

5 Supplementary Materials

Table S1: Main results with the accuracy metric.

Backbone	Baseline	Segmentation			Magnification		
		Multitask	TTT	TTT	Multitask	TTT	TTT
		(online)			(online)		
Camelyon (biased setting)							
H-optimus-1	0.90	0.90	0.90	0.90	0.94	0.94	0.93
Kaiko-Midnight	0.93	0.94	0.94	0.94	0.90	0.90	0.91
Virchow2	0.94	0.96	0.96	0.96	0.96	0.96	0.96
UNI2-h	0.87	0.87	0.87	0.87	0.85	0.85	0.85
Prov-GigaPath	0.83	0.83	0.83	0.83	0.84	0.84	0.84
Phikon-v2	0.59	0.59	0.59	0.59	0.60	0.60	0.60
Tolkach-ESCA (biased setting)							
H-optimus-1	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Kaiko-Midnight	0.97	0.96	0.96	0.96	0.97	0.97	0.97
Virchow2	0.97	0.97	0.97	0.97	0.97	0.97	0.97
UNI2-h	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Prov-GigaPath	0.94	0.94	0.94	0.94	0.95	0.95	0.95
Phikon-v2	0.94	0.94	0.94	0.94	0.94	0.94	0.94
Chaoyang $\rightarrow$ NCT							
H-optimus-1	0.88	0.86	0.86	0.86	0.88	0.88	0.89
Kaiko-Midnight	0.93	0.97	0.97	0.97	0.97	0.97	0.97
Virchow2	0.78	0.85	0.85	0.85	0.90	0.90	0.91
UNI2-h	0.87	0.85	0.85	0.85	0.83	0.83	0.84
Prov-GigaPath	0.76	0.79	0.79	0.79	0.83	0.83	0.84
Phikon-v2	0.90	0.73	0.73	0.73	0.89	0.89	0.88
NCT $\rightarrow$ Chaoyang							
H-optimus-1	0.82	0.88	0.88	0.88	0.91	0.91	0.90
Kaiko-Midnight	0.70	0.79	0.79	0.79	0.74	0.74	0.77
Virchow2	0.67	0.66	0.66	0.66	0.68	0.68	0.69
UNI2-h	0.68	0.68	0.68	0.68	0.62	0.62	0.60
Prov-GigaPath	0.87	0.86	0.86	0.86	0.72	0.72	0.77
Phikon-v2	0.46	0.45	0.45	0.45	0.48	0.48	0.48

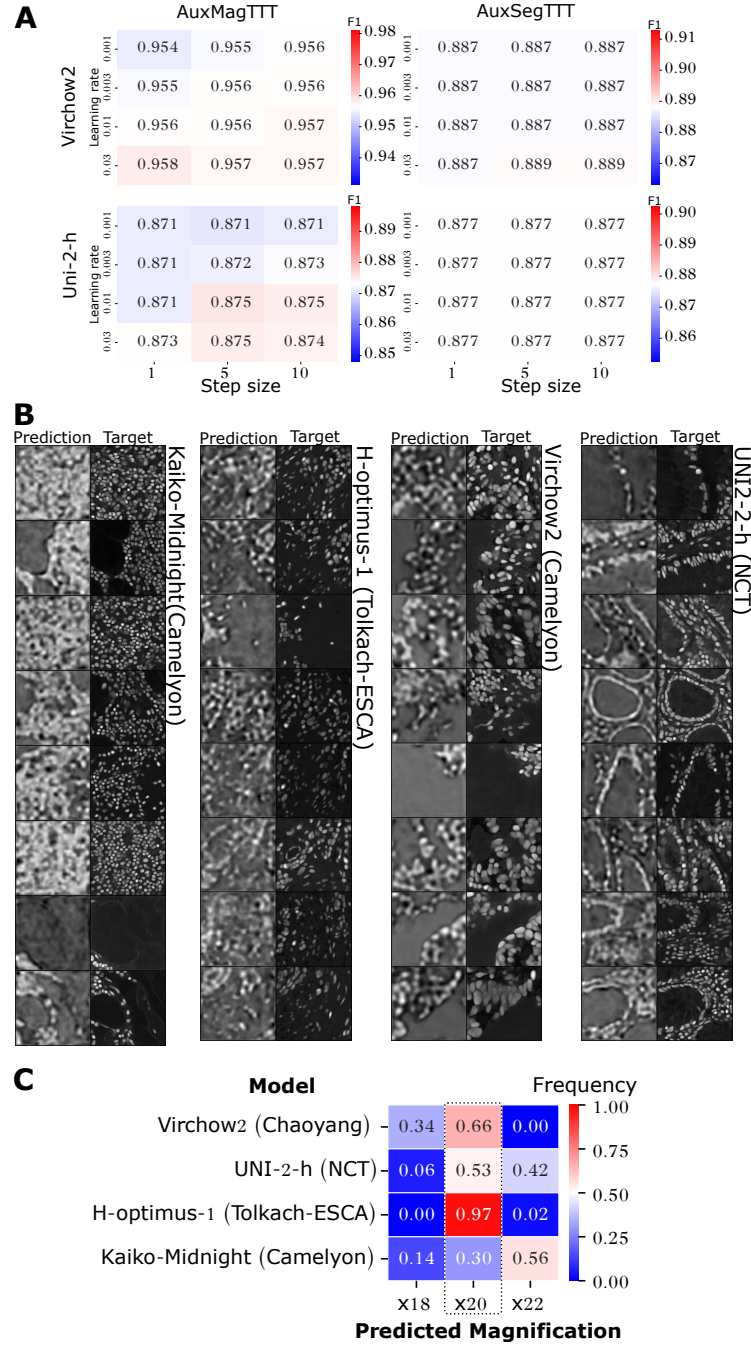


Fig. S1: (A) Results of a grid-search across different TTT-step learning rates and TTT-step sizes shown for two selected Backbones for both AuxMagTTT and AuxSegTTT. Original grid-search was performed on the validation datasets but shown here are the results on the NCT dataset test split with models trained on the Chaoyang dataset. (B) Selected examples of the nuclei segmentation output of several backbone-dataset combinations. (C) Heatmap showing the accuracy of the auxiliary magnification task on the OOD test set (ground truth for all instances is the native x20 magnification level) of selected backbones trained on different datasets.