Supplementary Material for ECLIPSE: Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning

Beomyoung Kim[1,2] Joonsang Yu[1] Sung Ju Hwang[2]

NAVER Cloud, ImageVision[1] KAIST[2]

{beomyoung.kim,joonsang.yu}@navercorp.com, sjhwang82@kaist.ac.kr

A. Additional Analysis

A.1. Impact of Class Ordering on Performance

We delve deeper into the robustness of our method concerning class ordering. We carry out experiments on the ADE20K panoptic segmentation `100-10` scenario, employing 10 different class orderings. Note that we randomly shuffle the base class order 10 times to generate these varied orderings. In Figure S1, PQ distributions are illustrated through boxplots. Remarkably, our ECLIPSE demonstrates resilience to changes in class ordering, consistently outperforming other methods.

A.2. Continual Panoptic Segmentation under the Disjoint Setting

The seminal work [1] introduced two different settings, *disjoint* and *overlap*. Since the overlap setting is more challenging and realistic, we mainly followed it in our main paper. Here, we provide the experimental results on ADE20K [9] continual panoptic segmentation under the *disjoint* setting. Table S1 shows the superiority of ECLIPSE compared to existing continual panoptic segmentation methods.

A.3. Continual Panoptic Segmentation on COCO Dataset

We validate our approach on the COCO panoptic segmentation benchmark [7], comprising 100K training and 5K validation images spread across 133 classes. For the incremental protocol, we designate 83 base classes and increment by an additional 50 classes. We note that the class ordering of COCO panoptic segmentation consists of things and stuff in sequence. To conduct a more meaningful validation, we randomly shuffled this order:

```
[1, 3, 10, 47, 58, 9, 88, 16, 126, 120, 17, 129, 35,
119, 59, 57, 54, 90, 75, 38, 80, 48, 131, 56, 95, 25,
43, 2, 68, 110, 32, 14, 29, 11, 7, 52, 83, 102, 84, 73,
5, 45, 117, 93, 87, 46, 118, 34, 61, 19, 77, 111, 63,
98, 130, 66, 79, 97, 33, 86, 127, 104, 64, 49, 36, 6,
91, 50, 112, 8, 65, 132, 92, 27, 122, 22, 51, 85, 115,
28, 89, 70, 62, 12, 101, 108, 125, 123, 39, 81, 20, 40,
41, 114, 128, 74, 18, 99, 100, 60, 30, 124, 69, 37, 13,
23, 116, 55, 26, 121, 71, 67, 106, 133, 42, 107, 105,
109, 82, 103, 76, 94, 24, 15, 78, 53, 21, 96, 72, 113,
44, 31, 4].
```

Our method is compared against three baseline methods (MiB [1], PLOP [5], and CoMFormer [2]), all utilizing the ResNet-50 backbone network under the *overlap* setting. As demonstrated in Table S2, our approach exhibits superior performance with considerably fewer trainable parameters.

A.4. Exploring Pre-trained Knowledge

To demonstrate the potential for further improving ECLIPSE, we explore the impact of using more advanced frozen parameters of the base model. We study the effect of the frozen parameters in continual segmentation using various pre-trained weights from Cityscape [4], Mapillary Vistas [8], and COCO [7] panoptic segmentation datasets. By default, we used the ImageNet pre-trained weights only for the backbone network (ResNet-50 [6]). As shown in Table S3, using pre-trained weights on larger datasets (*e.g.*, Cityscape→Mapillary→COCO) results in more noticeable performance improvements. This result demonstrates the potential of our approach to further enjoy the more powerful expandability of the model.

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, 3

[2] Fabio Cermelli, Matthieu Cord, and Arthur Douillard. Comformer: Continual learning in semantic and panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3010–3020, 2023. 1, 2, 3

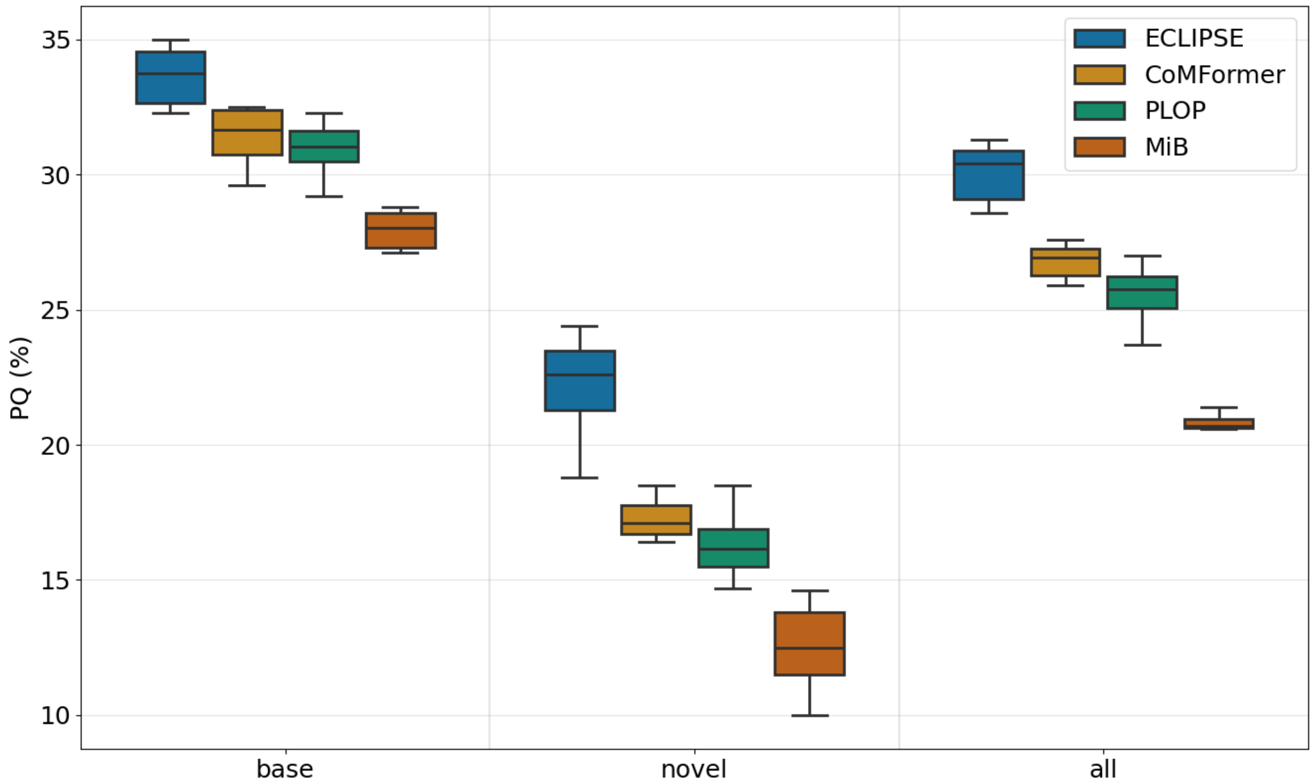

Figure S1. PQ distributions for 10 different class-orderings in the ADE20K panoptic segmentation `100-10` scenario.

Method	Backbone	Trainable Params	KD	**100-5** (11 tasks)			**100-10** (6 tasks)			**100-50** (2 tasks)		
				1-100	*101-150*	*all*	*1-100*	*101-150*	*all*	*1-100*	*101-150*	*all*
MiB [1]	R50	44.9M	✓	20.5	4.3	15.1	27.7	7.1	20.8	33.7	10.5	26.0
PLOP [5]	R50	44.9M	✓	19.2	8.8	15.8	28.9	10.6	22.8	34.8	12.4	27.4
CoMFormer [2]	R50	44.9M	✓	20.1	8.2	16.1	29.7	10.3	23.3	34.7	13.2	27.6
ECLIPSE	R50	**0.60M**		34.4	8.9	25.9	34.4	10.2	26.4	35.2	13.3	27.9

Table S1. **Continual Panoptic Segmentation** results on ADE20K dataset in PQ under the *disjoint* setting. *KD* denotes using distillation strategies, which demands more trainable parameters and computational overhead. All methods use the same network of Mask2Former [3] with ResNet-50 [6] backbone.

[3] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 2

[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1, 3

[5] 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, 2, 3

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 1, 3

[8] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999,

Method	Backbone	Trainable Params	KD	**83-5** (11 tasks)			**83-10** (6 tasks)		
				1-83	*84-133*	*all*	*1-83*	*84-133*	*all*
MiB [1]	R50	43.9M	✓	29.3	25.6	27.9	34.8	28.0	30.3
PLOP [5]	R50	43.9M	✓	34.0	27.1	31.4	37.7	31.1	35.2
CoMFormer [2]	R50	43.9M	✓	34.2	27.3	31.6	37.7	31.5	35.4
ECLIPSE	R50	**0.60M**		**36.9**	**31.7**	**34.9**	**38.1**	**34.5**	**36.7**

Table S2. **Continual Panoptic Segmentation** results on COCO [7] panoptic segmentation dataset where the total number of classes is 133 in PQ under the *overlap* setting. *KD* denotes using distillation strategies, which demands more trainable parameters and computational overhead.

Pretrained	**100-5** (11 tasks)			**100-10** (6 tasks)			**100-50** (2 tasks)		
	1-100	101-150	all	1-100	101-150	all	1-100	101-150	all
-	41.1	16.6	32.9	41.4	18.8	33.9	41.7	23.5	35.6
Cityscape [4]	41.7	16.9	33.2	42.2	18.9	34.5	42.2	23.8	35.9
Mapillary [8]	42.5	17.2	34.0	42.9	19.8	35.2	43.0	24.1	36.3
COCO [7]	46.1	18.9	37.0	46.4	22.3	38.4	44.2	29.0	39.1

Table S3. **Exploring pre-trained knowledge.** At the beginning of the continual learning process, we employ the pre-trained parameters to explore stronger frozen parameters of the base model.

2017. 1, 3

[9] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017. 1