

sTable 1. Computational efficiency comparison of different methods on the cephalometric dataset

Method	Params (M)	FLOPs (G)	GPU Memory (GB)	Inference Time (ms/image)
U-Net	31.04	12.35	0.85	11.2
ResNet-50 (Regression baseline)	23.52	15.60	1.12	14.5
Swin-T (Single-branch Transformer)	27.51	11.84	1.35	18.2
TransFuse	26.34	14.21	1.28	16.8
Swin-Unet	27.17	11.56	1.42	19.5
DS-TransUNet	34.25	18.92	1.78	24.1
Res-SwinFusion (Ours)	42.18	23.45	1.95	22.4

sTable 2. Subgroup quantitative evaluation (MRE in mm) under challenging cases

Challenging Subsets	CNN Baseline	Swin-Unet	Res-SwinFusion (Ours)
Abnormal Anatomy	3.42 ± 0.51	2.98 ± 0.42	2.21 ± 0.35
Low Contrast/Quality	2.85 ± 0.38	2.45 ± 0.31	1.94 ± 0.28
Ambiguous Landmarks	4.12 ± 0.65	3.24 ± 0.48	2.56 ± 0.41
Abnormal Anatomy	3.42 ± 0.51	2.98 ± 0.42	2.21 ± 0.35
Low Contrast/Quality	2.85 ± 0.38	2.45 ± 0.31	1.94 ± 0.28