Novel Class Discovery without Forgetting: Supplementary Material

1 Ablation Experiments

We systematically remove the two ingredients in our methodology and report the results in Tab. 1. We evaluate with CIFAR-10 and CIFAR-100-20-80 settings. The 20-80 setting is the most practical among the various setting on CIFAR-100 as it has lower number of labeled classes, and more unlabeled classes. We note that removing the pseudo-latent replay (PLR, Sec. 3.2), causes significant catastrophic forgetting. Selectively turning off the mutual-information based regularizer (MIR, Sec. 3.3) reduces the performance of class discovery. These results brings out the efficacy of our constituent methodological components. We note that our proposed method significantly closes the gap with the upper-bound which has access to the labeled and unlabeled data together, during training.

Table 1: Results on ablating Pseudo Latent Replay (PLR) and Mutual Information Regularizer (MIR) from our methodology.

Settings			CIFAR-10-5-5			CIFAR-100-20-80		
	PLR	MIR	Lab	Unlab	All	Lab	Unlab	All
Upper-bound	-	-	98.24	94.15	96.2	88.85	53.71	71.28
Ours	✓	×	92.12	85.57	88.85	84.95	43.34	64.15
	×	✓	34.12	91.08	62.6	6.75	50.97	28.86
	✓	✓	92.72	90.32	91.52	84.8	49.67	67.24

2 Sensitivity Analysis on Mixing Coefficient

The pseudo-latent representations $\boldsymbol{z}_p$ that we generate using Algo. 1 uses a mixing coefficient α to do a linear combination of the inverted latent representation $\boldsymbol{z}_L$ and the corresponding class mean $\boldsymbol{z}^c_\mu$: $\boldsymbol{z}_p = \alpha \boldsymbol{z}_L + (1-\alpha)\boldsymbol{z}^c_\mu$. α is indeed sampled from $Beta(\gamma, \rho)$. As we increase γ, $\boldsymbol{z}_p$ will be closer to the inversed latents, while increasing ρ the class means will have more importance. When $\gamma = \rho = 1$, we get uniform samples between 0 and 1. We experiment with these three configurations in Tab. 6. For the simple CIFAR-10 dataset, we see negligible effect on varying α. On the harder CIFAR-100 dataset, we see that sampling α uniformly or closer to the class means helps to retain past knowledge without affecting class discovery. This result emphasises the fact that class means imparts semantic information and its interpolation with inversed latents provides diverse pseudo latents.

Table 2: The mixing coefficient α is sampled from $Beta(\gamma, \rho)$. We vary γ, ρ to measure the sensitivity of our method to the degree of mixing.

$\alpha \sim Beta(\gamma, \rho)$		CIFAR-10-5-5			CIFAR-100-80-20			CIFAR-100-50-50			CIFAR-100-20-80		
γ	ρ	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All
1	1	92.92	90.93	91.93	64.6	76.17	70.39	73.74	55.22	64.48	84.75	47.07	65.91
100	1	92.52	91.31	91.92	64.8	12.76	38.78	74.06	42.95	58.51	84.95	40.33	62.64
1	100	92.72	90.32	91.52	65.03	77.03	71.03	73.18	55.66	64.42	84.8	49.67	67.24