Supplementary Material
Mitigating Background Shift in Class-Incremental Semantic Segmentation

Gilhan Park, WonJun Moon, SuBeen Lee, Tae-Young Kim, Jae-Pil Heo*

Sungkyunkwan University, South Korea

A Appendix

Table A1: Results on Pascal VOC 10-1 (11 steps) and 5-3 (6 steps) *overlapped* setting. † indicates the reproduced results from SSUL [2].

Method	10-1 (11 steps)			5-3 (6 steps)		
	1-10	11-20	All	1-5	6-20	All
CNN-based Methods						
ILT† [10]	7.1	3.6	5.5	22.5	31.6	29.0
MiB† [1]	20.0	20.1	20.1	57.1	42.6	46.7
SDR [11]	32.4	17.1	25.1	-	-	-
PLOP† [5]	44.0	15.5	30.5	17.5	19.2	18.7
RECALL [9]	59.5	46.7	54.8	-	-	-
RCIL [14]	55.4	15.1	34.3	-	-	-
Joint	79.8	72.6	78.2	76.9	77.6	77.4
Transformer-based Methods						
MiB [1]	-	-	-	33.4	43.2	42.9
INC [13]	77.62	60.33	70.16	-	-	-
Ours	**80.95**	**71.98**	**77.19**	**77.5**	**77.5**	**78.1**
Joint	83.33	84.31	83.81	79.6	85.2	83.8

A.1 Additional quantitative results

Scenario Study with Small Number of Inital Classes. The experimental results for scenarios with a small initial class set are reported in Tab. A1. Experiments are conducted with Pascal VOC dataset. The 10-1 steps and 5-3 steps configuration present a more challenge due to its increased susceptibility to forgetting. This challenge arises from the initial step having a limited class set, in contrast to the subsequent steps which encompass a broader class set. In such a difficult scenario, our method shows consistent results, achieving a significant performance improvement. Specifically, in the scenario of steps 5-3,our method exceeds MiB [1] by 44.04%p in the initial step (1-5) and 34.24%p in the continual steps (6-20). Additionally, in the 10-1 scenario, the initial step (1-10) was 3.33%p higher and the continuous steps (11-20) were 11.65%p higher than the previous

* Corresponding author

Table A2: Ablation study for each component on Pascal VOC 15-1 in *disjoint* setting. SPL, AFD, $\mathcal{L}_{\text{Sep}}$ denote Selective Pseudo-Labeling, Adaptive Feature Distillation, and Separating objective, respectively.

	SPL	AFD	$\mathcal{L}_{\text{Sep}}$	Pascal VOC 15-1 (6 steps) 1-15	16-20	All
(a)	-	-	-	77.5	54.4	72.7
(b)	✓	-	-	81.0	58.7	76.3
(c)	-	✓	-	78.3	54.9	73.4
(d)	-	-	✓	79.1	61.6	75.6
(e)	✓	✓	-	81.4	59.7	76.9
(f)	✓	-	✓	80.9	63.7	77.4
(g)	-	✓	✓	79.1	62.1	75.7
(h)	✓	✓	✓	**81.5**	**64.7**	**78.1**

state-of-the-art (SOTA) model, INC [13]. This demonstrates the effectiveness of our distillation techniques, i.e., selective pseudo-labeling and adaptive feature distillation, in significantly addressing the background shift towards old classes. Furthermore, the improved performance in subsequent steps with a more extensive class set demonstrates the advantages of leveraging background knowledge to initialize new class tokens and the decoupling of shared knowledge, particularly in these scenarios. This approach effectively addresses the background shift towards new classes.

Ablation Study in the Disjoint setting. As discussed in Sec. 4.1, the disjoint setting is relatively simpler challenges compared to the overlapped setting, as future class objects do not appear in the images of the current step. In this subsection, we further investigate each component within the disjoint setting as shown in Tab. A2. In short, we find that an ablation study within a disjoint setting consistently validates the effectiveness of all components.

Results with CNN Backbone. We evaluate the effectiveness of our method using a CNN architecture. Specifically, we utilize a DeepLabv3 [3] segmentation network with a ResNet-101 [7] backbone, which is pre-trained on ImageNet [4]. For other configurations, we follow PLOP [5]. We evaluate our method on two datasets: Pascal VOC [6] and ADE20k [16]. In the Pascal VOC overlapped setting, experiments are conducted in the 15-5 (2 steps) and 15-1 (6 steps) scenarios, and similarly, in the ADE20k dataset overlapped setting, we perform experiments in the 100-10 (6 steps). As shown in Tab. A3, our method clearly outperforms all previous methods, demonstrating the strengths also with CNN backbones.

Hyperparameter Ablation. A threshold τ is set on the output of the previous model to generate pseudo-labels. To compare between different strategies of

Table A3: Results on Pascal VOC 15-5 (2 steps) and 15-1 (6 steps) in the *overlapped* setting, along with ADE20k 100-10 (6 steps) in the *overlapped* setting, all utilizing a CNN backbone. † indicates the reproduced results from [5, 13, 15]

	Pascal VOC						ADE20k		
	15-5 (2 steps)			15-1 (6 steps)			100-10 (6 steps)		
	1-15	16-20	All	1-15	16-20	All	1-100	101-150	All
CNN-based Methods									
MiB† [1]	76.4	50.0	70.1	34.2	13.5	29.3	38.2	11.1	29.2
SDR† [11]	75.4	52.6	69.9	44.7	21.8	39.2	28.9	7.4	21.7
PLOP† [5]	75.7	51.7	70.1	65.1	21.1	54.6	39.4	13.6	30.9
REMIND [12]	76.1	50.7	70.1	68.3	27.2	58.5	39.0	21.3	33.1
RCIL [14]	78.8	52.2	72.4	70.6	23.7	59.4	39.3	17.6	32.0
SPPA [8]	78.1	52.9	72.1	66.2	23.3	56.0	41.0	12.5	31.5
RBC† [15]	76.6	52.8	70.9	69.5	38.4	62.1	39.0	21.7	33.3
Ours	**78.8**	**54.3**	**73.3**	**76.1**	**39.1**	**67.8**	**42.8**	**22.6**	**36.1**
Joint	79.1	72.6	77.4	79.1	72.6	77.4	43.9	27.2	38.3

Table A4: Ablation on Pascal VOC 15-1 *overlapped* setting to evaluate the effect of hyperparameters.

λ_{LGKD}	1-15	16-20	All	λ_{Ortho}	1-15	16-20	All	Thr. τ	1-15	16-20	All
5	82.0	69.1	79.5	0.5	82.4	70.2	80.1	0.6	82.6	71.9	80.5
10	82.3	70.8	80.1	0.1	82.3	71.5	80.2	**0.7**	**82.6**	**72.2**	**80.6**
20	82.6	71.4	80.4	0.05	82.5	72.2	80.5	0.8	82.4	71.3	80.2
25	**82.6**	**72.2**	**80.6**	0.01	82.5	72.1	80.5	PLOP	79.6	68.2	77.4
50	82.3	70.14	79.9	$\lvert C^t \rvert / \lvert C^{1:t} \rvert$	**82.6**	**72.2**	**80.6**	SSUL	81.5	70.0	79.3

selecting hyperparameter τ, we conduct ablation studies with different pseudo-labeling methods, such as employing a sigmoid function [2] and using entropy function [5]. We reported the results of threshold selection in Tab. A4 (right). As observed, we find that setting a threshold *tau* to 0.7 is suitable for continual scenarios, ensuring the reliability of pseudo-labels thereby reducing the number of noisy labels. Additionally, we also point out the robustness of our method to different constant τ values. λ_{LGKD} (ranging from 5 to 50) and λ_{ortho} (from 0.5 to 0.01) are also examined in the left and the middle columns. These results further demonstrate the robustness of the model to hyperparameter variations, showing the consistent performances across different hyperparameter values.

A.2 Additional qualitative results

Additional qualitative results for SPL. Fig. A1 shows additional visualization results of the object identifier O which is introduced in Sec. 3.3. These results illustrate the weakness of conventional pseudo-labeling methods; ambiguous pixels, *e.g.*, sofa, chair, plant, and cat, are labeled as background due to the inaccuracy of the old model. However, our selective pseudo labeling using object identifier O is shown to be capable of detecting ambiguous pixels.

Additional qualitative results on Pascal VOC. We present additional qualitative results in the Pascal VOC 15-1 *overlapped* setting in Fig. A2. Our proposed methods demonstrate their advantages in terms of robustness against forgetting and platsticity in learning new classes, outperforming both our baseline (detailed in Sec. 4.3 and MiB [1]. For instance, while the baseline and MiB encounter difficulties in retaining old knowledge, often resulting in a background shift or overfitting to new classes (a-step 5, b-step 6, and c-step 3), our method consistently recognizes old classes. Furthermore, in contrast to our method, which correctly learns the new class 'sofa', we find that MiB [1] (especially b-step 4) has difficulty learning the new class. These results confirm the effectiveness of our method in preventing background shifts towards both old and new classes.

Qualitative results on ADE20k. Fig. A3 and Fig. A4 show qualitative results of our method, baseline, and MiB [1] on the 100-10 setting on ADE20k [16]. The models at the final training step are used to produce visualized results. The baseline uses conventional pseudo-labeling, vanilla feature distillation, and output distillation. These techniques are detailed in Sec. 4.3. As can be observed, both the baseline and MiB are vulnerable to background shifts, often misclassifying objects as the background class (black) or as other object classes that are visually similar (*e.g.*, house as building).

On the other hand, we find that our method is capable of preserving old class knowledge, as our distillation methods alleviate error propagation throughout the continual steps, effectively addressing the background shift towards old classes.

A.3 Limitation

Our method, particularly the selective pseudo-labeling strategy, effectively classifies ambiguous pixels for old classes, yet further development is needed for future classes. We believe that attempting to address limitations through analysis of our method will help direct future research on the CISS problem.

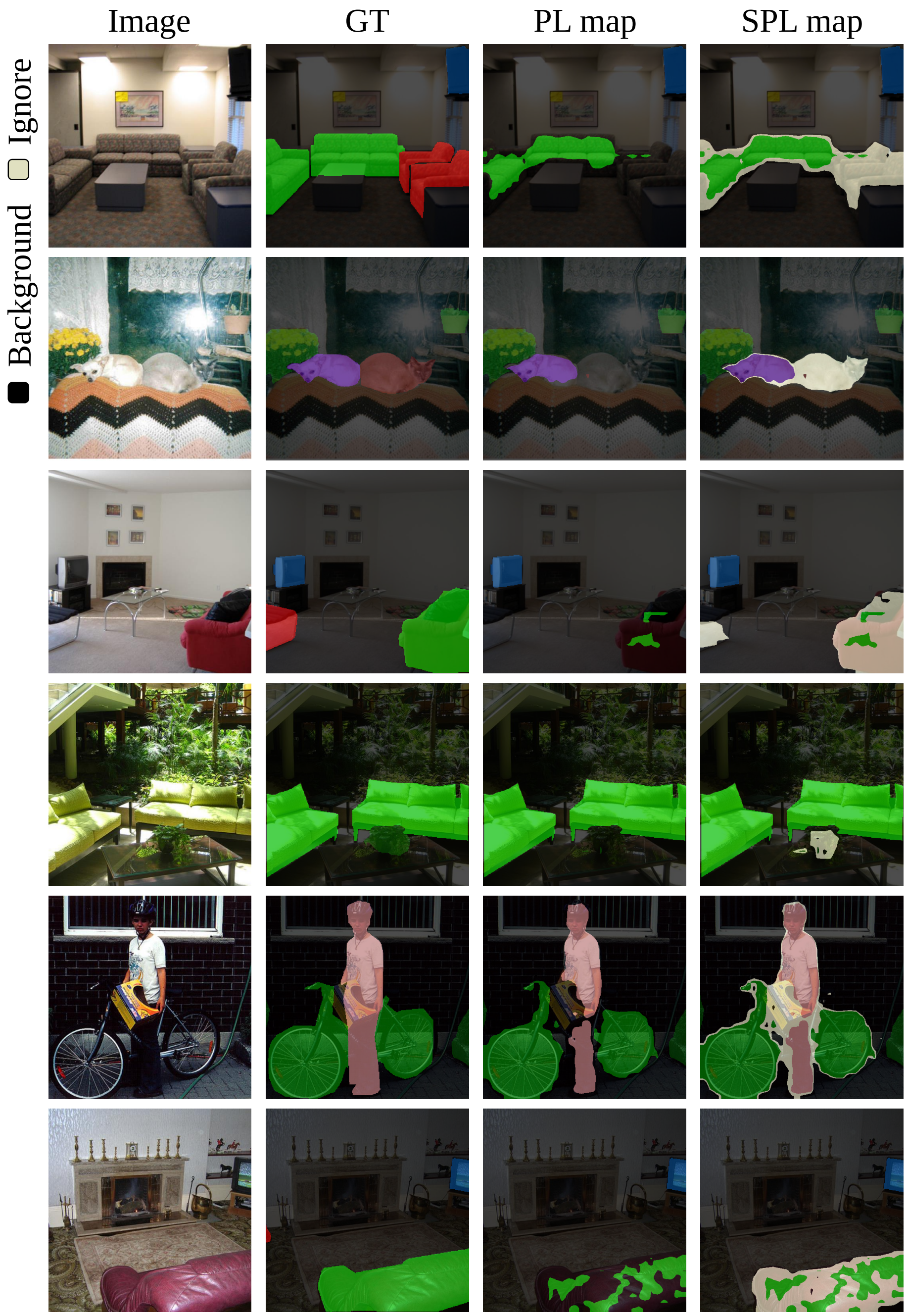

Fig. A1: Visualization of Pseudo-Labeling (PL) map and Selective Pseudo-Labeling (SPL) map. By comparing these two, we highlight the primary objective of selective strategies in pseudo-labeling: to effectively exclude pixels that might be mislabeled from the training process.

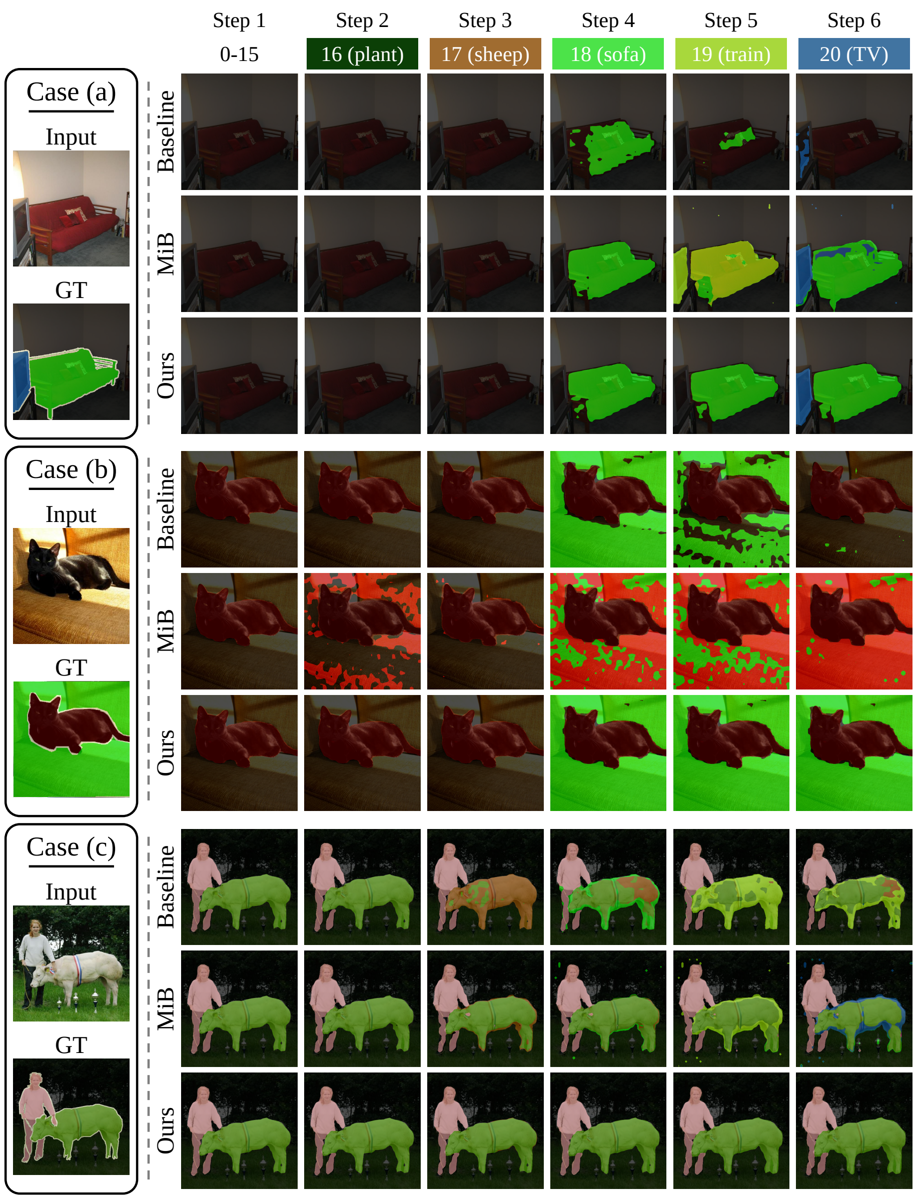

Fig. A2: Comparison of qualitative results on the 15-1 protocol of the Pascal VOC between the baseline, MiB, and ours.

Fig. A3: Qualitative results on the 100-10 protocol of the ADE20k.

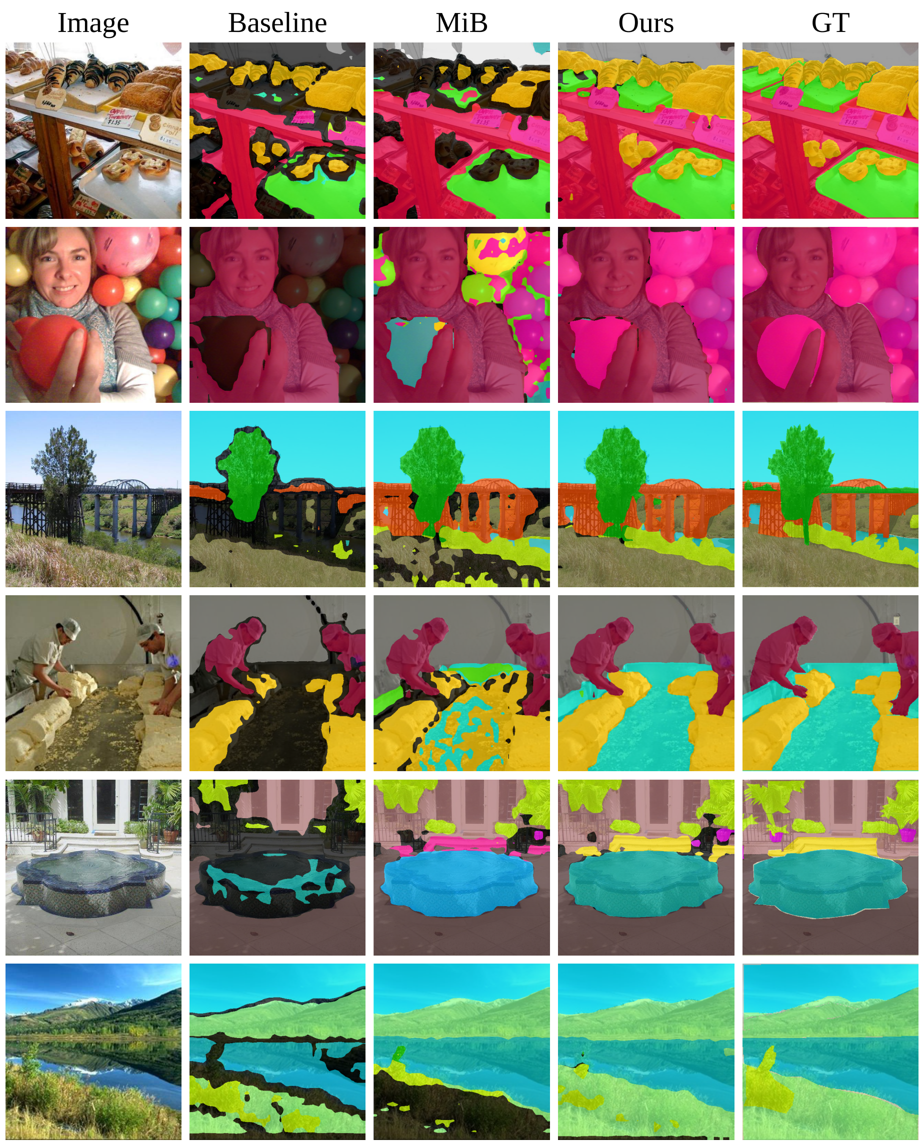

Fig. A4: Qualitative results on the 100-10 protocol of the ADE20k.

References

1. Cermelli, F., Mancini, M., Bulo, S.R., Ricci, E., Caputo, B.: Modeling the background for incremental learning in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9233–9242 (2020)
2. Cha, S., Yoo, Y., Moon, T., et al.: Ssul: Semantic segmentation with unknown label for exemplar-based class-incremental learning. Advances in neural information processing systems **34**, 10919–10930 (2021)
3. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
5. Douillard, A., Chen, Y., Dapogny, A., Cord, M.: Plop: Learning without forgetting for continual semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4040–4050 (2021)
6. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. International journal of computer vision **111**, 98–136 (2015)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. Lin, Z., Wang, Z., Zhang, Y.: Continual semantic segmentation via structure preserving and projected feature alignment. In: European Conference on Computer Vision. pp. 345–361. Springer (2022)
9. Maracani, A., Michieli, U., Toldo, M., Zanuttigh, P.: Recall: Replay-based continual learning in semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 7026–7035 (2021)
10. Michieli, U., Zanuttigh, P.: Incremental learning techniques for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision workshops. pp. 0–0 (2019)
11. Michieli, U., Zanuttigh, P.: Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1114–1124 (2021)
12. Phan, M.H., Phung, S.L., Tran-Thanh, L., Bouzerdoum, A., et al.: Class similarity weighted knowledge distillation for continual semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16866–16875 (2022)
13. Shang, C., Li, H., Meng, F., Wu, Q., Qiu, H., Wang, L.: Incrementer: Transformer for class-incremental semantic segmentation with knowledge distillation focusing on old class. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7214–7224 (2023)
14. Zhang, C.B., Xiao, J.W., Liu, X., Chen, Y.C., Cheng, M.M.: Representation compensation networks for continual semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7053–7064 (2022)

15. Zhao, H., Yang, F., Fu, X., Li, X.: Rbc: Rectifying the biased context in continual semantic segmentation. In: European Conference on Computer Vision. pp. 55–72. Springer (2022)
16. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ade20k dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 633–641 (2017)