

Supplementary Table 1. Patient-level segmentation statistics (tested on n=43 patients).

Model	Dice	IoU	Recall	Precision
UNet	90.25±2.02	83.18±2.65	94.05±2.08	87.46±2.21
UNet++	87.48±3.02	79.30±3.49	95.71±2.44	81.49±3.57
FCN	86.88±5.35	78.30±7.51	90.91±5.49	84.36±5.83
Deeplabv3	86.89±5.31	78.32±7.41	91.04±5.25	84.25±5.78
Swin	90.43±3.57	83.32±5.37	95.57±3.01	86.33±4.94
TSTNet	93.16±2.91	87.95±4.49	95.39±3.11	91.71±3.69

IoU: intersection over union ; Dice: Dice Coefficient. Bolden numbers indicate the best performance.

Supplementary Table 2. Model complexity and clinical utility of CBCT segmentation. (tested on n=43 patients).

Models	Unet	Unet++	FCN	Deepla bv3	Swin	TSTNst (w/o aug)	TSTNet	TSTNst (multi-scale)	Human Experts
#Params(M)	29.62	34.95	49.48	68.10	121.17	121.43	121.43	121.43	-
Inf-T(s)	~0.072	~0.085	~0.076	~0.066	~0.071	~0.052	~0.479	~3.070	~600- 900

#Params: number of parameters in the neural networks. Inf-T: end-to-end inference time of models. ~ means approximately.

“w/o aug” denotes TSTNet with standard augmentation, i.e., without cropping and domain specific augmentation.

“ multi-scale” denotes TSTNet with multi-scale inference.

Supplementary Table 3. Model complexity and clinical utility of IOS segmentation. (tested on n=100 patients).

Models	Baseline	DC-Net	IOSNet	Human Experts
#Params(M)	0.4	1.5	1.5	-
Inf-T(s)	~321	~24	~24	~600-900

#Params: number of parameters in the neural networks. Inf-T: end-to-end inference time of models.

~ means approximately.

Supplementary Table 4. Ablation studies of TSTNet (tested on n=43 patients).

Model	Dice	IoU	Recall	Precision
TSTNet	96.49	93.22	97.75	95.27
w/o aug	95.54	91.46	96.58	94.52
w/o loss	96.35	92.97	97.93	94.92
w/o TEC	96.3	92.87	97.92	94.73
w/o aug+loss+TEC (Swin)	95.02	90.52	97.69	92.5

IoU: intersection over union ; Dice: Dice Coefficient. Bolden numbers indicate the best performance.

“w/o aug” denotes TSTNet with standard augmentation, i.e., without cropping and domain specific augmentation.

“w/o loss” denotes TSTNet with cross entropy loss but without the upernet loss and FCN loss.

“w/o TEC” denotes TSTNet without the TEC module.

“w/o aug+loss+TEC(Swin)” denotes the original Swin transformer.

Supplementary Table 5. Ablation study of IOSNet (tested on n=100 patients).						
Model	Mandible			Maxillary		
	mIoU(%)	ACC _f (%)	ACC _a (%)	mIoU(%)	ACC _f (%)	ACC _a (%)
IOSNet	90.37	96.82	97.96	92.50	97.11	98.27
w/o centroid loss	87.83	95.73	96.95	90.95	96.45	97.69
w/o boundary loss	90.15	96.62	97.77	91.91	96.82	98.03
w/o both	90.29	96.75	97.88	91.48	96.93	98.11
IOSNet*	92.44	97.31	98.41	93.73	97.36	98.53
IOSNet**	94.23	98.14	98.81	95.70	98.35	99.05

w/o: without. mIoU: mean Intersection over Union, ACC_f: per-face accuracy, ACC_a: average-area accuracy.

IOSNet*: IOSNet trained on 28,000 scans without smoothing. IOSNet**: IOSNet* with smoothing.

The remaining are trained on a small dataset with 4271 IOS scans.

Supplementary Table 6. Curvature-based segmentation on 3D CBCT meshes. (tested on 50 patients).			
Method	Gaussian Curvature-based method	DBSCAN	Ours
Success Rate	42%	24%	94%

Success Rate: The probability of successfully separating CBCT mesh into the maxillary and mandible.

Supplementary Table 7. Evaluation of the registration between reconstructed CBCT meshes and IOS scans.(tested on 30 groups data).

Method	Super4pc	AAICP	Ours
Fitness	0.112	0.0573	0.3100
Inlier_rmse	0.0858	0.0866	0.0823
Correspondence_set size	7933.9	4009.5	21270.9
Success Rate	13.3%	6.67%	90%

Fitness, which measures the overlapping area. The higher the better.

Inlier_rmse, which measures the RMSE of all inlier correspondences. The lower the better.

Supplementary Table 8. Multimodal fusion results of the proposed method. (tested on 20 individual cases).

Case	ASSD (mm)	CD (mm)	HD (mm)
Case01	0.14±0.04	0.17±0.02	0.35±0.22
Case02	0.18±0.02	0.21±0.02	0.40±0.05
Case03	0.18±0.03	0.21±0.03	0.39±0.08
Case04	0.14±0.02	0.17±0.02	0.31±0.06
Case05	0.25±0.07	0.27±0.07	0.51±0.10
Case06	0.11±0.02	0.14±0.02	0.21±0.04
Case07	0.17±0.04	0.19±0.05	0.35±0.13
Case08	0.14±0.02	0.17±0.02	0.30±0.09
Case09	0.14±0.03	0.17±0.02	0.32±0.07
Case10	0.10±0.02	0.14±0.02	0.23±0.06
Case11	0.12±0.06	0.15±0.06	0.28±0.16
Case12	0.11±0.05	0.14±0.04	0.26±0.12
Case13	0.10±0.02	0.14±0.02	0.23±0.05
Case14	0.13±0.04	0.16±0.03	0.31±0.08
Case15	0.21±0.09	0.23±0.07	0.48±0.23
Case16	0.15±0.08	0.18±0.06	0.34±0.17
Case17	0.12±0.03	0.15±0.03	0.26±0.07
Case18	0.15±0.02	0.17±0.02	0.30±0.06
Case19	0.22±0.03	0.24±0.03	0.57±0.13
Case20	0.21±0.09	0.24±0.07	0.59±0.55

ASSD:Average symmetric surface distance; HD: Housdoff distance; CD: Chamfer distance

