Supplementary Material of
"Modeling Inter-Class and Intra-Class Constraints in Novel Class Discovery"

Wenbin Li[1], Zhichen Fan[1], Jing Huo[1]*, Yang Gao[1]
[1]State Key Laboratory for Novel Software Technology, Nanjing University, China

1. Unknown Number of Novel Classes

The results reported in the main paper are based on a common assumption, *i.e.*, the number of novel classes C^u is known a priori, in NCD. However, in many real applications, it is somewhat difficult to obtain this priori in advance. That is to say, when facing with a real unlabelled dataset, we may cannot foretell the number of clusters accurately. To address this issue, many previous NCD works relax this restriction by using an estimation algorithm proposed in DTC [3] to estimate the number of unlabelled classes before discovering novel classes. For example, DualRank [4] and UNO [1] are both in this way. To be specific, DTC splits a probe subset from the labelled dataset, and then runs a semi-supervised k-means algorithm multiple times on the union of the probe subset and the unlabelled dataset by varying the number of clusters. The optimal number of novel classes is finally obtained by optimizing the cluster quality indices on the probe subset and unlabelled dataset.

In order to further demonstrate the effectiveness of our method in a more realistic setting, we first use the aforementioned estimation algorithm to find an estimation of novel classes on the CIFAR100-20 dataset split, denoted by $\hat{C^u}$. Next, we compare our proposed method with UNOv2 using the estimated number of novel classes and report the final clustering results in Tab. 1. It can be found that our method consistently outperforms UNOv2 and other comparison methods by a large margin and our method can achieve more stable performance (*i.e.*, with smaller standard deviations), when the ground truth number of novel classes is unknown.

Table 1. Experimental results with the estimated number of novel classes on CIFAR100-20, using the task-aware evaluation protocol. We also report the result with the true number of unlabelled classes. Results are reported in clustering accuracy (%) in form of mean and standard deviation (averaged over 3 runs). For simplicity, we fix $\alpha = 0.05$ and $\beta = 0.01$. **Best** results are highlighted in each column. †Our reproduced result.

Method	Class number	
	$C^u = 20$	$\hat{C^u} = 23$
DTC [3]	67.3±1.2	64.3
RS [2]	73.2±2.1	70.5
RS+ [2]	75.2±4.2	71.2
UNOv1 [1]	85.0±0.6	75.1
UNOv2† [1]	90.5±0.7	71.5±1.8
IIC (Ours)	**92.4±0.2**	**85.1±0.9**

2. Varying the Number of Clusters

To verify the effectiveness of the proposed method in more complex scenarios, we consider dataset splits with multiple different levels of difficulty and conduct additional experiments in Tab. 2. In the similar way of dividing CIFAR100 into CIFAR100-20 and CIFAR100-50 in the main paper, we adjust the number of classes contained in the labelled and unlabelled subsets to obtain various different dataset splits, such as CIFAR100-30 and CIFAR100-80.

As shown in Tab. 2, we compare the proposed method with UNOv2 [1] using an increasing number of unlabelled classes on CIFAR100 and report the corresponding task-aware clustering accuracy. Despite the gradual decrease in performance of both methods with increasing difficulty of the NCD task, our method consistently achieves superior results compared with UNOv2, with a minimum improvement of 1.2% in ACC. In this sense, we are able to demonstrate that the proposed method can achieve solid performance even facing with more challenging NCD tasks.

3. Speeding up the Computation of Inter-Class Constraint

From the experimental results and analyses in the main paper, it can be concluded that our proposed symmetric Kullback-Leibler divergence (sKLD) based constraints are simple yet effective. However, the calculation speed will be very slow if directly using the equations mentioned in Section 3.2, especially when the number of samples is large. It can be found that the reason for the enormous time cost

*Corresponding author

Table 2. Experimental results with an increasing number of unlabelled classes on CIFAR100, using the task-aware evaluation protocol. Results are reported in clustering accuracy (%) in form of mean and standard deviation (averaged over 3 runs). For simplicity, we fix the hyperparameters $\alpha = 0.05$ and $\beta = 0.01$ for the inter-class constraint and the intra-class constraint, respectively. **Best** results are highlighted in each column. †Our reproduced result.

Method	#Unlabelled classes						
	20	30	40	50	60	70	80
UNOv2† [1]	90.5±0.7	76.7±2.0	66.8±1.7	62.3±1.4	57.9±0.9	56.0±1.1	56.2±1.0
IIC (Ours)	**92.4±0.2**	**83.8±1.6**	**71.1±1.4**	**65.8±0.9**	**60.6±0.4**	**57.7±1.0**	**57.4±0.9**

is that calculating the inter-class sKLD loss term requires nested loops with the division operation (Eq. (1) - Eq. (4) in main paper). Therefore, we have to make additional designs for the objective of inter-class constraint to speed up the computation. Taking Eq. (2) as an example, we can simply rewrite the fraction into a subtraction formula as

$$\begin{aligned} D_{\mathrm{KL}}(\boldsymbol{p}_i^l \| \boldsymbol{p}_j^u) &= \sum_{k=1}^{C^l+C^u} \boldsymbol{p}_i^l(k) \log \frac{\boldsymbol{p}_i^l(k)}{\boldsymbol{p}_j^u(k)} \\ &= \sum_{k=1}^{C^l+C^u} \boldsymbol{p}_i^l(k) (\log \boldsymbol{p}_i^l(k) - \log \boldsymbol{p}_j^u(k)) \\ &= \sum_{k=1}^{C^l+C^u} \boldsymbol{p}_i^l(k) \log \boldsymbol{p}_i^l(k) - \sum_{k=1}^{C^l+C^u} \boldsymbol{p}_i^l(k) \log \boldsymbol{p}_j^u(k). \end{aligned}$$

In this more efficient way, the inter-class sKLD loss term over a mini-batch can be calculated faster with the matrix multiplication on GPU, promoting our method to be better applied on large-scale datasets (*e.g.*, ImageNet).

References

[1] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9284–9292, 2021. 1, 2

[2] Kai Han, Sylvestre-Alvise Rebuffi, Sébastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020. 1

[3] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8401–8409, 2019. 1

[4] Bingchen Zhao and Kai Han. Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 22982–22994, 2021. 1