Online Continuous Generalized Category Discovery

- Supplementary Materials -

Keon-Hee Park[1], Hakyung Lee[2], Kyungwoo Song[2†], and Gyeong-Moon Park[1†]

[1] Kyung Hee University, Republic of Korea
{khpark,gmpark}@khu.ac.kr
[2] Yonsei University, Republic of Korea
{hakyunglee0417, kyungwoo.song}@yonsei.ac.kr

Table 1: Dataset configuration for the proposed online continuous generalized category discovery. S.Ratio denotes the ratio of samples.

Dataset	All Classes	Base Session	Inc Session (Known + Novel)
		# of Classes (S.Ratio)	# of Classes (S.Ratio)
CUB200	200	160 (0.8)	160 (0.2) + 40 (1.0)
FGVC-Aircraft	100	80 (0.8)	80 (0.2) + 20 (1.0)
CIFAR-100	100	80 (0.8)	80 (0.2) + 20 (1.0)

A Further Implementation Details

To clearly define the scenario configuration, Table 1 demonstrates the data configuration for the proposed online continuous generalized category discovery. The incremental session comprised samples from both known categories and novel categories. We conducted 3 simulations for all experiments using different random seeds and reported the averages.

B Details of Comparison Methods

FRoST is designed to address class incremental novel category discovery, where the labeled dataset and unlabeled dataset have no overlap. FRoST utilizes an auxiliary classifier to pseudo-label novel samples and requires prior knowledge of the novel categories to expand the classifier and initialize the auxiliary classifier. In our experiments, we refrained from providing prior knowledge to ensure a fair comparison. Instead, we relied on calculating the similarity between features and classifier nodes to discover novel categories using affinity propagation.

† Corresponding authors

Table 2: Further study of the layers containing LoRA on CUB200.

Layers	CUB200			
From → To	M_{all}	M_{old}	M_{new}	$F \downarrow$
$1 \rightarrow 5$	55.90	65.18	22.97	17.61
$3 \rightarrow 7$	**56.01**	**65.33**	21.36	**17.05**
$5 \rightarrow 9$	54.58	63.66	22.71	18.79
$7 \rightarrow 11$	55.19	64.19	22.86	18.34
8 → 12	54.97	62.98	**23.67**	19.61

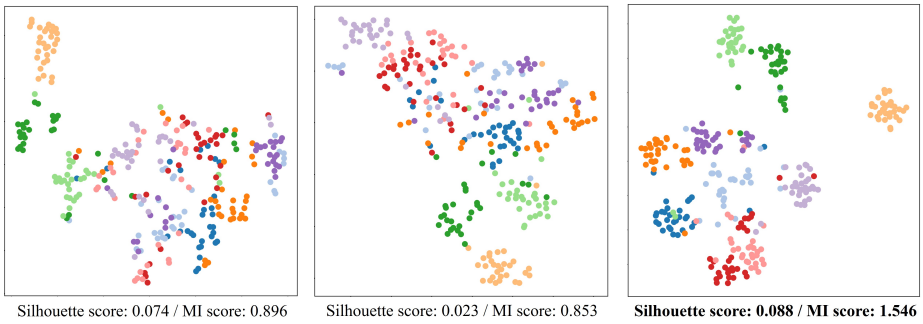
Table 3: Further experiment to evaluate the proposed online continuous generalized category discovery in various settings.

[Base / Incremental Sessions] [# of Classes(S.Ratio)]	Method	CUB200			
		M_{all}	M_{old}	M_{new}	$F \downarrow$
[180(0.9) / 180(0.1) + 20(1.0)]	FRoST [ECCV' 22]	43.61	47.32	10.71	33.63
	CGCD [ICCV' 23]	61.22	66.75	12.24	18.33
	DEAN (Ours)	**65.60**	**68.92**	**36.22**	**14.46**
[160(0.8) / 160(0.2) + 40(1.0)]	FRoST [ECCV' 22]	23.98	28.72	5.45	50.85
	CGCD [ICCV' 23]	40.40	47.88	11.18	35.61
	DEAN (Ours)	**59.56**	**66.33**	**33.11**	**16.26**
[140(0.7) / 140(0.3) + 60(1.0)]	FRoST [ECCV' 22]	32.61	44.64	5.16	34.89
	CGCD [ICCV' 23]	32.78	45.77	3.16	37.00
	DEAN (Ours)	**54.60**	**64.62**	**26.36**	**17.21**
[120(0.6) / 120(0.4) + 80(1.0)]	FRoST [ECCV' 22]	31.00	47.74	6.71	30.10
	CGCD [ICCV' 23]	27.18	41.94	5.77	40.57
	DEAN (Ours)	**47.76**	**64.00**	**23.77**	**17.22**
[100(0.5) / 100(0.5) + 100(1.0)]	FRoST [ECCV' 22]	27.12	50.23	4.53	28.84
	CGCD [ICCV' 23]	21.03	38.90	3.56	43.85
	DEAN (Ours)	**43.46**	**63.99**	**19.21**	**17.26**
[60(0.3) / 60(0.7) + 140(1.0)]	FRoST [ECCV' 22]	18.78	53.86	4.39	20.05
	CGCD [ICCV' 23]	13.73	31.76	6.33	45.82
	DEAN (Ours)	**27.27**	**56.88**	**15.12**	**19.38**

CGCD CGCD proposed continuous generalized category discovery, where the model continuously fine-tunes unlabeled datasets after training on a labeled dataset. CGCD assumes the overlapping between labeled and unlabeled datasets and does not require prior knowledge of novel categories. And, CGCD employs the proxy anchor-based contrastive loss for effective deep metric learning and requires a split network to categorize unlabeled data into old and novel categories.

C Further Analysis for the PET

We conducted an ablation study to determine the optimal position for parameter-efficient tuning. As shown in Table 2, while training LoRA from the 3rd layer to the 7th layer ($3 \rightarrow 7$) effectively mitigated severe forgetting, adopting these layers resulted in poor performance in novel category clustering. However, inserting



(a) FRoST [ECCV'22] **(b)** CGCD [ICCV'23] **(c) DEAN (Ours)**

Fig. 1: Comparison feature visualization for the novel categories on CUB200. We visualized 10 categories sequentially selected among the 40 novel categories.

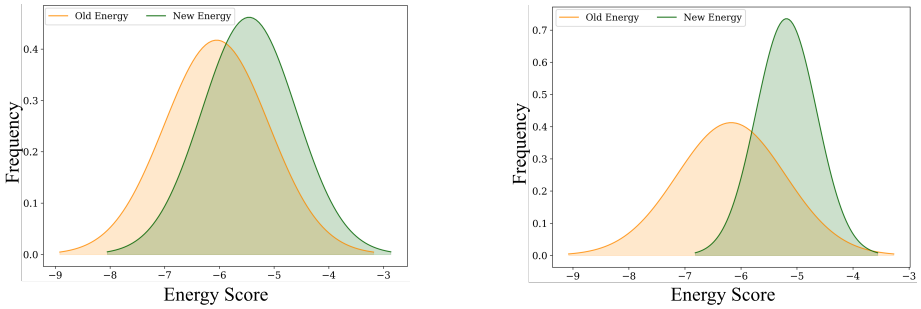
LoRA into the last five layers ($8 \rightarrow 12$) recorded the highest performance in novel category clustering. Based on this observation, we selected the last five layers to facilitate learning novel categories.

D Qualitative Evaluation of the Proposed Online Continuous Generalized Category Discovery

To evaluate the robustness of the proposed method, DEAN, we conducted experiments under various OCGCD settings. We scaled down the number of classes and samples in the base session following a specific ratio. As shown in Table 3, decreasing the number of classes and samples in the base session made the model struggle to discover novel categories. FRoST and CGCD suffered severe performance declination due to the decreasing proportions of classes and samples in the base session. However, DEAN consistently achieved outstanding performance across all measurements. This experimental result demonstrates that our method can maintain performance even when the model cannot be trained with sufficient data and classes during the base session.

E Visualization to Validate Novel Category Discovery

We compared feature visualizations among FRoST, CGCD, and the proposed DEAN to validate the clustering performance on CUB200. Additionally, we measured the silhouette score and mutual information-based score to compare the effectiveness of clustering. The silhouette score indicates the quality of clustering, with a higher score indicating better clustering. A mutual information-based score measures the agreement between clustering labels and ground truth labels. a higher score indicates more effective labeling through clustering. As shown in Figure 1, FRoST and CGCD faced challenges in clustering novel categories due to online continuous learning. Their performance metrics were notably lower



(a) w/o EC Loss **(b)** w/ EC Loss

Fig. 2: Comparing the normal distributions of energy scores using data estimated as unknown during the incremental session. Old energy and new energy denote the energy scores calculated by the known and unknown nodes from the online head, respectively.

Table 4: Ablation study of diverse variance values for variational feature augmentation. σ_l, σ_{B^t}, and σ_u denote the standard deviation of feature vectors from the entire labeled dataset, provided batch-wise data, and detected unseen data, respectively.

Variance σ^2	CUB200						
	M_{all}	M_{old}	M_{new}	$F\downarrow$	M_{all}^{PS}	M_{old}^{PS}	M_{new}^{PS}
w/o VFA	57.37	64.98	27.63	17.61	57.31	85.19	46.47
$\hat{h}_u^t \sim \mathcal{N}(h_u, \sigma_l^2)$	55.56	65.44	26.86	17.35	75.72	93.39	64.41
$\hat{h}_u^t \sim \mathcal{N}(h_u, \sigma_{B^t}^2)$	58.96	66.20	27.92	16.60	75.55	93.39	64.48
$\boldsymbol{\hat{h}_u^t \sim \mathcal{N}(h_u, \sigma_u^2)}$	**59.56**	**66.33**	**33.11**	**16.26**	**76.93**	**94.05**	**65.75**

compared to DEAN. For instance, DEAN achieved a mutual information-based score of 1.546, outperforming FRoST and CGCD significantly. This result highlights the ability of DEAN to effectively cluster novel categories and assign accurate labels compared to prior methods.

F Further Analysis of Energy-based Contrastive Loss

To validate the effectiveness of the proposed energy-based contrastive loss, we conducted an ablation study using the energy scores of data estimated as unknown during the incremental session. Specifically, we calculated two different energy scores: old energy, based on the known nodes, and new energy, based on the unknown nodes from the online head. As shown in Figure 2, the application of the energy-based contrastive loss resulted in a more discernible difference between old energy and new energy. This experimental result demonstrates that the proposed energy-based contrastive loss assists the model in detecting data from novel categories, thereby facilitating the learning of novel categories.

Table 5: Further experiment to evaluate the proposed online continuous generalized category discovery in various settings.

Augmentations	CUB200			
	M_{all}	M_{old}	M_{new}	$F \downarrow$
AutoAug [CVPR' 19]	57.90	65.76	27.20	16.28
TrivialAug [ICCV '21]	56.98	65.51	23.64	16.81
ME [CVPR' 21]	57.73	66.27	24.35	19.02
VFA (Ours)	**59.56**	**66.33**	**33.11**	**16.26**

Table 6: Comparison of VFA against existing augmentations.

G Further Analysis of Variational Feature Augmentation

We introduced variational feature augmentation to enhance pseudo-labeling via accurate clustering. To analyze its effectiveness, we conducted an ablation study with diverse variance values on CUB200. Variance values were calculated from feature vectors of the entire labeled dataset, the batch-wise data provided during online learning, and the data detected as unknown. As shown in Table 4, adopting variance values estimated from unknown data recorded superior clustering performance compared to other variance values. This experimental result demonstrates that leveraging variance information from estimated unknown data helps the model discover novel categories and improve pseudo-labeling accuracy. Moreover, We conducted a comparison study between our VFA and AutoAug [1], TrivialAug [3], and Moment Exchange (ME) [2] which do not require label information as VFA does. Table 6 shows the promising performance of VFA, while the others faced significant shifts in feature space due to unstable online learning, hindering effective clustering.

References

1. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
2. Li, B., Wu, F., Lim, S.N., Belongie, S., Weinberger, K.Q.: On feature normalization and data augmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2021)
3. Müller, S.G., Hutter, F.: Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 774–782 (2021)