Supplementary Materials for SSUL: Semantic Segmentation with Unknown Label for Exemplar-based Class-Incremental Learning

Anonymous Author(s)
Affiliation
Address
email

1 Additional Experimental Details

1.1 The details on datasets

Pascal VOC 2012 consists of 13,487 images, and it is divided as 10,582 images for training, 1,449 images for validation and 1,456 images for test dataset. **ADE20K** is a large scale dataset for semantic segmentation of scenes, including 25,210 images. It is also grouped as 20,210 images for the training set, 2,000 images for the validation set, and 3,000 images for the testing set. As stated in the manuscript, we followed exactly same experimental settings with PLOP [2].

2 The More Details of Experiments on Pascal VOC 2012

2.1 The details of experimental results of Pascal VOC 2012

	bg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mIoU
10-1 (11 tasks)																						
SSUL	83.22	88.43	39.48	86.07	68.54	82.81	84.41	88.96	86.15	36.37	67.16	20.20	65.95	32.44	59.14	80.28	26.19	36.89	18.79	38.93	32.53	58.23
SSUL-M	86.73	90.36	39.53	87.72	68.36	81.94	89.81	89.27	86.44	35.15	67.41	27.65	69.54	46.99	67.80	80.70	25.93	43.11	23.77	56.97	46.34	62.45
15-1 (6 tasks)																						
SSUL	85.86	90.06	41.63	88.74	69.94	79.35	90.44	88.85	92.76	36.84	78.21	59.53	90.49	87.66	82.54	86.14	28.44	44.45	17.65	31.97	20.21	66.27
SSUL-M	89.49	90.23	39.95	89.41	71.97	80.10	93.79	88.00	93.08	36.86	81.43	59.41	90.33	86.97	85.97	85.89	29.86	58.64	23.76	61.90	45.16	70.58
5-3 (6 tasks)																						
SSUL	86.49	73.10	37.84	85.10	65.05	79.49	41.21	59.68	67.67	12.58	43.94	37.13	61.67	35.69	61.22	78.54	35.61	46.74	21.00	34.18	43.85	52.75
SSUL-M	88.35	80.21	37.13	84.98	66.68	80.12	58.45	64.79	66.72	14.45	48.51	38.88	61.87	33.32	65.88	77.90	33.54	46.96	24.77	50.02	49.31	55.85

Table 1: Details of Pascal VOC 2012.

Table 1 shows the summarized results of Pascal VOC 2012 by each class name.

2.2 The details of experimental results of class orderings

15-1 (6 tasks)	**Class Ordering**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SSUL																				
1-15	78.06	74.46	73.32	75.45	77.43	71.72	76.56	81.50	75.06	74.19	76.94	80.30	78.69	78.30	72.72	77.65	74.83	75.29	74.71	74.11
16-20	28.54	29.25	23.48	38.03	34.54	32.83	18.97	20.23	36.35	23.02	30.50	35.21	40.31	15.88	37.26	27.44	26.35	27.25	46.56	50.66
all	66.27	63.70	61.45	66.54	67.22	62.46	62.85	66.91	65.84	62.00	65.88	69.57	69.55	63.44	64.28	65.70	63.29	63.85	68.00	68.53
SSUL-M																				
1-15	78.92	75.25	74.25	76.82	78.39	72.89	77.25	81.99	75.59	74.25	77.31	80.07	78.56	78.73	75.30	78.25	76.27	75.49	74.59	74.17
16-20	43.86	46.26	54.53	61.88	51.93	60.00	39.09	36.11	54.23	54.08	44.70	43.90	48.91	38.27	52.98	44.28	55.42	52.62	51.23	58.45
all	70.58	68.35	69.55	73.27	72.09	69.82	68.16	71.06	70.51	69.43	69.55	71.46	71.50	69.10	69.98	70.17	71.30	70.05	69.03	70.42

Table 2: Details of class orderings.

Table 2 shows the numerical details for SSUL and SSUL-M of Figure 3(b) in the manuscript. Note that we strictly followed the class orderings of Pascal VOC 2012 as done in PLOP [2]:

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

3 The Additional Experimental Results on ADE20K

3.1 Comparison with the jointly trained model in "ADE 100-5" and "ADE 100-10"

Table 3: Experimental results on ADE20K.

	ADE 100-5 (11 tasks)			ADE 100-10 (6 tasks)		
Method	0-100	101-150	all	0-100	101-150	all
ILT [3]	0.08	1.31	0.49	0.11	3.06	1.09
MiB [1]	36.01	5.66	25.96	38.21	11.12	29.24
PLOP [2]	39.11	7.81	28.75	40.48	13.61	31.59
SSUL	**42.03**	**15.80**	**33.35**	**42.10**	**16.02**	**33.46**
SSUL-M	**42.53**	**15.85**	**34.00**	**42.17**	**16.03**	**33.89**
Joint	44.30	28.20	38.90	44.30	28.20	38.90

To show the competitiveness of our proposed methods (SSUL and SSUL-M), we additionally trained "Joint" as an upper bound of CISS in ADE20K. Table 3 shows the experimental results of "Joint" with other baselines (the result of "ADE 100-5" and "ADE 100-10" for other baselines is exactly same with Table 2 in the manuscript). We clearly observe that the performance of SSUL and SSUL-M is not only overwhelming the performance of other baselines, but also nearly catching up with the upper bound.

3.2 Experimental results on the simple task sequence scenario

Table 4: Experimental results on ADE20K. SSUL-M denotes the result using exemplar memory.

	ADE 100-50 (2 tasks)		
Method	0-100	101-150	all
ILT [3]	18.29	14.40	17.00
MiB [1]	40.52	17.17	32.79
PLOP [2]	41.87	14.89	32.94
SSUL	42.13	13.32	32.59
SSUL-M	42.20	13.95	32.80
Joint	44.30	28.20	38.90

Table 4 shows the experimental results on "ADE 50-50". Note that it is not practical and the more simple scenario than others in the manuscript. Note that SSUL and SSUL-M also achieves almost competitive compared to other baselines.

3.3 Analysis of Qualitative Results

In Figure 1, we visualized the qualitative results from ADE20K 100-10 (6 tasks) scenario. We argue that we seldom suffer from the background semantic shift issue on ADE20K because its clear and dense labels for both things and stuff. Consequently, the false-positive predictions are noticeably reduced compared to the results on Pascal VOC 2012. As in Figure 1, the *unknown* label ($i.e.$, *black* pixels) is correctly transformed to the label to be learned in the future ($e.g.$, *fan* in step-5 and *plate* in step-6) while keeping the previously learned knowledge.

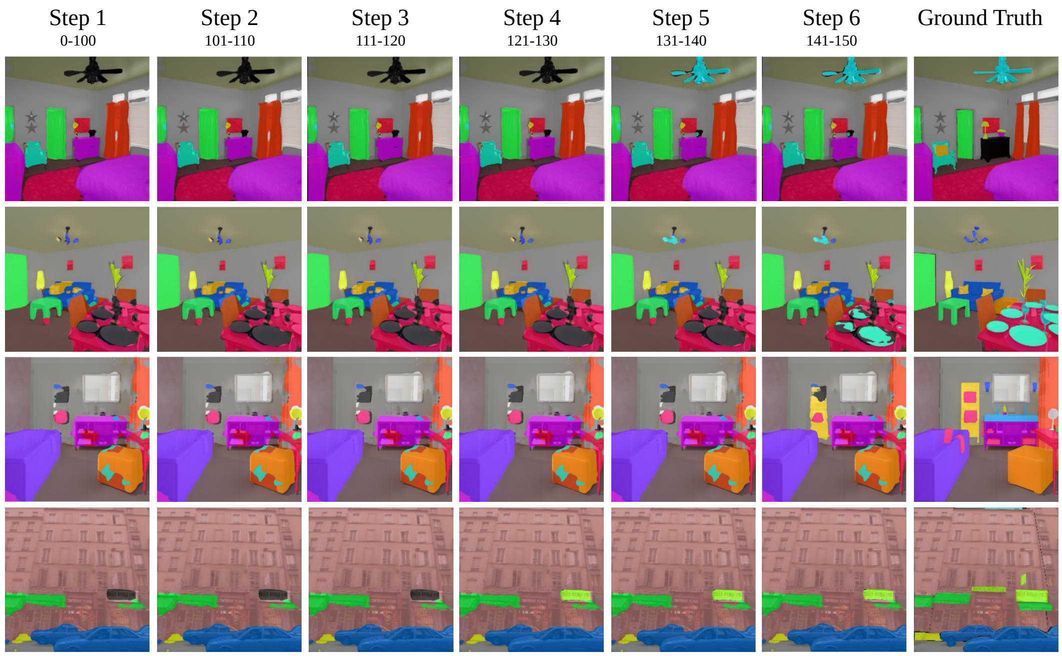

Figure 1: Qualitative results of SSUL-M on ADE20K.

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.

[2] Arthur Douillard, Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Plop: Learning without forgetting for continual semantic segmentation. *arXiv preprint arXiv:2011.11390*, 2020.

[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.