Appendix A   Mean and Deviation Results

We collect test set segmentations on each dataset and run 100 bootstrap iterations. We show $\mu \pm \sigma^2$ for models trained using $\mathcal{L}_{Dice}$ (left) and $\mathcal{L}_{CE}$ (right).

Table 3: Results when training with $\mathcal{L}_{Dice}/\mathcal{L}_{CE}$ as base loss, adding $\mathcal{L}_{clDice}$ and $\mathcal{L}_{clCE}$.

	$\mathcal{L}_{Dice}$		$\mathcal{L}_{Dice} + 0.5 \cdot \mathcal{L}_{clDice}$		$\mathcal{L}_{Dice} + \mathcal{L}_{clCE}$	
	DSC	cl-DSC	DSC	cl-DSC	DSC	cl-DSC
HRF	78.88±0.40	77.46±0.72	75.63±0.53	83.79±0.51	79.37±0.37	78.97±0.64
DR-HAGIS	69.28±0.82	68.73±0.77	66.74±0.81	74.06±0.81	70.79±0.86	70.03±0.84
TREND	63.76±0.75	65.23±0.54	61.13±0.64	71.35±0.47	64.64±0.62	67.08±0.50
ASOCA	84.62±0.87	84.70±0.79	83.33±0.93	84.77±0.86	84.75±0.96	85.04±0.83

	$\mathcal{L}_{CE}$		$\mathcal{L}_{CE} + 0.5 \cdot \mathcal{L}_{clDice}$		$\mathcal{L}_{CE} + \mathcal{L}_{clCE}$	
	DSC	cl-DSC	DSC	cl-DSC	DSC	cl-DSC
HRF	79.28±0.29	76.92±0.71	73.34±0.56	85.50±0.51	79.46±0.29	77.83±0.79
DR-HAGIS	68.91±1.06	67.28±0.89	66.05±0.98	78.80±0.74	70.77±1.02	68.04±0.83
TREND	62.89±0.69	61.36±0.63	58.41±0.50	73.24±0.54	64.60±0.67	64.20±0.66
ASOCA	85.37±0.74	85.85±0.67	59.40±2.27	80.96±1.34	85.50±0.86	86.21±0.66