverse matrix of upper jaw.

4.2 Implementation Details

The optimization of our framework comprises two stages, including the optical flow model and fine-tuned process. In the first stage, the transformer in our optical flow network contains six transformer blocks for both self-attention and cross-attention mechanisms. The flow propagation module is implementation by a self-attention block across the feature map. Both upsampling module and LocalIteration module are learnable. The LocalIteration module and our optical flow model are both configured to execute 3 iterations. In the training stage, the optical flow network and other baselines are optimized by Adam optimizer with the learning rate started from $1e^{-2}$ and was trained 30,000 steps with a batch size of 32. In the inference stage, the optical flow network repeat the forward function and update the rendered image three times. In the second stage, the parameter of camera is optimized with fixed 200 iterations by Adam with initial 0.5 learning rate. The rendering size is 512 × 512. The 2D segmentation network is developed with DeepLabv3 [Yurtkulu *et al.*, 2019] and the 3D segmentation network is constructed via the previous work [Zheng *et al.*, 2022]. Currently, the rendering pipeline is constructed based on Pytorch3D [Ravi *et al.*, 2020] and Pytorch [Paszke *et al.*, 2019]. All experiments run on a server with eight Nvidia 3090 GPUs, an AMD EPYC 7402 24-core processor, and 252 GB RAM.

4.3 Comparison Results

The comparative results on our dental dataset are presented in Table 1. Notably, the zero-shot model Megapose struggles to infer the correct spatial relationship between the upper and lower jaws. This limitation likely stems from two factors: (1) the occlusion of most teeth in intraoral images, which obscures critical geometric cues, and (2) the absence of precise focal length or depth map inputs, leaving the model prone to erroneous pose estimations. Meanwhile, traditional algorithms and DGCNN exclusively rely on 3D input data. However, their performance is constrained by the inherent ambiguity in intraoral scenes—while most photographs depict occluded biting positions, multiple plausible jaw alignments exist, leading to suboptimal outcomes for these methods. While in our framework, focal length is also optimized with a weak range prior and the depth information is implicitly constrained by chamfer distance in the optimization process.

VI-Net demonstrates improved accuracy by incorporating 2D image data, though it occasionally fails to resolve fine-grained positional relationships. In contrast, our framework achieves superior performance across nearly all evaluation metrics. This advantage derives from our direct supervision strategy: the predicted camera pose and jaw positioning are explicitly constrained by aligning with the tooth contours visible in intraoral photographs, ensuring anatomically consistent results.

4.4 Ablation Study

Effect of Optical Flow model. To test the effect of our optical flow model, we evaluate the accuracy of the relative position estimation across successive iterations. As illustrated in Table 2, when refinement via the point renderer is removed, the overall accuracy improves as the number of iteration steps increases. This is rational because we cannot get the correct camera poses under the wrong relative position. With better camera poses, we can get better relative position. In the first row of the table, refinement by our point renderer is performed without leveraging the coarse initialization from the result of the optical flow model. These results demonstrate that our gradient-based optimization requires a robust initial pose estimate from the optical flow model to achieve convergence.

Effect of Point Renderer. To test the effect of our point renderer, we show the teeth edges in Fig. 6. The green one is the edge detected in the intraoral photograph. The purple one is the edge rendered by the 3D teeth in the position predict by the optical flow model. The red one is the edge rendered by the 3D teeth in the position optimized by the point renderer. It should be noticed that the optimization is on the basis of the coarse result. Compare to the coarse result from the optical flow model. The refinement result after the optimization is closer to the ground truth. A comparative analysis of the final rows in Tables 2 and 1 reveals that the point renderer refinement significantly improves positional accuracy, achieving closer alignment with ground truth values across all three Euler angles and translational axes.

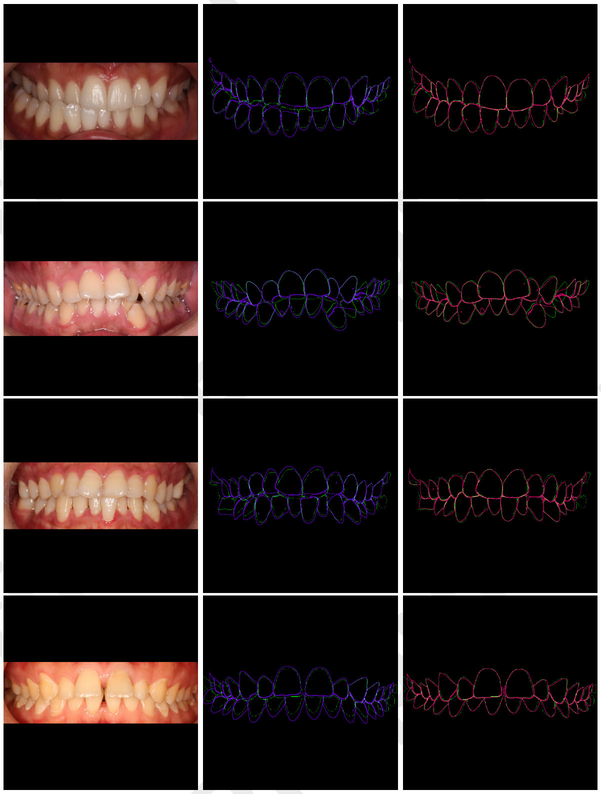

Figure 6: Ablation studies on the influence of the iterative optimization using differentiable rendering. The green line is the detection result from the image, while the purple line is the result of the optical flow model and the red line is the result of differentiable rendering.

5 Conclusion

We propose a novel framework with one data-driven optical flow model and one customized point renderer. It can visualize the restoration treatment designed by the dentists. Our framework efficiently leverages the similarity between the virtual data and the real intraoral photographs to achieve robust estimation capabilities. The refinement stage established by point renderer in our framework make the result more accurate. The pose estimation component in our framework can contribute to lots of potential application in dentistry.

References

[Alharkan, 2024] Hamad M Alharkan. Integrating digital smile design into restorative dentistry: A narrative review of the applications and benefits. *The Saudi Dental Journal*, 36(4):561–567, 2024.

[Chen and Wang, 2024] Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *arXiv preprint arXiv:2401.03890*, 2024.

[Cheng *et al.*, 2024] Ri Cheng, Ruian He, Xuhao Jiang, Shili Zhou, Weimin Tan, and Bo Yan. Context-aware iteration policy network for efficient optical flow estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 1299–1307, 2024.

[Deng *et al.*, 2019] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *IEEE Computer Vision and Pattern Recognition Workshops*, 2019.

[Dosovitskiy *et al.*, 2015] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2758–2766, 2015.

[Durvasula *et al.*, 2023] Sankeerth Durvasula, Adrian Zhao, Fan Chen, Ruofan Liang, Pawan Kumar Sanjaya, and Nandita Vijaykumar. Distwar: Fast differentiable rendering on raster-based rendering pipelines. *arXiv preprint arXiv:2401.05345*, 2023.

[Fan *et al.*, 2024] Zhimin Fan, Pengcheng Shi, Mufan Guo, Ruoyu Fu, Yanwen Guo, and Jie Guo. Conditional mixture path guiding for differentiable rendering. *ACM Transactions on Graphics (TOG)*, 43(4):1–11, 2024.

[Feng *et al.*, 2025] Brandon Y Feng, Rodrigo Ferrer-Chávez, Aviad Levis, Jason J Wang, Katherine L Bouman, and William T Freeman. Exoplanet detection via differentiable rendering. *arXiv preprint arXiv:2501.01912*, 2025.

[Gao and Qi, 2024] Ruicheng Gao and Yue Qi. A brief review on differentiable rendering: Recent advances and challenges. *Electronics*, 13(17):3546, 2024.

[Gavic *et al.*, 2024] Lidia Gavic, Mihaela Budimir, and Antonija Tadin. The association between self-esteem and aesthetic component of smile among adolescents. *Progress in Orthodontics*, 25(1):9, 2024.

[Horn and Schunck, 1981] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.

[Huang *et al.*, 2022] Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer: A transformer architecture for optical flow. In *European conference on computer vision*, pages 668–685. Springer, 2022.

[Ilg *et al.*, 2017] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2462–2470, 2017.

[Jafri *et al.*, 2020] Zeba Jafri, Nafis Ahmad, Madhuri Sawai, Nishat Sultan, and Ashu Bhardwaj. Digital smile design-an innovative tool in aesthetic dentistry. *Journal of oral biology and craniofacial research*, 10(2):194–198, 2020.

[Jung *et al.*, 2023] Hyunyoung Jung, Zhuo Hui, Lei Luo, Haitao Yang, Feng Liu, Sungjoo Yoo, Rakesh Ranjan, and Denis Demandolx. Anyflow: Arbitrary scale optical flow with implicit neural representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5455–5465, 2023.

[Kato *et al.*, 2020] Hiroharu Kato, Deniz Beker, Mihai Morariu, Takahiro Ando, Toru Matsuoka, Wadim Kehl, and Adrien Gaidon. Differentiable rendering: A survey. *arXiv preprint arXiv:2006.12057*, 2020.

[Labbé *et al.*, 2022] Yann Labbé, Lucas Manuelli, Arsalan Mousavian, Stephen Tyree, Stan Birchfield, Jonathan Tremblay, Justin Carpentier, Mathieu Aubry, Dieter Fox, and Josef Sivic. Megapose: 6d pose estimation of novel objects via render & compare. *arXiv preprint arXiv:2212.06870*, 2022.

[Li *et al.*, 2022] Yiyuan Li, Yiqiang Yu, Yue Feng, and Weicai Liu. Predictable digital restorative workflow for minimally invasive esthetic rehabilitation utilizing a virtual patient model with global diagnosis principle. *Journal of Esthetic and Restorative Dentistry*, 34(5):769–775, 2022.

[Lin *et al.*, 2023] Jiehong Lin, Zewei Wei, Yabin Zhang, and Kui Jia. Vi-net: Boosting category-level 6d object pose estimation via learning decoupled rotations on the spherical representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14001–14011, 2023.

[Mildenhall *et al.*, 2021] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[Omar and Duarte, 2018] Doya Omar and Carolina Duarte. The application of parameters for comprehensive smile esthetics by digital smile design programs: A review of literature. *The Saudi dental journal*, 30(1):7–12, 2018.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Phan *et al.*, 2018] Anh Viet Phan, Minh Le Nguyen, Yen Lam Hoang Nguyen, and Lam Thu Bui. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural Networks*, 108:533–543, 2018.

[Ravi *et al.*, 2020] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020.

[Sun *et al.*, 2010] Deqing Sun, Stefan Roth, and Michael J Black. Secrets of optical flow estimation and their principles. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 2432–2439. IEEE, 2010.

[Vicini *et al.*, 2022] Delio Vicini, Sébastien Speierer, and Wenzel Jakob. Differentiable signed distance function rendering. *ACM Transactions on Graphics (TOG)*, 41(4):1–18, 2022.

[Wall *et al.*, 2003] Michael E Wall, Andreas Rechtsteiner, and Luis M Rocha. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer, 2003.

[Wang and Li, 2023] Yi Wang and Yun Li. Differentiable rendering approach to mesh optimization for digital human reconstruction. In *2023 28th International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE, 2023.

[Xu *et al.*, 2023] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofighi, Fisher Yu, Dacheng Tao, and Andreas Geiger. Unifying flow, stereo and depth estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Yurtkulu *et al.*, 2019] Salih Can Yurtkulu, Yusuf Hüseyin Şahin, and Gozde Unal. Semantic segmentation with extended deeplabv3 architecture. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2019.

[Zaborowicz *et al.*, 2024] Katarzyna Zaborowicz, Marcel Firlej, Ewa Firlej, Maciej Zaborowicz, Kamil Bystrzycki, and Barbara Biedziak. Use of computer digital techniques and modern materials in dental technology in restoration: A caries-damaged smile in a teenage patient. *Journal of Clinical Medicine*, 13(18):5353, 2024.

[Zhang *et al.*, 2024] Yibo Zhang, Lihong Wang, Changqing Zou, Tieru Wu, and Rui Ma. Diff3ds: Generating view-consistent 3d sketch via differentiable curve rendering. *arXiv preprint arXiv:2405.15305*, 2024.

[Zheng *et al.*, 2022] Youyi Zheng, Beijia Chen, Yuefan Shen, and Kaidi Shen. Teethgnn: semantic 3d teeth segmentation with graph neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 29(7):3158–3168, 2022.