Contents

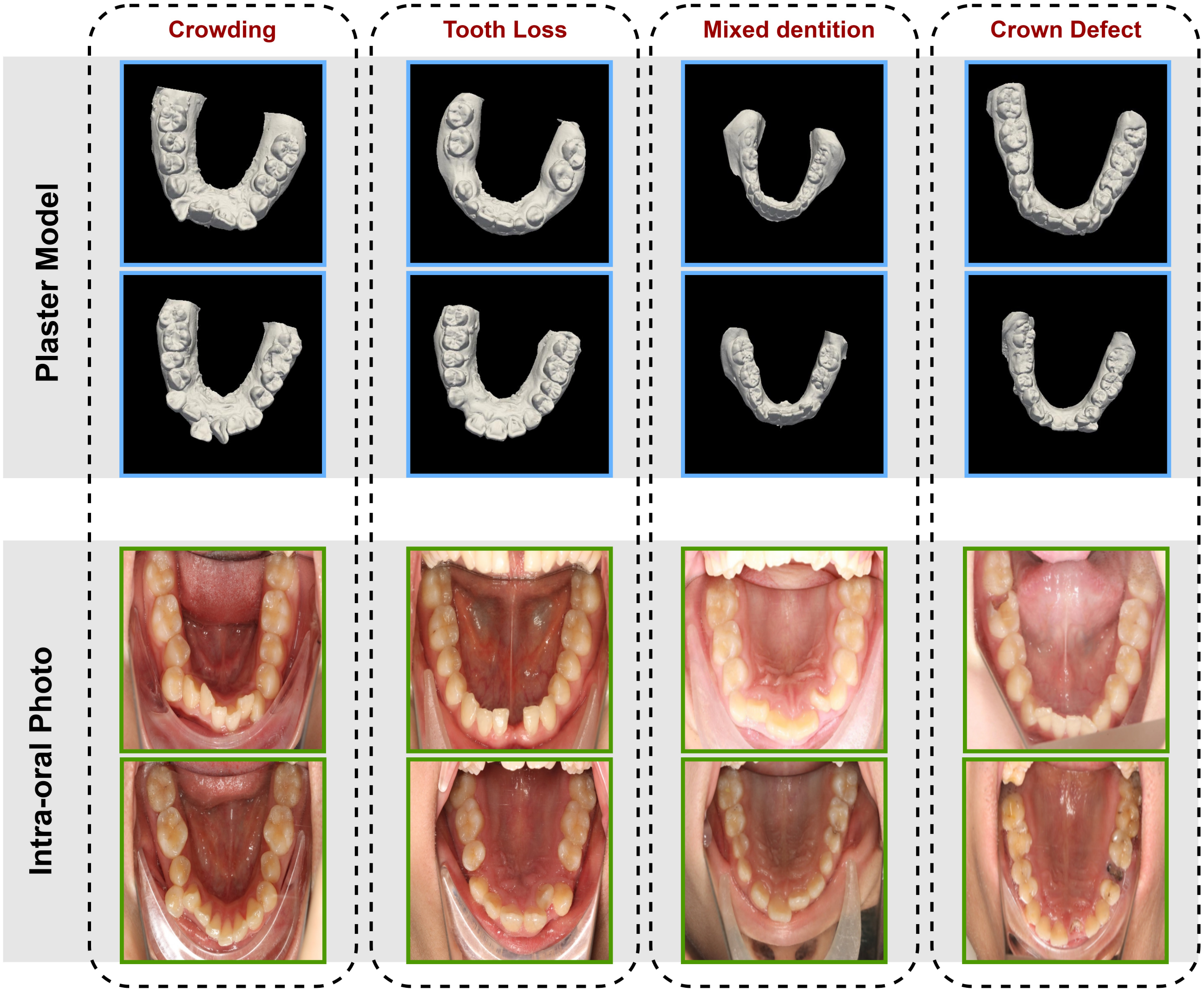

Figure 1. Visualization of Plaster70K and RGB0.8K

Name	Split			dental malformations			
	Train	Val	Test	Crowding	Tooth Loss	Mixed dentition	Crown Defect
Challenge80K	60k	10k	10k	-	-	-	-
Plaster70K	50k	10k	10k	44.1%	19.4%	4.5%	4.6%
RGB0.8K	0.6k	0.1k	0.1k	88.8%	8.4%	14.3%	2.5%

Table 1. Dataset statistics over the three subsets of IO150K.

A. Dataset Statistics

We summarize the dataset statistics over the three subsets of IO150K in Table 1 (we omit the statistic of dental malformations for Challenge80K, which is generated from previous open-sourced [1]) and visualize some examples from our collected Plaster70K and RGB0.8K in Figure 1.

Method	Epoch	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	mIoU
DeepLab-v3	30	0.76	0.72	0.78	0.85	0.86	0.81	0.74	NaN	0.76	0.75	0.82	0.84	0.84	0.87	0.85	NaN	0.80
Segformer	30	0.76	0.73	0.77	0.85	0.87	0.80	0.72	NaN	0.76	0.76	0.81	0.87	0.89	0.90	0.88	NaN	0.81
Segmenter	30	0.63	0.61	0.63	0.74	0.70	0.68	0.59	NaN	0.62	0.59	0.68	0.74	0.70	0.81	0.78	NaN	0.68
Swin-L	30	0.60	0.58	0.63	0.69	0.64	0.67	0.67	NaN	0.55	0.52	0.57	0.63	0.62	0.73	0.76	NaN	0.64
SwinV2-G	30	0.52	0.48	0.52	0.59	0.57	0.68	0.67	NaN	0.49	0.45	0.53	0.60	0.66	0.76	0.76	NaN	0.60
BeiT-B	30	0.78	0.76	0.77	0.83	0.74	0.76	0.77	NaN	0.79	0.80	0.82	0.84	0.76	0.81	0.82	NaN	0.78
ViT-Adapter-L	30	0.78	0.76	0.77	0.83	0.74	0.76	0.77	NaN	0.79	0.80	0.82	0.84	0.76	0.81	0.82	NaN	0.79
TeethSEG	5	0.80	0.79	0.82	0.86	0.86	0.79	0.71	NaN	0.78	0.78	0.83	0.88	0.90	0.93	0.91	NaN	0.84

Table 2. Tooth segmentation results (mIoU) compared with SOTA methods on the IO150K out-of-the-distribution (o.o.d.) test splits

Method	Epoch	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	mIoU
DeepLab-v3	30	0.76	0.72	0.78	0.85	0.86	0.81	0.74	NaN	0.76	0.75	0.82	0.84	0.84	0.87	0.85	NaN	0.80
Segformer	30	0.49	0.52	0.64	0.89	0.86	0.85	0.74	NaN	0.45	0.50	0.65	0.84	0.83	0.84	0.57	NaN	0.69
Segmenter	30	0.48	0.44	0.47	0.62	0.66	0.58	0.58	NaN	0.52	0.49	0.49	0.60	0.63	0.57	0.55	NaN	0.55
Swin-L	30	0.43	0.50	0.58	0.55	0.49	0.46	0.58	NaN	0.41	0.51	0.43	0.35	0.35	0.51	0.59	NaN	0.48
SwinV2-G	30	0.58	0.43	0.51	0.56	0.53	0.43	0.43	NaN	0.60	0.37	0.44	0.42	0.41	0.42	0.32	NaN	0.46
BeiT-B	30	0.61	0.48	0.46	0.49	0.53	0.39	0.35	NaN	0.61	0.50	0.51	0.53	0.48	0.37	0.33	NaN	0.47
ViT-Adapter-L	30	0.83	0.89	0.89	0.92	0.92	0.91	0.83	NaN	0.85	0.88	0.86	0.92	0.92	0.89	0.72	NaN	0.85
TeethSEG	5	0.91	0.93	0.88	0.93	0.94	0.92	0.87	NaN	0.86	0.94	0.90	0.94	0.93	0.93	0.91	NaN	0.91

Table 3. Tooth segmentation results (mIoU) compared with SOTA methods on the IO150K RGB test splits

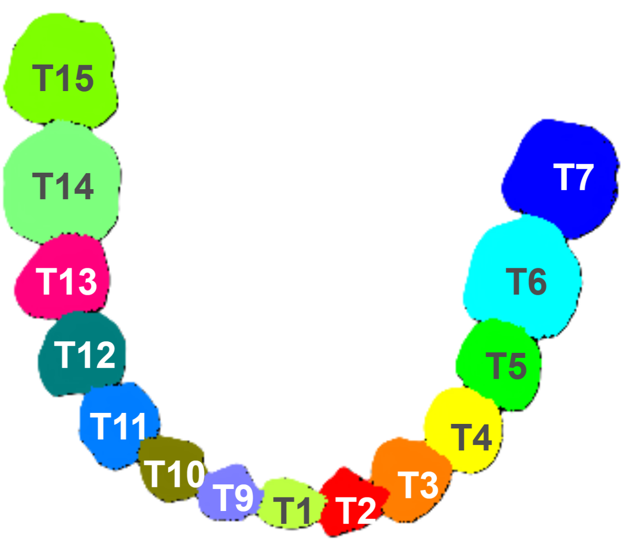

Figure 2. Classes Definition and Attention Mask in APK.

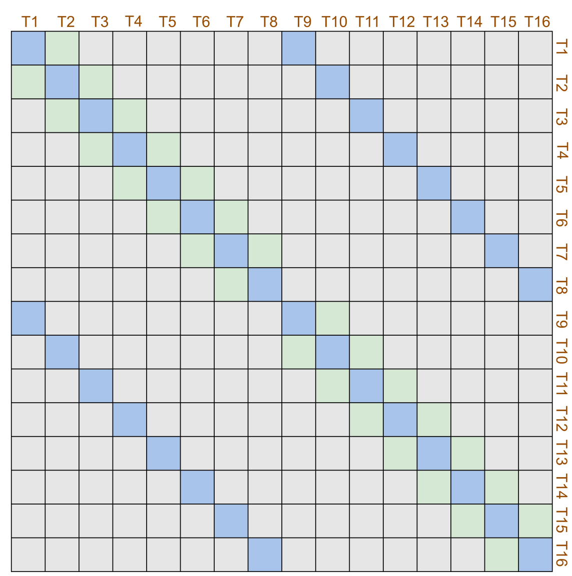

Figure 3. The attention mask of the APK layer.

B. The Classes Definition and The Attention Mask for The APK Layer

The pre-defined order of Tooth IDs is presented in Figure 2. In Figure 3, we visualize the attention mask that only enables interaction between adjacent (green) and contralateral (blue) teeth. We set the attention score of "$-inf$" for the locations with a gray color in the mask and 0 for the locations with green or blue.

C. Supplementary Results of The O.O.D. Test and The RGB Test.

We provide the detailed quantitative results of each tooth (Table 2, Table 3) and visualizations (Table 4, Table 5) for the o.o.d. test and RGB test. We find previous methods can still provide clear and appropriate segmentation boundaries but fail to provide accurate tooth IDs in complex situations of orthodontic treatment, especially in the RGB test.

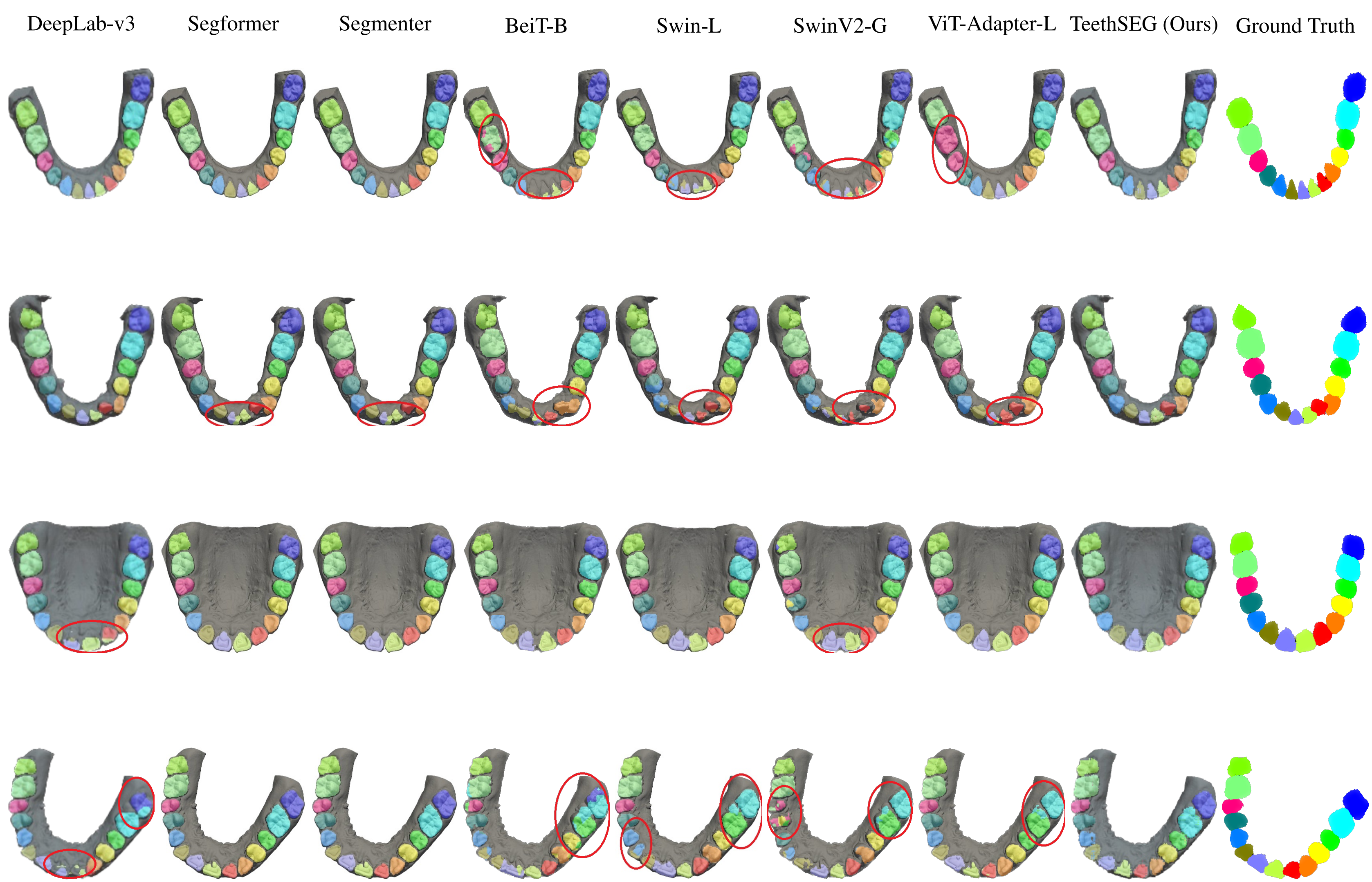

Table 4. The visual comparison of segmentation results (o.o.d test), as well as the corresponding ground truth.

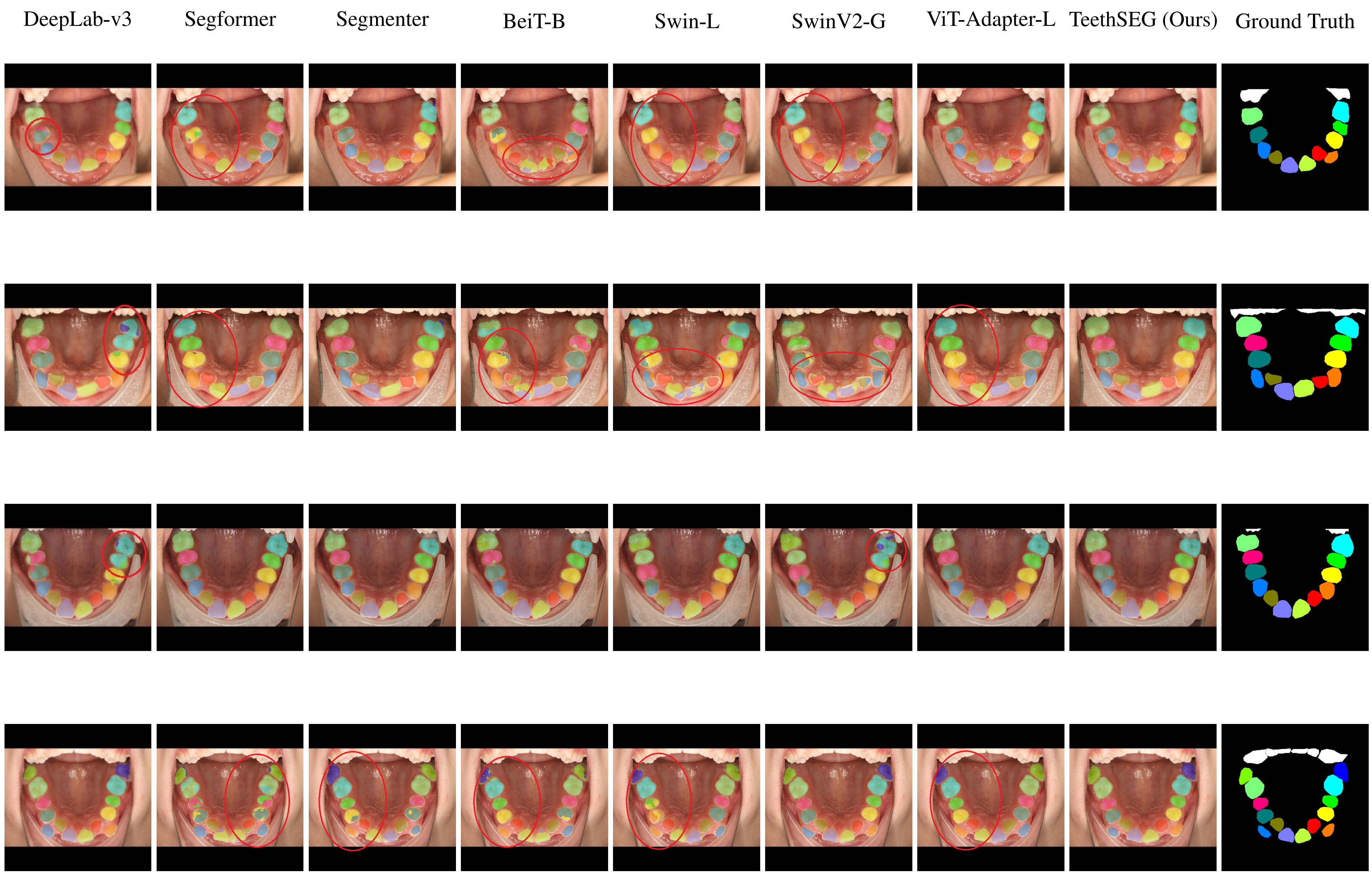

Table 5. The visual comparison on RGB test. Previous methods fail to provide accurate tooth IDs in complex situations of orthodontic treatment.

Variant	Res.	Attention Layers M	Embed Dim. D	MSA Blocks N_{MSA}	Naive Upscalers N_{Up}	mIoU
(a)	224×224	3	768	3	2	0.83
(b)	336×336	1	768	3	2	0.82
(c)	336×336	2	768	3	2	0.85
(d)	336×336	3	256	3	2	0.79
(e)	336×336	3	512	3	2	0.89
(f)	336×336	3	768	1	4	0.81
(g)	336×336	3	768	2	3	0.89
(h)	336×336	3	768	3	2	**0.91**

Table 6. Ablations on Hyper-parameters.

D. Ablations on Hyper-parameters

We design a series of variants of TeethSEG to study the best choice of hyper-parameters. In Table 6, (h) presents the performance of the best variant in this paper, (a) studies the influence of resolution of inputs, (b) and (c) explore the influence of the number of the transformer layers for the shallow fusion, (d) and (e) investigate the impact intermediate dimension of embeddings, (f) and (g) study the consequence of the different number of stacked MSA Blocks.

References

[1] Achraf Ben-Hamadou, Oussama Smaoui, Ahmed Rekik, Sergi Pujades, Edmond Boyer, Hoyeon Lim, Minchang Kim, Minkyung Lee, Minyoung Chung, Yeong-Gil Shin, et al. 3dteethseg'22: 3d teeth scan segmentation and labeling challenge. *arXiv preprint arXiv:2305.18277*, 2023. 2