Incremental Nuclei Segmentation from Histopathological Images via Future-class Awareness and Compatibility-inspired Distillation

Supplementary Material

1. More Ablation Studies

We compare the proposed compatibility-inspired distillation against different output-level distillations [1, 3, 4] to further show its advantage. As shown in Table 1, standard knowledge distillation exhibits the poorest performance as it does not correct for background shift. Both unbiased distillation and similarity transfer address the background shift well, which results in consistent improvements. Our approach not only resolves the background shift but also fully harnesses the knowledge pre-trained by the old model for the future, thus achieving the best performance.

Distillation loss	Old	New	Mean
Knowledge Distillation [3]	63.12	75.80	66.29
Unbiased Distillation [1]	68.01	77.43	70.36
Similarity Transfer [4]	68.55	77.52	70.79
Ours	**70.46**	**78.03**	**72.36**

Table 1. Comparison of compatibility-inspired distillation against different output-level incremental techniques on MoNuSAC 3-1.

2. More Hyper Parameter Analysis

Different Weights for Local POD Loss. The impact of multi-scale pooling distillation loss $\mathcal{L}_{pod}$ [2] depends on the selection of the weight λ. A high λ will hinder the model from learning new classes, while a small λ causes the model to lose constraints on features. Table 2 shows our parameter selection on different λ. We can see that, while the λ is set to 0.0001, our method achieves the best performance on both 1-1 and 3-1 settings.

Dataset	λ	1-1			2-1		
		Old	New	Mean	Old	New	Mean
MoNuSAC	0.1	**69.85**	63.69	65.23	**66.70**	70.30	68.50
	0.01	68.98	66.42	67.06	65.03	73.87	69.45
	0.001	68.50	68.63	68.60	64.87	75.23	70.05
	0.0001	68.11	**69.86**	**69.44**	63.79	76.53	**70.16**
	0.00001	67.81	68.91	68.44	61.03	**77.21**	69.12
CoNSeP	0.1	**68.61**	65.11	66.27	**75.92**	56.51	69.45
	0.01	68.05	67.87	67.93	75.11	59.66	69.96
	0.001	67.34	69.01	68.45	74.69	62.33	70.57
	0.0001	66.64	70.84	**69.44**	74.29	64.65	**71.08**
	0.00001	65.10	**70.89**	68.96	72.91	**65.10**	70.30

Table 2. Statistical analysis of different λ applied on MoNuSAC and CoNSeP datasets.

Different Thresholds for Future-class Awareness. The choice of τ_u, whether high or low, will impact the model's determination of future classes and consequently affect the quality of learned features. As shown in Table 3, the suitable threshold for our module performing future-class awareness is 0.8.

Dataset	τ_u	1-1	2-1	2-2	3-1	Average
MoNuSAC	0	64.23	66.70	66.31	68.97	66.55
	0.6	68.32	68.44	70.34	70.36	69.36
	0.7	69.14	70.06	71.28	72.42	70.72
	0.8	69.44	70.16	71.32	72.36	**70.82**
	0.9	68.98	70.11	71.07	71.97	70.53
CoNSeP	0	67.54	69.38	-	-	68.46
	0.6	68.92	68.03	-	-	68.47
	0.7	68.98	70.85	-	-	69.91
	0.8	69.44	71.08	-	-	**70.26**
	0.9	69.03	70.11	-	-	69.57

Table 3. Statistical analysis of different thresholds τ_u applied on MoNuSAC and CoNSeP datasets.

3. Appendix of Endpoints Weight Fusion

Taking the final model θ_{old} from the previous incremental step as the starting point and the new model θ_{new} as the endpoint, the parameter fusion is then carried out using the following formula:

$$\theta_{new} = \alpha_t \theta_{new} + (1 - \alpha_t)\theta_{old}$$
$$\alpha_t = \sqrt{\frac{N_{new}}{N_{old}}}$$

where the number of old classes N_{old} and new classes N_{new} reflect the contributions of the old model and the new model, respectively.

4. More Visualizations

As shown in Figure 1 and Figure 2, we provide more visual comparisons with more incremental segmentation methods including EWF [6], IDEC [7], CoNuSeg [5], REMINDER [4] and MiB [1], which further proves the superiority of our proposed method for achieving a better balance between stability and plasticity.

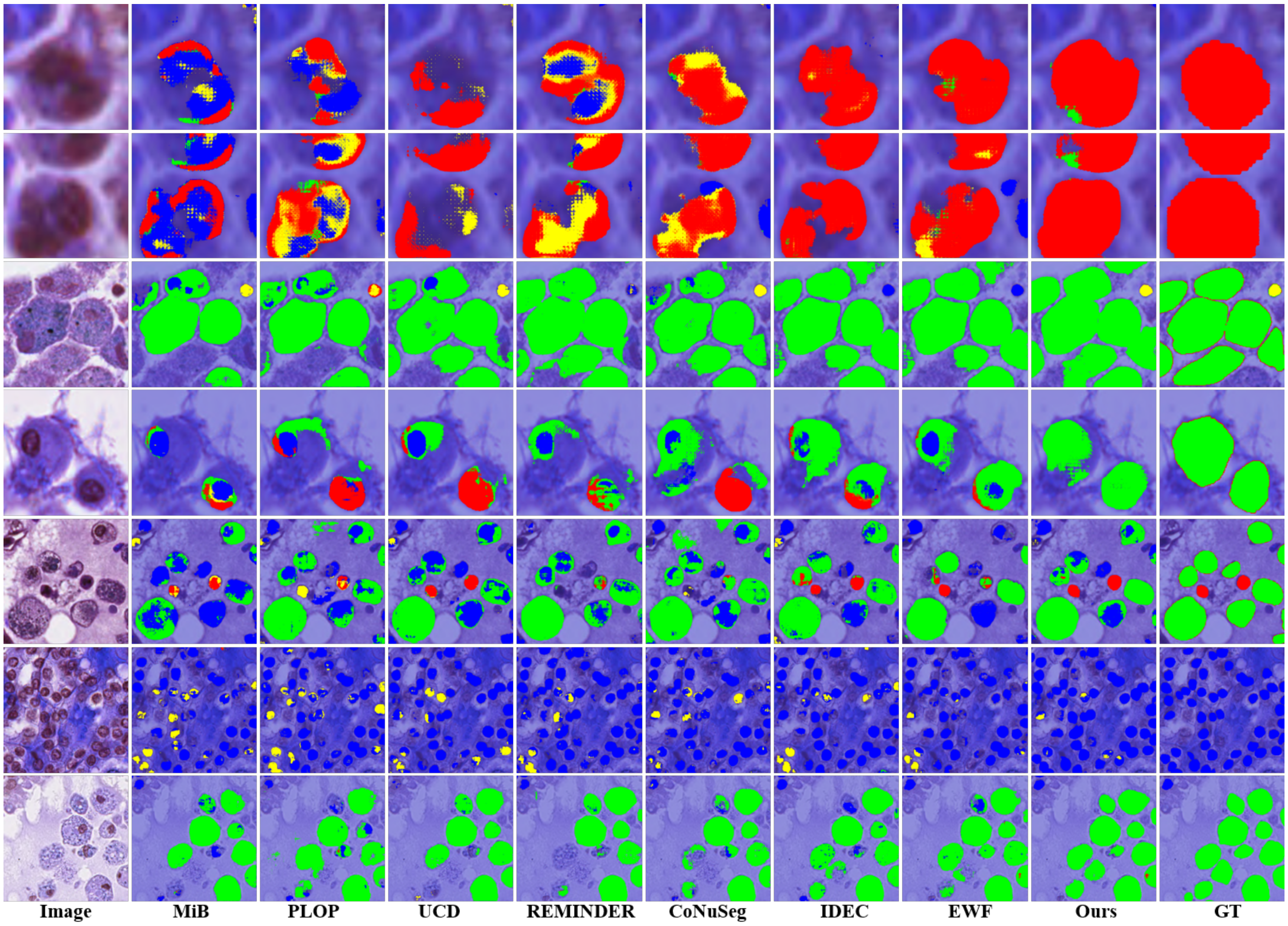

Figure 1. More visual comparisons with different state-of-the-art methods in incremental nuclei image segmentation.

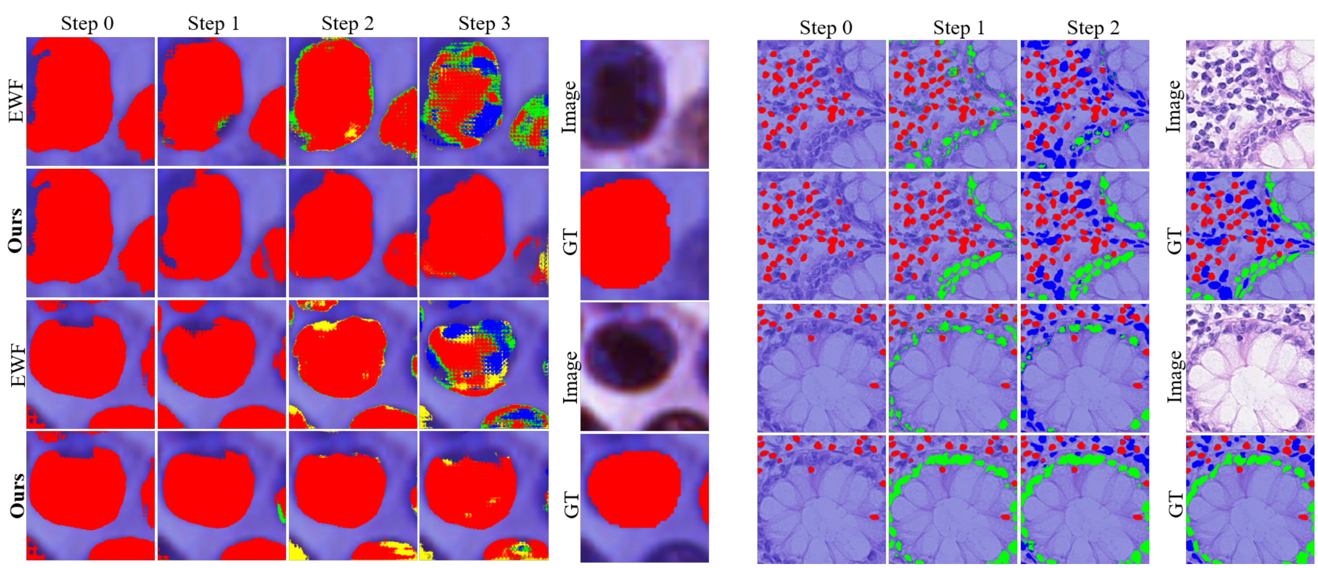

Figure 2. More visualizations of our method and EWF at different incremental steps on MoNuSAC 1-1.

References

[1] Fabio Cermelli, Massimiliano Mancini, Samuel Rota Bulo, Elisa Ricci, and Barbara Caputo. Modeling the background for incremental learning in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9233–9242, 2020. 1

[2] Arthur Douillard, Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Plop: Learning without forgetting for continual semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4040–4050, 2021. 1

[3] Umberto Michieli and Pietro Zanuttigh. Incremental learning techniques for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision workshops*, pages 0–0, 2019. 1

[4] Minh Hieu Phan, Son Lam Phung, Long Tran-Thanh, Abdesselam Bouzerdoum, et al. Class similarity weighted knowledge distillation for continual semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16866–16875, 2022. 1

[5] Huisi Wu, Zhaoze Wang, Zebin Zhao, Cheng Chen, and Jing Qin. Continual nuclei segmentation via prototype-wise relation distillation and contrastive learning. *IEEE Transactions on Medical Imaging*, 2023. 1

[6] Jia-Wen Xiao, Chang-Bin Zhang, Jiekang Feng, Xialei Liu, Joost van de Weijer, and Ming-Ming Cheng. Endpoints weight fusion for class incremental semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7204–7213, 2023. 1

[7] Danpei Zhao, Bo Yuan, and Zhenwei Shi. Inherit with distillation and evolve with contrast: Exploring class incremental semantic segmentation without exemplar memory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1