

Table 1: *Details of the experimental settings.* We do not perform heavy hyperparameter tuning to ensure the generalizability of our best practices. The ensemble is not used, which can further improve performances. *For Brain, Vessel, and Pancreas, we train each model with 250,000 steps given their larger training set. We provide all the codes, including data pre-processing, data loading, model training, and evaluation at <https://github.com/yuhui-zh15/TransSeg>

Hyperparam	Value	Hyperparam	Value
Batch Size	16	Patch Size	$16 \times 16 \times 5$
Loss Function	DiceFocal	Optimizer	AdamW
Learning Rate	3e-5	Weight Decay	5e-2
Scheduler	Slanted Triangular	Warm-up Steps	20
Step	25,000*	Hyperparam Tuning	No
GPUs	8 Titan RTX	Time	8 Hours
Ensemble	No multi-model, multi-view, multi-scale ensemble		
Train Data	Random Zoom ($[0.5 \times, 2 \times]$), Random Crop (if $\text{Zoom} > 1 \times$), Normalize Intensity ($[-175, 250]$ (CT) or $[0, \text{MAX}]$ (MRI) $\rightarrow [-1, 1]$), Random Flip ($p = 0.1$), Random Rotation ($p = 0.1$), Random Shift Intensity ($[-0.1, +0.1]$ with $p = 0.5$), Pad (if $\text{Zoom} < 1 \times$)		
Inference Data	Normalize Intensity (same as training)		