Online Continuous Generalized Category Discovery

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

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

Abstract. With the advancement of deep neural networks in computer vision, artificial intelligence (AI) is widely employed in real-world applications. However, AI still faces limitations in mimicking high-level human capabilities, such as novel category discovery, for practical use. While some methods utilizing offline continual learning have been proposed for novel category discovery, they neglect the continuity of data streams in real-world settings. In this work, we introduce *Online Continuous Generalized Category Discovery* (**OCGCD**), which considers the dynamic nature of data streams where data can be created and deleted in real time. Additionally, we propose a novel method, **DEAN**, **D**iscovery via **E**nergy guidance and feature **A**ugmentatio**N**, which can discover novel categories in an online manner through energy-guided discovery and facilitate discriminative learning via energy-based contrastive loss. Furthermore, DEAN effectively pseudo-labels unlabeled data through variance-based feature augmentation. Experimental results demonstrate that our proposed DEAN achieves outstanding performance in proposed OCGCD scenario. The implementation code is available at https://github.com/KHU-AGI/OCGCD.
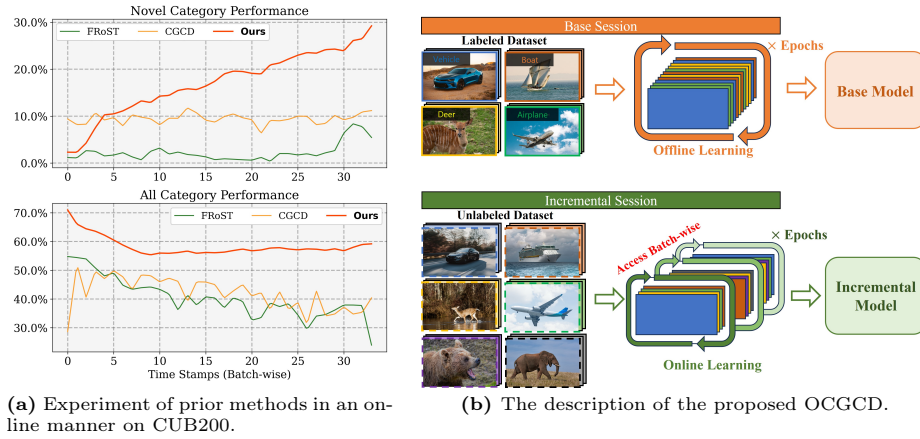
Keywords: Online Continual Learning · Generalized Category Discovery · Energy-Guided Discovery · Variance-based Feature Augmentation

1 Introduction

The rapid progress of deep neural networks in computer vision has greatly facilitated the widespread adoption of artificial intelligence into real-world applications. Nevertheless, the networks still have challenges in embodying high-level human capabilities such as object recognition and grouping. To emulate these sophisticated human abilities, there has been a lot of interest in the field of category discovery. Specifically, generalized category discovery (GCD) [40] demands

---
[†] Corresponding authors

(a) Experiment of prior methods in an online manner on CUB200.

(b) The description of the proposed OCGCD.

Fig. 1: Figure 1a shows that existing methods recorded poor performance in online training, suggesting that prior methods cannot handle online continual learning. Figure 1b shows the proposed scenario, OCGCD. As our scenario assumes batch-wise online learning, the model suffers severe forgetting and poor novel category discovery.

advanced recognition and grouping skills to discover unknown categories within unlabeled datasets. In general, existing methods for GCD rely on an impractical assumption that they can access both labeled and unlabeled datasets simultaneously during training, which prevents these methods from meeting real-world application demands.

In this context, a new scenario of continuous category discovery [19,22,37,42, 44] has emerged to address the unrealistic assumption by adopting offline continual learning. Despite its great advantage not requiring the assumption, these offline-based methods have inherent limitations in handling the continuous nature of data streams in the real world. Current approaches to offline continual learning primarily focus on static data streams, thus failing to adequately accommodate the dynamic nature of data streams where data can be created and deleted in an online manner. To verify this limitation, we conducted an empirical analysis using the CUB200 [41] dataset, dividing it into 160 labeled categories and 40 unlabeled categories. We evaluated the performance of existing methods [22,37] for continuous category discovery in an online manner. As shown in Figure 1a, we observed that existing methods [22,37] exhibited inferior performance in identifying novel categories, thereby resulting in poor overall performances. This indicates that the offline-based GCD methods have difficulty in clustering novel data effectively in the online learning scenario.

In light of this observation, in this paper, we introduce a novel online learning scenario in continuous generalized category discovery, called **Online Continuous Generalized Category Discovery (OCGCD)**. As shown in Figure 1b, OCGCD involves batch-wise training of unlabeled data in an online continual manner without any prior knowledge of the unknown data. Therefore, our novel

scenario closely replicates real-world data streams characterized by dynamic creation and elimination of data during training [14]. Here we summarize the key challenges of our novel OCGCD scenario as follows: 1) *Challenging category discovery.* The model struggles to retain the knowledge of trained categories because of the continuous nature of the data, making novel category discovery challenging. 2) *Noisy pseudo-labeling.* Online learning restricts access to the entire training dataset, leading to inaccurate category assignments and the generation of severe noise pseudo-labels for unlabeled data. 3) *Severe catastrophic forgetting.* Online learning erases observed data after training, resulting in severe forgetting of previously acquired knowledge [33]. Hence, our novel OCGCD scenario requires addressing the aforementioned challenges to achieve successful learning outcomes.

To address the challenges of OCGCD, we propose a novel approach called **DEAN**, **D**iscovery via **E**nergy Guidance and Feature **A**ugmentatio**N**, which utilizes the energy score [28] to discover unknown data and further discriminate whether the data is seen or unseen and employs feature augmentation for better pseudo-labeling. By comparing the energy values of known and unknown data, DEAN effectively identifies unknown data among them, facilitating the clustering of novel categories. On top of that, we introduce an energy-based contrastive loss to enhance learning from unknown samples, promoting the acquisition of more discriminative knowledge during online learning. In addition, we propose a novel variance-based feature augmentation, a simple yet effective approach for accurate novel clustering and pseudo-labeling. In our scenario, since the model can access only batch-wise training data at once, it struggles to cluster data accurately. Variance-based feature augmentation (VFA) temporally extends the number of feature vectors through variational information to promote accurate clustering. To mitigate catastrophic forgetting, we employ parameter-efficient tuning, especially LoRA [17] in our framework while preserving the pre-trained weights. Extensive experiments demonstrate that our proposed method, DEAN, enables the network to discover novel categories more effectively in the OCGCD scenario than existing methods.

We summarize the main contributions of this paper as follows:

- We introduce a novel Online Continuous Generalized Category Discovery (OCGCD) scenario that effectively reflects the online characteristics of real-world data. To tackle this scenario, we propose a novel method called Discovery via Energy Guidance and Feature AugmentatioN (DEAN).
- To the best of our knowledge, this is the first work to introduce the energy score for novel category discovery. We propose the energy-based contrastive loss to enhance online learning of discovered unknown data.
- For effective pseudo-labeling, we propose a new variance-based feature augmentation, called VFA. The proposed VFA enhances the sample clustering, leading to the improvement of pseudo-label quality.
- Comprehensive experiments demonstrate that our novel method achieves significant performance improvements in the OCGCD scenario compared to state-of-the-art models.

2   Related Work

2.1   Novel Category Discovery

Novel Category Discovery (NCD) focuses on the task of identifying and categorizing unlabeled data [39]. To effectively discover and classify data, this field requires advanced object recognition and clustering skills. Most of the NCD methods fall into two training scenarios: disjoint training and joint training. The disjoint training scenario, also known as a two-stage approach, starts by pre-training the model on a labeled dataset, followed by fine-tuning on an unlabeled dataset from a different domain [13, 15, 16, 29]. The joint training scenario adopts a more integrated approach, simultaneously utilizing both labeled and unlabeled datasets during the training process [7, 12, 18, 46, 47]. While this scenario outperforms disjoint training approach [29], its dependence on the availability of labeled data makes it less feasible in real-world applications. Accordingly, more research has been directed towards designing scenarios where they mitigate constraints of NCD for better alignment with real-world settings. Generalized Category Discovery (GCD) [40] broadens the setting of NCD by allowing test data to contain both known and unknown classes, in contrast to NCD which focuses on exclusively unknown classes. Moreover, to better represent real-world scenarios, several methods [19, 22, 37, 42, 44] have been proposed that integrate continual learning into category discovery. However, existing methods predominantly utilizes an offline continual learning framework, which remains an impractical assumption for real-world applications.

We propose an online version of continuous GCD that tackles the challenges of online learning through energy-guided clustering and energy-based contrastive loss. Energy-guided clustering utilizes the trained classifier to extract energy from each sample, efficiently discovering novel categories. Moreover, applying energy-based contrastive loss enables the model to effectively learn feature knowledge, thereby enhancing the capacity of the model in online continual learning.

2.2   Online Continual Learning

Online Continual Learning (OCL) is a more realistic continual learning scenario where the model learns from data arriving in small, sequential batches in real-time, without access to previously seen batches [30]. A key aspect of continual learning is to mitigate catastrophic forgetting and balance existing knowledge while adapting to new knowledge [8, 32, 33]. The dynamic distribution of online setting intensifies this challenge, and various approaches have been introduced to address this issue. MIR [1] proposed a memory retrieval technique for replay, selecting mostly interfered samples. OCM [10] deals with forgetting by maximizing mutual information between past and current data to learn more transferable features. i-Blurry [23] introduces a realistic setting that combines disjoint and blurry class distributions, mitigating forgetting by employing sample-wise importance sampling. Furthermore, Si-Blurry [34] proposed a new blurry task scenario, acknowledging the stochastic properties in real-world data distributions.

In this paper, we propose a novel setting that integrates online continual learning with generalized category discovery. By emphasizing the online learning, OCGCD presents a new perspective in the field of category discovery. It is a more realistic and challenging scenario than existing approaches in GCD, taking into account the continuous flow of real-world data.

2.3   Noise Label

Deep neural networks can overfit to noisy label data, leading to a substantial decrease in the robustness and generalization ability [2,43]. Consequently, tackling noisy label data becomes essential, especially in category discovery where pseudo-labeling of unlabeled data plays a pivotal role in identifying new categories. Recent studies have primarily focused on discriminating between clean data and noisy label data. DivideMix [26] use a Gaussian Mixture Model (GMM) generated from the training dataset to differentiate between clean and noisy data, while AugDesc [35] applies various data augmentations to the training dataset to amplify the distinction. SplitNet [21] introduces a compact learnable module that separates clean and noisy label data.

In OCGCD, learning occurs without prior knowledge of novel categories, relying only on the current batch-wise data to predict the number of clusters and generate pseudo-labels based on this prediction. This scenario makes it difficult to apply existing techniques that depend on utilizing the training dataset or employing learnable modules. Therefore, we propose Variance-based Feature Augmentation (VFA) to effectively separate clean and noisy labels in our framework. VFA can diversely augment (re-generate) the feature vectors of given data, enabling the creation of discriminative clusters for accurate pseudo-labeling. Unlike CGCD [22], which requires training additional modules for the discovery, VFA facilitates end-to-end training without the additional modules or extra learning.

3   Online Continuous Generalized Category Discovery

3.1   Problem Formulation and Method Overview

Problem Formulation. In online continuous GCD, the training dataset $\mathcal{D} = \{\mathcal{D}^{base}, \mathcal{D}^{inc}\}$ is given sequentially, where $\mathcal{D}^{base} = \{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}_l\}_{i=1}^{|\mathcal{D}^{base}|}$ is the labeled dataset for the base session and $\mathcal{D}^{inc} = \{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}_u\}_{i=1}^{|\mathcal{D}^{inc}|}$ is the unlabeled dataset for the incremental session. The proposed novel scenario assumes an overlap between the labeled category set $\mathcal{Y}_l$ and the unlabeled category set $\mathcal{Y}_u$, as it is a sub-category of the generalized category discovery. In this setting, the network trains on $\mathcal{D}^{base}$ and on $\mathcal{D}^{inc}$ in an offline and online manner, respectively. Our proposed online continuous GCD aims to facilitate the discovery of novel categories in real-world scenarios while preserving the clustering knowledge obtained from the offline training of the labeled dataset.
Method Overview. Figure 2 shows an overview of the proposed method, where the pre-trained vision transformer [6] via self-supervised learning serves as the
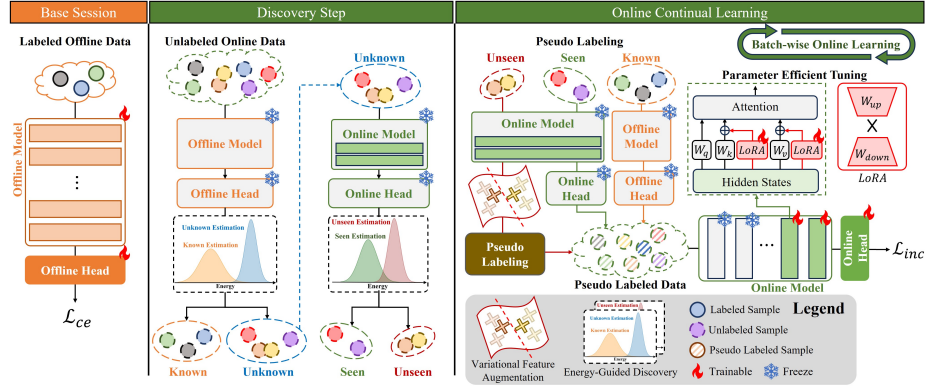
Fig. 2: Overall process of the proposed DEAN framework. The energy-guided discovery splits unlabeled data into known, seen, and unseen data for better novel category discovery, while variance-based feature augmentation enhances the clustering of unseen data. $\mathcal{L}_{ec}$ facilitates better discriminative learning in the online continual learning.
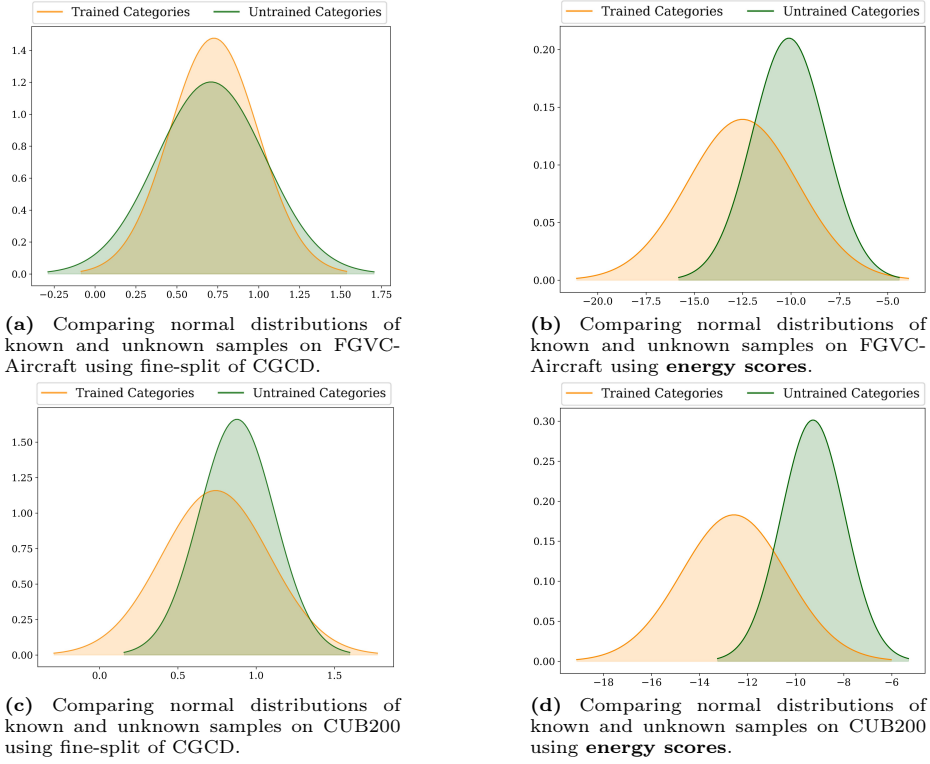
backbone network. We initially train the network using the labeled dataset in the base session (Section 3.2). We introduce energy-guided discovery to identify unknown samples from the given data. Our energy-guided discovery involves a two-step process to determine whether data belong to known or unknown categories and whether the unknown data belong to previously seen or unseen categories (Section 3.3). We split the data into known, seen, and unseen categories through energy-guided discovery. For pseudo-labeling of unseen data, we propose variance-based feature augmentation to improve the clustering of unseen data, resulting in better pseudo-labeling (Section 3.4). Subsequently, the network trains pseudo-labeled data in an online manner using an energy-based contrastive loss for better discriminative learning (Section 3.5).

3.2 Supervised Learning at Base Session

During the base session, we train the offline model θ^{off} consisting of the feature extractor $f^{off}(\cdot) : \mathcal{X} \to \mathbb{R}^d$ and classifier head $g^{off}(\cdot) : \mathbb{R}^d \to \mathcal{Y}_l$ in an offline manner. As the offline network $\theta^{off} = \{f^{off}, g^{off}\}$ trains on the labeled dataset during the base session, we employ the cross-entropy loss for supervised learning. While existing novel category discovery methods often utilize self-supervised [11] and supervised [20] contrastive loss for representation learning without a classifier head, our approach utilizes a classifier for clustering. The formulation of the base session training is as follows:

$$\mathcal{L}_{ce} = -\mathbb{E}_{(\mathcal{X}, \mathcal{Y}_l) \sim \mathcal{D}^{base}} \frac{1}{|\mathcal{D}^{base}|} \sum_{i=1}^{|\mathcal{D}^{base}|} y_i \cdot \log \delta(g^{off}(f^{off}(x_i))), \tag{1}$$

where $\delta(\cdot)$ is the softmax function. We train the model to capture representation knowledge during the base session training. After training the base session, we

(a) Comparing normal distributions of known and unknown samples on FGVC-Aircraft using fine-split of CGCD.

(b) Comparing normal distributions of known and unknown samples on FGVC-Aircraft using **energy scores**.

(c) Comparing normal distributions of known and unknown samples on CUB200 using fine-split of CGCD.

(d) Comparing normal distributions of known and unknown samples on CUB200 using **energy scores**.

Fig. 3: Validating the effectiveness of energy scores for novel category discovery without prior knowledge about novel categories by comparison with existing methods.

freeze the offline model and utilize the model at the discovery step to distinguish between known and unknown data, enabling accurate discovery. In addition, we initialize the online model, denoted as θ^{on}, based on the parameters of the offline model for the online continual learning on the unlabeled dataset.

3.3   Energy-Guided Discovery to Identify Unknowns

The discovery step aims to discover new categories within the provided batch-wise data. While CGCD, which assumes the absence of prior knowledge about novel categories, introduced a two-stage split process that depends on learning an additional split network to detect novel data, it has some limitations in our proposed scenario. CGCD includes the requirement for extra parameters and additional learning stage. Moreover, in our proposed online scenario where the model cannot access the entire training data, the split network faces difficulty in identifying novel samples due to severe forgetting of the split network. Liu *et al.* [28] proposed a method for detecting out-of-distribution (OOD) data by analyzing the difference in energy scores between in-distribution (ID) and OOD

data. We conducted experiments to compare the two-stage split process from CGCD with the energy-based discovery method to identify trained categories (known to the model) and untrained categories (unknown to the model).

In Figure 3, we found that the two-stage split of the CGCD struggled to identify both known and unknown categories in our proposed online learning scenario. Online learning with batch-wise data led to severe forgetting in the split network, resulting in poor detection of unknown samples. In contrast, energy-based discovery showed better performance in novel category discovery without extra parameters compared to CGCD. Moreover, as it does not require an additional learning phase, energy-based discovery enables end-to-end training. Inspired by this observation, we propose an energy-guided discovery approach for novel category discovery. To the best of our knowledge, this is the first work to utilize the energy score for novel category discovery.

During the incremental session, the network trains unlabeled batch-wise data $B^t = \{x_1, x_2, \ldots, x_{|B^t|}\}$, where t is the index of the batch. We classify unlabeled data B^t into known data, seen data, and unseen data using both the θ^{off} and θ^{on} through a two-stage process. In the first stage, we split the data into known and unknown categories. We obtain logits z using the offline model $\theta^{off}(\cdot) = g^{off}(f^{off}(\cdot))$ and calculate the energy scores using these logits. The formulation for calculating the energy scores is as follows:

$$z^t = \theta^{off}(B^t) \in \mathbb{R}^{|B^t| \times |\mathcal{Y}_l|}, \tag{2}$$

$$e^t = \{e_i^t | e_i^t = -\log \sum_{j=1}^{|\mathcal{Y}_l|} \exp(z_{i,j}^t)\}, \tag{3}$$

where $\exp(\cdot)$ denotes an exponential function, e^t denotes a set of energy scores, and $z_{i,j}$ denotes the logit from i-th sample corresponding to the j-th class. We divide the energy scores by the Gaussian Mixture Model (GMM). We initially set the GMM to have two clusters and use the energy score of each sample as input for clustering. We then label the cluster with a low energy score as "known" and the cluster with a high energy score as "unknown". This division allows us to split the unlabeled data into known and unknown data. In the second stage, we divide the unknown data into seen and unseen categories. Seen data comprise the data belonging to observed novel categories during the incremental session, while unseen data belong to undiscovered categories. To split the unknown data, we utilize the online model $\theta^{on}(\cdot) = g^{on}(f^{on}(\cdot))$ to extract energy scores. The process in the second stage is the same as the first stage, where we split the unknown data into seen and unseen categories based on their energy scores. For the initial batch of incremental sessions, the online model is identical to the offline model. Additionally, since the initial batch data is the first data of the incremental session, any data classified as unknown is assumed to be unseen data. The proposed energy-guided discovery splits unlabeled data into known, seen, and unseen categories effectively. Unlike prior methods which require sufficient data, it can identify novel categories with batch-wise data.

3.4   Pseudo-Labeling via Variance-based Feature Augmentation

After the discovery step, we pseudo-label known, seen, and unseen data to assign labels to each sample. For known and seen data, where the model has already learned knowledge for the categories, we utilize predictions from the offline model for known data and from the online model for seen data to assign pseudo-labels, respectively. Conversely, as the unseen data belong to undiscovered categories, the online model cannot predict proper categories for them and needs to discover novel categories. Moreover, due to the given batch-wise data, discovering novel categories to assign pseudo labels is challenging.

To address this challenge, we propose variance-based feature augmentation, which temporally augments feature vectors using variational knowledge from unseen data to acquire accurate novel clusters. We obtain feature vectors of the unseen data B_u^t using the online feature extractor $f^{on}(\cdot)$ denoted as $h_u^t = f^{on}(B_u^t)$. We calculate the variance of the unseen feature vectors and augment unseen feature vectors based on the variance. We employ each feature vector to augment features by the Gaussian distribution $\hat{h}_u^t = \{\hat{h}_u \sim \mathcal{N}(h_u, \sigma_u^2)\}$, $h_u \in h_u^t$. The formulation of the variance-based feature augmentation is as follows:

$$h_{aug}^t = \{h_u^t; \hat{h}_u^t\}, \quad \hat{h}_u^t = \{\hat{h}_u \sim \mathcal{N}(h_u, \sigma_u^2)\} \in \mathbb{R}^{(|B_u^t| \cdot K) \times d}, \tag{4}$$

where $\hat{h}_u$ and σ_u represent augmented feature vectors and the standard deviation of the unseen feature vectors, respectively. K denotes the hyperparameter for the number of augmented feature vectors, and h_{aug}^t denotes the concatenation of unseen feature vectors and augmented feature vectors. Novel categories are discovered based on h_{aug}^t. Augmented feature vectors expand the feature space with variational knowledge, which helps the model discover novel clusters.

We employ affinity propagation [9] to assign pseudo-labels to unseen data. Affinity propagation is a non-parametric clustering algorithm that does not require a predefined number of clusters, using the similarity matrix of the data. Given the absence of prior knowledge of novel categories, we extend the online classifier with the estimated number of novel clusters obtained by affinity propagation. The proposed variance-based feature augmentation is a simple yet effective approach that helps the model assign effective pseudo-labels based on accurate novel category discovery.

3.5   Online Continual Learning with Pseudo-Labeled Data

For online continual learning, we employ parameter-efficient tuning to learn novel representation knowledge while preserving previously learned knowledge, as it effectively mitigates catastrophic forgetting. We use cross-entropy loss with pseudo-labeled data in an online manner to capture novel knowledge. However, adapting the model with batch-wise data raises recency bias, where sufficient training through the given batch-wise data impedes discriminative knowledge. To address this, we propose an energy-based contrastive loss ($\mathcal{L}_{ec}$) for effective discriminative learning. The energy-based contrastive loss focuses on inducing

the online model to capture the knowledge of novel categories. It amplifies the energy from the known classifier while reducing the energy from the novel classifier. By selectively applying $\mathcal{L}_{ec}$ to novel data B_n^t including seen and unseen data, it helps the online model learn novel discriminative knowledge. The formulation of $\mathcal{L}_{ec}$ is as follows:

$$\mathrm{E}(f(x); g) = -\log \sum_{i=1}^{|\mathcal{Y}|} \exp(g_i(f(x))), \tag{5}$$

$$\mathcal{L}_{ec} = \frac{1}{|B_n^t|} \sum_{(x_n) \sim B_n^t} \log \left(1 + \frac{\mathrm{E}(f^{on}(x_n); g^{new})}{\mathrm{E}(f^{on}(x_n); g^{old})} \right), \tag{6}$$

where $\mathrm{E}(\cdot; \cdot)$ and $|\mathcal{Y}|$ denote energy function and the number of nodes at the classifier, respectively, $g_i(\cdot)$ represents the i-th node from the given classifier g, and g^{old} and g^{new} denote the nodes corresponding to known categories and discovered novel categories in the online classifier $g^{on}(\cdot)$, respectively. Eq. 5 represents the calculation of the energy score given the feature vector and classifier. The total loss $\mathcal{L}_{inc}$ for the incremental session can be summarized as follows:

$$\mathcal{L}_{inc} = \mathcal{L}_{ce} + \mathcal{L}_{ec}, \tag{7}$$

where $\mathcal{L}_{ec}$ represents energy-based contrastive loss, and $\mathcal{L}_{ce}$ denotes cross-entropy loss with pseudo-labeled data.

4 Experiments

4.1 Experimental Setup

Baselines and Implementation Details. Recently, many approaches have emerged in novel category discovery. However, many of them focus on self-supervised learning with non-parametric classifiers (*e.g.* K-means) [42] and rarely consider continual learning [45]. Some methods that consider continual learning assume prior knowledge for novel category discovery. Since the proposed EGD and fine-split from CGCD require a classifier to estimate the number of novel categories, non-parametric classifiers cannot be used. Additionally, existing methods not designed for continual learning are unsuitable baselines due to severe forgetting of known classes. For a fair comparison, we mainly compared our method with FRoST [37] and CGCD [22], which focus on continual learning and utilize parametric classifiers for novel category discovery.

We used ViT-B-16 [6] pre-trained on DINO-ImageNet [3] as the backbone network for all methods, including ours. We trained our method 30 epochs for the base session and 15 epochs for the incremental session. We utilized the AdamW optimizer with a weight decay of 1e-4 and a learning rate of 1e-3. We trained our model using a single RTX 3090 GPU with a batch size set to 64.

Datasets. We evaluated our method with state-of-the-art (SoTA) continuous category discovery methods on three datasets: CUB200 [41], FGVC-Aircraft [31], and CIFAR-100 [24]. For continual learning, we partitioned CUB200 and FGVC-Aircraft following the same split configuration in CGCD. For CIFAR-100, we divided 80 classes for known classes and 20 classes for novel classes. To follow the GCD setting, we divided the training samples into labeled and unlabeled per class, maintaining a ratio of 0.8 for the labeled to 0.2 for the unlabeled.

Evaluation Metrics. We evaluated the methods by measuring clustering accuracy on test dataset $\mathcal{D}$. We utilized the hungarian algorithm [25] to obtain the optimal permutation h^* that aligns the prediction y_i^* from the model with the ground truth label y_i:

$$h^* = \arg\min_h \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbb{I}(y_i = h(y_i^*)), \tag{8}$$

where $|\mathcal{D}|$ is the size of the test dataset. Clustering accuracy on $\mathcal{D}$ is defined as follows:

$$M = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbb{I}(y_i = h^*(y_i^*)). \tag{9}$$

We employed this metric to measure the clustering accuracy of all classes (M_{all}), known classes (M_{old}), and unknown classes (M_{new}) in the cases of test dataset $\mathcal{D}^{base}$ and $\mathcal{D}^{inc}$, respectively. Following continuous GCD approaches to evaluate the ability to maintain performance for known classes [22, 44], we employed the forgetting metric $F = M_{old}^{base} - M_{old}^{inc}$, where M_{old}^{base}, M_{old}^{inc} refers to known class clustering accuracy at the base session and incremental session. A higher clustering accuracy metric is better, and a lower forgetting metric is better. This ensures the model is effective for real-world applications.

Moreover, we evaluate the estimated pseudo-labels to validate their effectiveness in pseudo-labeling. Pseudo-labeling for unlabeled data is crucial as the model utilizes pseudo-labels to interpret the unlabeled data. The pseudo-labeling accuracy on the dataset $\mathcal{D}$ is defined as follows:

$$M = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbb{I}(y_i = h^*(\tilde{y}_i)), \tag{10}$$

where $\tilde{y}_i$ denotes the pseudo-label of the i-th sample.

4.2   Comparison with Baselines

We compared the clustering performance of our proposed method, DEAN, with SoTA approaches, as shown in Table 1. Since we considered an online scenario for category discovery, traditionally approached using offline methods, there is no previous work for comparison. Therefore, for a fair comparison, our study adapts

Table 1: Clustering performance comparison on CUB200, FGVC-Aircraft, and CIFAR-100. Best performance is highlighted in bold. All the baselines utilize the fine-split method from CGCD to handle the absence of prior knowledge.

Method	CUB200				FGVC-Aircraft				CIFAR-100			
	M_{all}	M_{old}	M_{new}	$F \downarrow$	M_{all}	M_{old}	M_{new}	$F \downarrow$	M_{all}	M_{old}	M_{new}	$F \downarrow$
Supervised	72.51	70.68	79.66	11.91	66.25	63.62	76.76	11.03	81.34	79.48	88.80	10.07
Fine-Tuning	1.96	1.94	2.01	80.52	2.06	1.30	5.10	73.35	3.68	3.24	5.45	86.26
FRoST [ECCV' 22]	23.98	28.72	5.45	50.85	14.44	15.65	9.60	48.31	6.13	6.21	5.80	69.24
CGCD[ICCV' 23]	40.40	47.88	11.18	35.61	14.08	15.67	7.75	60.00	2.64	2.76	2.15	86.74
DEAN w/ FS	55.57	65.37	17.29	17.62	40.32	49.14	5.10	25.43	52.27	62.11	12.90	27.55
DEAN (Ours)	**59.56**	**66.33**	**33.11**	**16.26**	**46.44**	**53.56**	**17.99**	**20.97**	**62.34**	**70.34**	**30.35**	**19.17**

Table 2: Ablation experiment on CUB200. The baseline denotes fine-tuning with the fine split from the CGCD.

Ablation	CUB200			
	M_{all}	M_{old}	M_{new}	$F \downarrow$
Baseline	1.96	1.94	2.01	80.52
+ LoRA	49.33	60.35	6.24	22.24
+ EGD	54.97	62.98	23.67	19.61
+ EC Loss	57.37	64.98	27.63	17.61
+ **VFA (Ours)**	**59.56**	**66.33**	**33.11**	**16.26**

Table 3: Ablation study for the hyper-parameter K of variance-based feature augmentation on CUB200.

Ablation	CUB200		
	M_{all}^{PS}	M_{old}^{PS}	M_{new}^{PS}
K=0	57.31	85.19	46.47
K=1	64.27	88.96	53.74
K=3	75.87	93.43	64.48
K=5	**76.93**	**94.05**	**65.75**
K=7	75.44	93.41	64.07
K=9	75.22	93.5	63.36

offline continuous NCD and GCD methods into OCGCD scenarios. The experiment includes two SoTA continuous category discovery methods: FRoST [37] and CGCD [22]. In addition, we include a supervised learning method as an upper bound and a fine-tuning method as a performance baseline. Note that in our proposed scenario, OCGCD, all baselines are not provided with prior knowledge about novel categories. Instead, we employed fine-split method from CGCD [22] which consists of two stages: initial split and fine split. During the initial split stage, unlabeled data is roughly divided into known and unknown categories based on cosine similarity. In the fine split stage, a split model is trained to distinguish between known and unknown data. The fine-split method estimates the number of novel categories conveying information about novel categories to the comparison methods.

The results shows that our method, DEAN, recorded a robust clustering performance, significantly surpassing existing continuous NCD and GCD methods. DEAN recorded consistently high metrics across M_{all}, M_{old}, and M_{new} while keeping forgetting metric F substantially low. This denotes our method can learn novel knowledge without forgetting. Moreover, it is noteworthy that DEAN w/ FS which denotes the adoption of fine-split from CGCD into DEAN instead

Table 4: Ablation study for the cluster accuracy of pseudo-labels.

Method	CUB200		
	M_{all}^{PS}	M_{old}^{PS}	M_{new}^{PS}
CGCD [ICCV' 23]	61.88	81.07	58.68
DEAN w/o ($\mathcal{L}_{ec}$, VFA)	57.23	85.11	45.80
DEAN w/o VFA	57.31	85.19	46.47
DEAN (Ours)	**76.93**	**94.05**	**65.75**

Table 5: Ablation experiment of parameter-efficient tuning on CUB200.

Method		CUB200			
		M_{all}	M_{old}	M_{new}	$F\downarrow$
DEAN (Ours)	Adapter [4]	56.96	64.75	26.53	17.64
	Prefix [27]	58.54	66.29	28.28	16.30
	LoRA [17]	**59.56**	**66.33**	**33.11**	**16.26**

of energy-guided discovery (EGD) also showed notable performance compared to the baselines. But, DEAN achieved significant performance gains across all datasets as the proposed EGD is applied. In contrast, existing continual methods showed shortcomings in clustering performance. Specifically, we observed extremely low clustering performance on CIFAR-100 dataset, low M_{all}, M_{old}, and M_{new} values, coupled with high F. On CIFAR-100 dataset, DEAN surpassed the most recent method CGCD assuming no prior knowledge for novel categories by 59.7%, 67.58%, 28.2%, and 67.57% for M_{all}, M_{old}, M_{new}, and F.

Such observations shows that while current GCD methods are vulnerable to online settings, DEAN can manage the balance required in the OCGCD scheme of retaining old knowledge while integrating the new. We believe this results shows the suitability of DEAN for real-world applications, where data is never static. We empirically confirmed that DEAN can evolve with the data it encounters, overcoming the challenges presented in the OCGCD scheme.

4.3 Ablation Study

The experimental results in Section 4.2 show that our DEAN achieved significant performance in OCGCD scenarios. Through further experiments on the CUB200, we provide a detailed analysis of the components proposed in our method.

Effectiveness of Components. We validate the effectiveness of adding each component of our method to the baseline model. The baseline denotes fine-tuning the backbone model with a data-split following settings in CGCD. As shown in Table 2, results reveal that all components have an important impact on performance. Employing LoRA to baseline shows significant improvement in M_{old}, F. This shows that our application of LoRA on the frozen model at incremental session effectively mitigates forgetting. While LoRA shows only a slight increase in M_{new}, additionally employing energy-guided discovery (EGD) and EC Loss markedly improves M_{new}. This confirms that adopting energy guidance will enhance the ability of the model to differentiate between old and new categories. Finally, the integration of VFA further boosts the model performance on M_{new}. This improvement reveals the importance of accurate pseudo-labeling since VFA is a technique applied after training. Each component is crucial, incrementally building upon the last to solidify the model's performance in our online category discovery scenario.

Hyperparameter Analysis of VFA. Variance-based Feature Augmentation (VFA) is our proposed method of temporarily augmenting feature vectors to acquire accurate pseudo-labeling. Our analysis evaluated the effect of hyperparameter K, the number of augmented feature vectors, on the accuracy of pseudo-labels. From Table 3 we observed that employment of VFA is beneficial, increasing novel category pseudo-label accuracy from 46.47% to 65.75%. Among VFA employment, we observed that K = 5 yields the best performance across all metrics. When K is less than 5, the smaller K worsened performance. Conversely, values of K greater than 5 did not benefit the model. We analyzed that since VFA augments features based on variance, as the number of K increases, unseen samples can be clustered together without being distinguished.

Pseudo-Labeling Accuracy. Accurate pseudo-labeling for clustering is important because it directly influences the performance of the model in incremental learning. Shown in Table 4, our proposed method, DEAN, significantly enhanced the clustering accuracy of pseudo-labels, achieving a 13.12% improvement in M_{all}^{PS} over the most recent method CGCD. Through the ablation study, we investigated the contributions of each step in DEAN. When DEAN adopts only parameter-efficient tuning, excluding EC loss and VFA, we observed a notable declination in M_{new}^{PS}. While each ablation step showed gradually accuracy enhancement, the application of VFA elevates the performance of DEAN above CGCD. Thus, the experimental result revealed that VFA played a vital role in distinguishing novel categories and confirming the effectiveness of our method.

Parameter Efficient Tuning. The ablation study in Table 5 assesses the impact of parameter-efficient tuning (PET) methods on the CUB200 dataset. Our method can employ various tuning methods such as Adapter [4] or Prefix-tuning [27], and in this experiment, we compared modules with LoRA [17], Adapter and Prefix-tuning. We set the size of all PET modules size to 5 and integrated them into the last 5 layers of ViT. Overall improvements are shown by adopting PET over existing methods reported in Table 1. Among PET methods, LoRA showed the highest clustering accuracy. The performance enhancement of LoRA may stem from its fast adaptation, which is highly effective in online continual learning scenarios that often depend on fewer samples for training.

5   Conclusion

In this paper, we introduced a novel scenario for category discovery, Online Continuous Generalized Category Discovery, which accounts for the dynamic nature of real-world data streams, leading to severe forgetting and poor novel category discovery. In addition, we proposed a novel framework, DEAN, which utilizes energy-based novel category discovery for the first time in this field and introduces variance-based feature augmentation to enhance accurate pseudo-labeling. Our DEAN demonstrated promising performance in the proposed novel scenario. However, DEAN could not guarantee inference performance midway through training. In future work, we plan to explore an online continual framework that can ensure inference performance during training through fast adaptation.

Acknowledgements

References

1. Aljundi, R., Belilovsky, E., Tuytelaars, T., Charlin, L., Caccia, M., Lin, M., Page-Caccia, L.: Online continual learning with maximal interfered retrieval. Advances in Neural Information Processing Systems **32** (2019)
2. Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., et al.: A closer look at memorization in deep networks. In: International conference on machine learning. pp. 233–242. PMLR (2017)
3. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the International Conference on Computer Vision (ICCV) (2021)
4. Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., Luo, P.: Adaptformer: Adapting vision transformers for scalable visual recognition. Advances in Neural Information Processing Systems **35**, 16664–16678 (2022)
5. Choe, S.A., Shin, A.H., Park, K.H., Choi, J., Park, G.M.: Open-set domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23943–23953 (2024)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
7. Fini, E., Sangineto, E., Lathuiliere, S., Zhong, Z., Nabi, M., Ricci, E.: A unified objective for novel class discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9284–9292 (2021)
8. French, R.M.: Catastrophic forgetting in connectionist networks. Trends in cognitive sciences **3**(4), 128–135 (1999)
9. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. science **315**(5814), 972–976 (2007)
10. Guo, Y., Liu, B., Zhao, D.: Online continual learning through mutual information maximization. In: International Conference on Machine Learning. pp. 8109–8126. PMLR (2022)
11. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 297–304. JMLR Workshop and Conference Proceedings (2010)

12. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Automatically discovering and learning new visual categories with ranking statistics. arXiv preprint arXiv:2002.05714 (2020)
13. Han, K., Vedaldi, A., Zisserman, A.: Learning to discover novel visual categories via deep transfer clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8401–8409 (2019)
14. He, J., Mao, R., Shao, Z., Zhu, F.: Incremental learning in online scenario. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13926–13935 (2020)
15. Hsu, Y.C., Lv, Z., Kira, Z.: Learning to cluster in order to transfer across domains and tasks. arXiv preprint arXiv:1711.10125 (2017)
16. Hsu, Y.C., Lv, Z., Schlosser, J., Odom, P., Kira, Z.: Multi-class classification without multi-class labels. arXiv preprint arXiv:1901.00544 (2019)
17. Hu, E.J., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2021)
18. Jia, X., Han, K., Zhu, Y., Green, B.: Joint representation learning and novel category discovery on single-and multi-modal data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 610–619 (2021)
19. Joseph, K., Paul, S., Aggarwal, G., Biswas, S., Rai, P., Han, K., Balasubramanian, V.N.: Novel class discovery without forgetting. In: European Conference on Computer Vision. pp. 570–586. Springer (2022)
20. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020)
21. Kim, D., Ryoo, K., Cho, H., Kim, S.: Splitnet: learnable clean-noisy label splitting for learning with noisy labels. arXiv preprint arXiv:2211.11753 (2022)
22. Kim, H., Suh, S., Kim, D., Jeong, D., Cho, H., Kim, J.: Proxy anchor-based unsupervised learning for continuous generalized category discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16688–16697 (2023)
23. Koh, H., Kim, D., Ha, J.W., Choi, J.: Online continual learning on class incremental blurry task configuration with anytime inference. arXiv preprint arXiv:2110.10031 (2021)
24. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
25. Kuhn, H.W.: The hungarian method for the assignment problem. Naval research logistics quarterly **2**(1-2), 83–97 (1955)
26. Li, J., Socher, R., Hoi, S.C.: Dividemix: Learning with noisy labels as semi-supervised learning. arXiv preprint arXiv:2002.07394 (2020)
27. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 4582–4597 (2021)
28. Liu, W., Wang, X., Owens, J., Li, Y.: Energy-based out-of-distribution detection. Advances in neural information processing systems **33**, 21464–21475 (2020)
29. Liu, Y., Tuytelaars, T.: Residual tuning: Toward novel category discovery without labels. IEEE Transactions on Neural Networks and Learning Systems (2022)
30. Mai, Z., Li, R., Jeong, J., Quispe, D., Kim, H., Sanner, S.: Online continual learning in image classification: An empirical survey. Neurocomputing **469**, 28–51 (2022)

31. Maji, S., Rahtu, E., Kannala, J., Blaschko, M., Vedaldi, A.: Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151 (2013)
32. McClelland, J.L., McNaughton, B.L., O'Reilly, R.C.: Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review **102**(3), 419 (1995)
33. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks: The sequential learning problem. In: Psychology of learning and motivation, vol. 24, pp. 109–165. Elsevier (1989)
34. Moon, J.Y., Park, K.H., Kim, J.U., Park, G.M.: Online class incremental learning on stochastic blurry task boundary via mask and visual prompt tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11731–11741 (2023)
35. Nishi, K., Ding, Y., Rich, A., Hollerer, T.: Augmentation strategies for learning with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8022–8031 (2021)
36. Park, K.H., Song, K., Park, G.M.: Pre-trained vision and language transformers are few-shot incremental learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23881–23890 (2024)
37. Roy, S., Liu, M., Zhong, Z., Sebe, N., Ricci, E.: Class-incremental novel class discovery. In: European Conference on Computer Vision. pp. 317–333. Springer (2022)
38. Seo, J., Lee, S.H., Lee, T.Y., Moon, S., Park, G.M.: Generative unlearning for any identity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9151–9161 (2024)
39. Troisemaine, C., Lemaire, V., Gosselin, S., Reiffers-Masson, A., Flocon-Cholet, J., Vaton, S.: Novel class discovery: an introduction and key concepts. arXiv preprint arXiv:2302.12028 (2023)
40. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7492–7501 (2022)
41. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
42. Wu, Y., Chi, Z., Wang, Y., Feng, S.: Metagcd: Learning to continually learn in generalized category discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1655–1665 (2023)
43. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning (still) requires rethinking generalization. Communications of the ACM **64**(3), 107–115 (2021)
44. Zhang, X., Jiang, J., Feng, Y., Wu, Z.F., Zhao, X., Wan, H., Tang, M., Jin, R., Gao, Y.: Grow and merge: A unified framework for continuous categories discovery. Advances in Neural Information Processing Systems **35**, 27455–27468 (2022)
45. Zhao, B., Han, K.: Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. Advances in Neural Information Processing Systems **34**, 22982–22994 (2021)
46. Zhong, Z., Fini, E., Roy, S., Luo, Z., Ricci, E., Sebe, N.: Neighborhood contrastive learning for novel class discovery. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10867–10875 (2021)
47. Zhong, Z., Zhu, L., Luo, Z., Li, S., Yang, Y., Sebe, N.: Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9462–9470 (2021)