

Supplementary Materials

Table S1. Comparison between the score results for the gold standard segmentation in the included databases with details for each sample. For each score, the average value (and corresponding standard deviation at 1 σ) is reported.

cases	SPECIFICITY	SENSITIVITY	PRECISION	F ₁ SCORE
coronacases 001	0.9990	0.428	0.779	0.553
coronacases 002	0.9996	0.669	0.903	0.769
coronacases 003	0.9998	0.535	0.984	0.693
coronacases 004	0.9999	0.511	0.902	0.652
coronacases 005	0.9993	0.801	0.723	0.760
coronacases 006	0.9988	0.849	0.657	0.741
coronacases 007	0.9985	0.693	0.567	0.623
coronacases 008	0.9993	0.563	0.851	0.678
coronacases 009	0.9994	0.483	0.733	0.582
coronacases 010	0.9979	0.621	0.767	0.686
CORONACASES overall	0.9992±0.0005	0.62±0.13	0.79±0.12	0.67±0.07
RC006	0.9987	0.863	0.246	0.382
RC017	0.9995	0.553	0.923	0.692
RC036	0.9993	0.618	0.911	0.736
RC058	0.9995	0.741	0.440	0.552
RC143	0.9996	0.920	0.834	0.910
GOLD STD OVERALL	0.9993±0.0003	0.74±0.14	0.67±0.28	0.65±0.18
Overall	0.9992±0.0005	0.66±0.15	0.75±0.20	0.67±0.12

Table S2. Details of the various state-of-art methods referenced in the paper.

Study	Technique	DataSet	Evaluation Metrics	Comments
Fan2020 [66]	InfNet		Dice, Spec, Sens, Prec	
Fan2020 [66]	SemiInfNet		Dice, Spec, Sens, Prec	
Muller2020 [49]	U-Net	COVID-19-CT-Seg	Dice, Sens, Spec	Intensive Data Augmentation
Jun2020 [52]	3D U-Net	COVID-19-CT-Seg, Many Other	Dice	Segmentation Benchmark: trained on Many Different Dataset, Validated on COVID-19-Seg
Jun2020 [52]	2D U-Net	COVID-19-CT-Seg, Many Other	Dice	Segmentation Benchmark: trained on Many Different Dataset, Validated on COVID-19-Seg
Qingsen2020 [65]	U-Net	861 COVID-19 CT scans	Dice, Sensitivity, Precision	