Supplementary Material

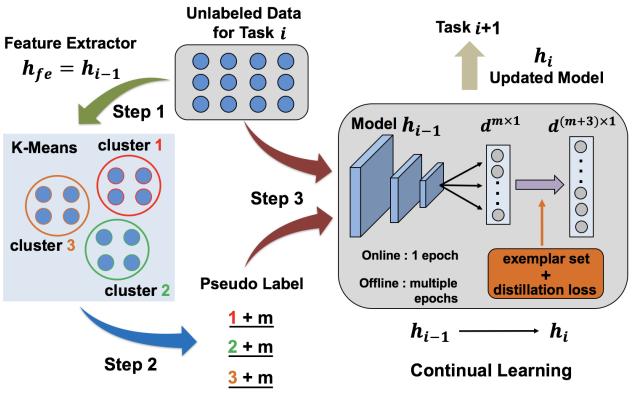

Figure 1. **Overview of the baseline solution to learn the new task i. h** refers to the model in different steps and **m** denotes the number of learned classes so far after task $i-1$. Firstly, the $\mathbf{h}_{i-1}$ (except the last fully connected layer) is applied to extract feature embeddings used for K-means clustering where the number **1**, **2**, **3** denote the corresponding cluster assignments. In step 2, the pseudo labels are obtained as **1+m**, **2+m**, **3+m** respectively. Finally in step 3, the unlabeled data with pseudo label is used together for continual learning without requiring human annotations.

1. Implementation Detail

In this section, we provide the detail for methods implemented in experimental parts including (1) the method to perform unsupervised continual learning and (2) existing "post-hoc" OOD detection methods, which will be illustrated in Section 1.1 and Section 1.2, respectively.

1.1. Unsupervised Continual Learning

In this work, we apply the baseline method proposed in [1] to perform unsupervised continual learning as shown in Figure 1, which includes three main steps: (1) Apply K-means clustering [9] on extracted features for all new task data using lower layers of the continual learning model updated from last incremental step. (2) Obtain the pseudo labels based on the cluster assignments. (3) Perform unsupervised continual learning and maintain the learned knowledge by using exemplar set [10] and knowledge distillation loss [4].

In our experimental part, we follow the same setting to use ResNet-32 [2] for CIFAR-100 [6] as the backbone network. The batch size of 128 with SGD optimizer, the initial learning rate is 0.1. We train 120 epochs for each incremental step and the learning rate is decreased by 1/10 for every 30 epochs. Exemplar size is set as $2,000$.

1.2. OOD detection

In our experiments, three existing "post-hoc" OOD detection methods are used for comparisons including MSP [3], ODIN [7] and Energy Score [8].

MSP directly applies the trained classification network to uses the maximum of softmax probability as the confidence score to discriminate between in-distribution and out-of-distribution data, which is regarded a strong baseline. Specifically, for each input data $\mathbf{x}$, we obtain the softmax output using trained classification model $\mathcal{F}_c$. The confidence score is calculated as

$$Conf = Max(Softmax(\mathcal{F}_c(\mathbf{x})))$$

ODIN further improves the performance by introducing the temperature scaling and input data pre-processing. Specifically, they proposed to calculate softmax probability using output value scaled by temperature $T > 1$ and use the maximum as confidence score.

Besides, to increase the difference between in-distribution and out-of-distribution data, they pre-process each input data by adding small perturbation

$$\tilde{\mathbf{x}} = \mathbf{x} - \epsilon sign(-\nabla_{\mathbf{x}} log(\frac{\mathcal{F}_c(\mathbf{x})}{T}))$$

Then, the final confidence score is given by

$$Conf = Max(Softmax(\frac{\mathcal{F}_c(\tilde{\mathbf{x}})}{T}))$$

In our experiment, we use $\epsilon = 0.001$ and $T = 1,000$.

Energy Score is proposed to replace the maximum of softmax probability as the confidence score. Specifically, considering the dimension of output logits $\mathcal{F}_c(\mathbf{x})$ is K. The energy function is used to calculate the confidence score, which is defined as

$$Conf = E(\mathbf{x}) = -T \times log(\sum_{i}^{K} e^{\mathcal{F}_c(\mathbf{x})/T})$$

where we use $T = 1,000$ as temperature scaling.

Methods	Step size 5			Step size 10			Step size 20		
	AUROC↑	**AUPR**↑	**FPR95**↓	**AUROC**↑	**AUPR**↑	**FPR95**↓	**AUROC**↑	**AUPR**↑	**FPR95**↓
baseline	0.754	0.959	0.793	0.736	0.915	0.824	0.729	0.874	0.814
LUCIR + Pseudo Labels	0.762	0.963	0.786	0.742	0.921	0.819	0.734	0.879	0.806

Table 1. Average AUROC, AUPR and FPR95 on CIFAR-100 with step size 5, 10 and 20.

2. Additional Experimental Results

For all experiments shown in the paper, we apply the **baseline** in [1] to perform unsupervised continual learning as illustrated in 1.1. In this section, we will demonstrate that our proposed OOD detection method can work with other existing continual learning methods to achieve higher performance. Following [1], we apply **LUCIR [5] + Pseudo Labels** to perform unsupervised continual learning and the results are shown in Table 1.

References

[1] Jiangpeng He and Fengqing Zhu. Unsupervised continual learning via pseudo labels. *arXiv preprint arXiv:2104.07164*, 2021. 1, 2

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1

[3] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Proceedings of International Conference on Learning Representations*, 2017. 1

[4] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. *Proceedings of the NIPS Deep Learning and Representation Learning Workshop*, 2015. 1

[5] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 831–839, 2019. 2

[6] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[7] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *Proceedings of International Conference on Learning Representations*, 2018. 1

[8] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 2020. 1

[9] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982. 1

[10] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. iCaRL: Incremental classifier and representation learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017. 1