This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Variable Few Shot Class Incremental and Open World Learning

Touqeer Ahmad   Akshay Raj Dhamija   Mohsen Jafarzadeh
Steve Cruz   Ryan Rabinowitz   Chunchun Li   Terrance E. Boult
Vision and Security Technology Lab, University of Colorado Colorado Springs, Colorado Springs, CO, USA
{touqeer, adhamija, mjafarzadeh, scruz, rrabinow, cli, tboult}@vast.uccs.edu

Abstract

Prior work on few-shot class incremental learning has operated with an unnatural assumption: the number of ways and number of shots are assumed to be known and fixed e.g., *10-ways 5-shots, 5-ways 5-shots, etc. Hence, we refer to this setting as Fixed-Few-Shot Class Incremental Learning (FFSCIL). In practice, the pre-specified fixed number of classes and examples per class may not be available, meaning one cannot update the model. Evaluation of FSCIL approaches in such unnatural settings renders their applicability questionable for practical scenarios where such assumptions do not hold. To mitigate the limitation of FFSCIL, we propose Variable-Few-Shot Class Incremental Learning (VFSCIL) and demonstrate it with* ***Up-to N-Ways, Up-to K-Shots*** *class incremental learning; wherein each incremental session, a learner may have up to* ***N*** *classes and up to* ***K*** *samples per class. Consequently, conventional FFSCIL is a special case of herein introduced VFSCIL. Further, we extend VFSCIL to a more practical problem of Variable-Few-Shot Open-World Learning* ***(VFSOWL)****, where an agent is not only required to perform incremental learning, but must detect unknown samples and enroll only those that it detects correctly. We formulate and study VFSCIL and VFSOWL on two benchmark datasets conventionally employed for FFSCIL* i.e., *Caltech-UCSD Birds-200-2011 (CUB200) and miniImageNet. First, to serve as a strong baseline, we extend the state-of-the-art FSCIL approach to operate in* ***Up-to N-Ways, Up-to K-Shots*** *class incremental and open-world settings. Then, we propose a novel but simple approach for VFSCIL/VFSOWL where we leverage the current advancements in self-supervised feature learning. Utilizing both benchmark datasets, our proposed approach outperforms the strong baseline on the conventional FFSCIL setting and newly introduced VFSCIL/VFSOWL settings. Our code is available at:* https://github.com/TouqeerAhmad/VFSOWL

Images Per Class	C101	C102	C103	C104	C105	5-Ways, 5-Shots	Up-to 5-Ways, Up-to 5-Shots
	0	8	9	7	0	**WAIT**	**Train 102,103,104**
	6	10	12	10	0	**WAIT**	**Train 101 + improve**
	8	15	20	15	4	**WAIT**	**Train 105 +improve**
	10	18	25	22	5	**Throw away most data and Train**	**Optionally Improve**
	Waste 5	*Waste 13*	*Waste 20*	*Waste 17*	*Waste 0*		

Figure 1. Operational example for Class Incremental Learning: *Fixed Few Shot (FFSCIL)* vs *Variable Few Shot (VFSCIL)*. Learning with a small amount of data is often critical, leading to increasing interest in few-shot learning. In this example, each row shows the number of images from unknown classes arriving each month. In an FFSCIL (say 5-ways 5-shots) setting, a learning agent must keep waiting 3 time periods until at least 5 classes have at least 5 images each – it then throws away the additional data and finally can start incremental learning. In the more realistic VFSCIL, a learning agent can learn when it has sufficient data and can exploit what data it has at that time, providing maximum utilization of available data. VFSCIL is still optimistic because it presumes someone other than the underlying learning system identified the unknown data to add. In addition to introducing VFSCIL, this paper also introduces *Variable Few Shot Open World Learning (VFSOWL)* where a system must identify unknown inputs, and only then can it learn them.

1 Introduction

As more and more deep models are being deployed in the real world, and computer vision keeps migrating from academia to industry, several challenging practical problems have emerged and are been explored extensively in recent years *e.g.*, task incremental learning [36, 81], class incremental learning [21, 48, 73, 82], continual learning [61], few-shot learning [88, 93, 103, 109], open-set recognition [12, 84], and open-world learning [13, 34].

Few-Shot Class Incremental Learning (FSCIL) is one such problem that was recently proposed by Tao et al. [88] where class incremental learning (CIL) is addressed in an even more challenging setup because it addresses the important practical constraint of limited data about new classes. We call their model Fixed FSCIL (FFSCIL) as upfront the system knows exactly N classes will be added and exactly K shots per class are available, where N,K are very small (5/10), see Fig. 1. Due to more constrained assumptions, FFSCIL suffers more over-fitting on the "few" classes. Like

CIL, it also suffers catastrophic forgetting of old classes. As a comparison, in general CIL, even the number of exemplars retained per class for replay are way more (generally more than 20) than the total samples per class (typically 5) in the incremental sessions of FFSCIL. FFSCIL and general CIL also assume a priori fixed number of classes (*i.e*., N-ways) are being introduced in each incremental session. FFSCIL further assumes that the number of shots per class is also fixed (*i.e*., K-shots). In a practical deployment setting, these assumptions rarely, if ever hold true, see Fig.1, requiring either delay or wasting data or both.

To relax these assumptions, we propose to investigate CIL and FSCIL in a variable format, in particular, **Up-to N-Ways, Up-to K-Shots** setting and in this paper explicitly focus on VFSCIL. More specifically, in Up-to N-Ways, Up-to K-Shots class incremental learning **(VFSCIL)**, an agent may have up-to N-ways in each incremental session and up-to K-shots per concept. Hence, the agent need not wait for a fixed number of ways and a fixed number of shots to become available. Inherently the conventional N-ways, K-shots setting, is a special case of herein introduced VFSCIL Up-to N-Ways, Up-to K-Shots.

Open-world learning [13,34] is an even challenging problem where an agent is required to not only learn incrementally but also detect and enroll unknown samples in each session. Open-world learning has been previously investigated in settings where the number of samples per class in each incremental session is quite high (more than 50 training samples per class). In this work, we investigate *few-shot open-world learning (FSOWL)* – an open-world setting similar to FSCIL, where the number of samples per concept is "few". We extend our VFSCIL setting for open-world learning, consequently formulating **Up-to N-Ways, Up-to K-Shots open-world learning (VFSOWL)**.

As we are the first to formulate Variable Few Shot Learning, no baseline pre-exists. We extend the state-of-the-art FFSCIL approach *Continually Evolved Classifier (CEC)* [104] as a strong baseline for VFSCIL, and VFSOWL. While possible it seems unlikely that weaker FFSCIL solutions would outperform this strong baseline on the new problem.

While formally novel, extending CEC is not our primary contribution. Following the trends of incorporating self-supervised learning for downstream tasks [16,34,37,41], we propose a novel approach where we investigate the fusion of self-supervised and supervised features as the representation. Specifically, we leverage the self-supervised models trained on ImageNet-2012 [85] or OpenImages-v6 [59] and a supervised model trained on data belonging to the base session where a relatively larger number of labeled images are available. Both supervised and self-supervised models are kept frozen after the initial training on the base session and disjoint unlabeled dataset, respectively, and serve as first-level feature extractors for the subsequent incremental sessions. A lightweight two-layer classification head is employed as the learnable module that adapts with data for each incremental session. The classification module is trained on the concatenation of independently normalized features emerging from images belonging to the respective incremental sessions. Since images are not directly fed to the classification module, several feature vectors per image are generated with conventional data augmentation techniques. To mitigate the catastrophic forgetting of old classes, we maintain a single mean vector per class as its representation. Our lightweight classification module trained on combined supervised and self-supervised features outperforms the state-of-the-art FSCIL methods on established benchmarks and is concurrently presented [2]. Here we demonstrate that our proposed approach outperforms the extended CEC baseline on the novel VFSCIL and VFSOWL settings and does so by a large, statistically significant margin.

Our Contributions: ***(a)*** Formalizing the novel and more realistic, Variable Few Shot paradigm for class incremental and open-world learning. ***(b)*** First to investigate self-supervised learning for the novel downstream tasks VFSCIL and VFSOWL. ***(c)*** Extending current FFSCIL SOTA CEC as a strong baseline for novel VFSCIL, VFSOWL settings. ***(d)*** Demonstrate our feature fusion approach outperforms the extended baseline on established benchmarks generally employed for FFSCIL and here adapted for new problems, setting new state-of-the-art performance. ***(e)*** Ablations study parameter choices and trade-offs between recognition performance and unknown detection.

2 Related Work

To the best of our knowledge, there is no published prior work on Up-to N-Ways, Up-to K-Shots learning. This section documents some of the work in closely related research problems of incremental learning, few-shot class incremental learning, open-world learning, and self-supervised learning. We discuss the most relevant works and how they are related to and different from our work.

2.1 Incremental Learning

Incremental learning is problem of learning new classes or representations without forgetting past classes and representations [1,3,6,11,23,33,42,44,49,50,62,67,70,71,74,76,80,89,94,96,99,109]. Several incremental learning solutions adapt both feature extractor and classifier jointly [107], others decouple feature extraction from inference subsystem [9], also, many methods add classes incrementally when feature extraction remains constant [10, 63, 87]. Models that use artificial neural networks forget what they have learned previously while learning new concepts [26,32,36,58,83,108] *i.e.*, *catastrophic forgetting*. When new samples of old classes come over time, the distribution of features may change in an unforeseen way, making the performance degrade dramatically on the new data, which is known as *concept* *drift* [15, 43, 101, 102]. Although incremental learning is related, these solutions require many samples to learn new classes, usually more than 20/50 samples per class per incremental session. Their performances drop significantly when the number of samples is small. Additionally, such general class incremental methods have a large memory budget for retaining exemplars to counter the catastrophic forgetting of old classes. Therefore, we do not compare the proposed method with general class incremental learning approaches.

2.2 Few-Shot Class Incremental Learning

Few-Shot Class Incremental Learning (FSCIL) is a special case of incremental learning where the number of samples per class is small [25, 31, 57, 88, 97]. Cheraghian et al. propose to use semantic information during training [30]. A recent work [109] proposed a random episode selection strategy that adapts the feature representation and a self-promoted prototype refinement mechanism, which strengthens the expressive ability of the new classes. In ERL [35], authors focused on the stability-plasticity dilemma and proposed *exemplar relation distillation incremental learning* framework to balance the tasks of old-knowledge preservation and new-knowledge acquisition. Centroid-Based Concept Learning (CBCL) method is proposed in [7]. In [104], authors devised a decoupled learning strategy for representations and classifiers where only the classifiers are updated in each incremental session to avoid knowledge forgetting. To propagate context information between classifiers learned on individual incremental sessions, they employed a graph model and proposed *Continually Evolved Classifier (CEC)*. To the best of our knowledge, CEC is the best performing FSCIL approach to date. We have thoroughly compared our proposed approach against the most recent FSCIL methods, including CEC, on established FSCIL benchmarks in [2]. Herein, we only present a summary table of performance comparison on FSCIL for completeness before focusing on the newly introduced VFSCIL, and VFSOWL problems.

2.3 Open-World Learning

Open-world learning is the problem of detecting and learning new classes incrementally [12, 14, 22, 34, 47, 55, 77, 84]. The biggest difference between open-world learning and incremental learning is that in incremental learning, labels of all new samples are provided, while an open-world learning agent only gets labels for those samples that it has predicted as novel. Some papers [51, 60, 69, 72, 86] used the term "open-world" in their titles, but they never showed how to learn new classes and are inconsistent with the original definitions of [12]. To our best knowledge, there is no published work on few-shot open-world learning (FSOWL). For comparison, we extend the SOTA FSCIL approach (CEC) for FSOWL as a strong baseline and compare our proposed method on benchmarks conventionally adapted for FSCIL.

2.4 Self-Supervised Learning

Self-supervised learning is an active research area where many approaches have emerged in recent years to learn better feature representations without any supervision and labeling. Self-supervised learning has been accomplished by solving a *pretext task* [38–40, 54, 75, 78, 79, 105, 106] using *contrasting loss* [27, 45], by *clustering* [17, 18] the underlying deep features, or by knowledge distillation [20, 64]. Generally, models learned in a self-supervised manner are evaluated on the downstream task of object recognition, using ImageNet-2012 [85], by training a classification head. However, there have been recent studies where self-supervised learning has also been explored for other downstream tasks such as incremental/open-world learning [16, 34], continual learning [37], and novel class discovery and recognition [41]. Inspired by these recent advances, we explore the suitability of self-supervised learning for challenging VFSCIL, and VFSOWL problems. While there has been a recent surge in self-supervised approaches [5, 8, 24, 29, 53, 64–66, 68, 92, 95, 98], in this paper we have used self-supervised features based on Moco-v2 [28], and DeepCluster-v2 [18].

3 Problem Statement

Herein, we first review the formulation of few-shot class incremental learning (FSCIL). Then we extend it for few-shot open-world learning (FSOWL). Subsequently, we generalize them to make both problems **variable** (Up-to N-ways, Up-to K-shots) rather than fixed (N-ways, K-shots), which are the primary problems in this paper.

Few-Shot Class Incremental Learning (FSCIL) Like CIL, in FSCIL, the objective of the underlying model is also to learn new concepts while retaining the knowledge of old ones. But unlike CIL, in FSCIL, very few labeled samples per class become available to the learner in each incremental session. Following [88, 104], let $\{\mathcal{D}^0_{train}, \mathcal{D}^1_{train}, \cdots, \mathcal{D}^n_{train}\}$ be the training sets for n incremental sessions and class labels for i-th session i.e., $\mathcal{D}^i_{train}$ is denoted by $\mathcal{C}^i$. The classes added in different sessions do not have any overlap i.e., $\forall i, j$ where $i \neq j, \mathcal{C}^i \cap \mathcal{C}^j = \emptyset$. After each incremental session i, the model is evaluated on test data belonging to the current session and classes seen in all previous sessions, i.e., $\mathcal{C}^0 \cup \mathcal{C}^1 \cdots \cup \mathcal{C}^i$. In FSCIL, it is conventional [88] to have way more training data in the base session ($\mathcal{D}^0_{train}$) than in the incremental sessions where *N*-way *K*-shot setting is employed, i.e., each incremental session has *N* fixed number of classes and only *K* samples per class are available. For our approach, we assume there exists another unlabeled data set $\mathcal{D}^u_{train}$ that is disjoint with data for any of the sessions in FSCIL and used for self-supervised training.

Few-Shot Open-World Learning (FSOWL) Open-world learning is a general problem where instead of providing the agent with all the training samples from an incremental session, the agent is required to detect training samples from the subsequent session as unknown/novel. The agent then can be provided with the labels for all the training samples irrespective of its detection performance (*i.e.*, incremental learning + unknown detection), or can be provided labels for only samples that have been detected correctly (*i.e.*, true open-world learning). This work explores open-world learning in a few-short setting where the agent is provided fewer samples in each incremental session like FSCIL. The agent is not only evaluated on its incremental recognition performance on all the classes enrolled so far ($\mathcal{C}^0 \cup \mathcal{C}^1 \cdots \cup \mathcal{C}^i$), but also its capability of detecting samples belonging to these classes as *knowns* and samples from classes of subsequent incremental session ($\mathcal{C}^{i+1}$) as *unknowns*. Like FSCIL, the number of ways and shots per concept would also be fixed in FSOWL.

Fixed vs Variable In a fixed version of the above problems (FFSCIL/FFSOWL), the values of N and K are prespecified and fixed for the experiment. While this simplifies experimental protocols and was a good first start in this problem space, these assumptions rarely hold true for practical deployment applications. In the variable (Up-to N-ways, Up-to K-shots) variants of FSCIL and FSOWL, we relax the assumptions of having a fixed number of classes per incremental session and a fixed number of shots per class. In the variable version, there can be up-to N-ways *i.e.*, we can have as few as one class in an incremental session. Similarly, there could be up-to K-shots *i.e.*, and we can have only a single sample for a class. The evaluation for VFSCIL and VFSOWL stays the same as that for FSCIL and FSOWL, respectively. An agent is still evaluated on all the test samples belonging to the classes enrolled so far. In VFSCIL, an agent can use the supplied labels to determine the number of classes for that round; in VFSWOL, agents determine the unknowns, which means they can have different numbers of classes in any round complicating evaluation. A VFSOWL agent may fail to detect and enroll samples from a specific class but is still evaluated on that specific class's test samples. Consistently with FSCIL protocol [88], we assume the base training session has relatively much more data.

4 Method

The overall architecture of our proposed approach ***Fe**w-**S**hot **S**elf-**S**upervised **S**ystem* (FeSSSS) is comprised of: the feature extraction, the feature fusion, the linear classifier, and the *unknown* detection module for open-world learning. Below, we describe each module of our pipeline, code is available on GitHub along with the extended baseline and the experimental setup files for proposed framework.

4.1 Feature Extractors

A typical deep learning model can be thought of as a composition of a feature extractor $\hat{x} = f(x;\theta)$ followed by a classification head $c(\hat{x};\phi)$; where θ, and ϕ are learnable parameters, and x (an image), $\hat{x}$ (feature vector) are the inputs for respective modules. During training, these parameters are learned using data in a supervised or self-supervised manner depending upon the setting. In our hybrid framework FeSSSS, we train one deep model ($f_s(x;\theta_s)$, $c_s(\hat{x};\phi_s)$) using data from base session $\mathcal{D}^0_{train}$ in a supervised manner, and another network ($f_{ss}(x;\theta_{ss})$, $c_{ss}(\hat{x};\phi_{ss})$) is trained on $\mathcal{D}^u_{train}$ in a self-supervised manner. We do not assume any fixed task for the self-supervised model and it can be learned in any conventional manner, i.e., using a *pretext task*, employing *contrastive loss*, through *clustering*, or *knowledge distillation*. Once the two models are trained fully on their respective datasets, their classification heads are discarded and outputs from feature extractors $f_s(x;\theta_s)$, $f_{ss}(x;\theta_{ss})$ are normalized to have unit L^2 norm yielding ($\bar{x}_s, \bar{x}_{ss}$) which are then used as input to the lightweight classification module.

4.2 Feature Fusion and Classification Module

Our model operating in the incremental setting is comprised of a lightweight network $l_c(\theta_c)$ that has two fully connected layers followed by a Softmax. It takes concatenated normalized feature vectors $\bar{x}_t$ = ($\bar{x}_s|\bar{x}_{ss}$) as input and provides the probability vector for n classes that have been enrolled up to the current incremental session. The number of nodes in the intermediate feature fusion layer is set to half of the feature dimension of the concatenated vector $\bar{x}_t$, whereas the number of output nodes is equal to the total number of classes enrolled so far and grow with each incremental session. In each incremental session, up-to N new nodes are added where nodes being added are different in each incremental session, but never more than N and could also be zero. In an open-world learning scenario, only classes detected as *unknowns* are enrolled, and the agent may end up enrolling no new classes in an incremental session as it was unable to detect those samples as *unknowns*. The lightweight module is initially trained with data from base classes, giving the initial training of feature fusion considerably more data.

In each incremental session the lightweight model $l_c(\theta^i_c)$ is initialized with weights from the previous session θ^{i-1}_c and *up-to N* more nodes are added to the output layer. The weights for these new connections are randomly initialized. After training each incremental session, the model is evaluated on test samples belonging to all classes that have been enrolled so far. Importantly, the weights between the input normalized concatenated features are retained so that the system continues to better learn feature fusion over time. The new nodes, while randomly initialized, can exploit those fused features. If the system only used a simple linear classifier (linear layer), even retaining weights for known classes would not allow learning to fuse since the new classes would have no access to that information.

4.3 Unknown Detection

We use a simple Softmax thresholding approach [46] for detecting unknowns, determined based on the fixed true positive rate (TPR) on the base session; the chosen threshold is then used through all the subsequent incremental sessions. More sophisticated novelty/unknown detection approaches can be plugged directly into our pipeline. Still, here we focus on introducing the new problem setup and not after getting the absolute best performance.

5 Experiments and Results

We conducted our experiments on two established benchmark datasets which are commonly employed for FSCIL *i.e*., Caltech-UCSD Birds-200-2011 (CUB200) [91], and miniImageNet [85]. We exclude CIFAR100 [56]), also often used for FSCIL evaluation because, in our opinion, is impractical (32×32 sized images). Thus it does not add much to evaluation. Below we list the details about datasets, experimental settings and subsequently provide results comparing our approach against the extended baseline.

5.1 Data Sets

Caltech-UCSD Birds-200-2011 CUB200 [91] is a fine-grained image classification dataset originally comprised of 5994 training and 5794 test images belonging to 200 classes of birds. Following FSCIL protocol [88] we use the first 100 classes as the base session, and the remaining 100 are distributed in either 15 or 30 incremental sessions depending upon the experimental setting. However, unlike [88] the 100 incremental classes are not distributed equally; rather each incremental session may have up-to N classes. For training, the base session is still comprised of 30 samples per class, whereas for the incremental session, we explore 4 settings, where N and K could be 5 or 10. Specifically we generated (i) Up-to 10-ways, Up-to 10-shots, (ii) Up-to 10-ways, Up-to 5-shots, (iii) Up-to 5-ways, Up-to 5-shots, and (iv) Up-to 5-ways, Up-to 10-shots, and for each experimental setting 5 different experiments are generated. There are 15 incremental sessions for each of the first two settings and 30 for the other two. It should be noted that all test examples belonging to the enrolled classes at any given incremental session are used for evaluation for class incremental learning and the incremental module of the open-world agent.

miniImageNet miniImageNet is a small subset of ImageNet-2012 [85] comprised of 100 classes, each having 600 images; 500 training, and 100 test images. Tao et al. [88] split the 100 classes into 60 base and 40 incremental classes. Following [88] we retain the same 60 classes as the base session while the remaining 40 classes are distributed in either 7 or 14 incremental sessions depending upon the experimental setting. We employ the same 4 experimental settings for miniImageNet as described for CUB200.

5.2 Training Details

For incremental learning experiments, the baseline *CEC* approach is trained following the same training setup as in authors' original code [104] and changes are made to make it work with different numbers of incremental sessions and variables N and K. For open-world learning, a thresholding approach is incorporated to detect and enroll the unknowns while other training parameters such as learning rate schedule, number of epochs, temperature etc. stay the same for each incremental session as reported in [104]. It should be noted that CEC adapts a cosine layer as the classification layer instead of Softmax, so a fixed TPR is employed for a fair comparison between CEC and the proposed to choose their respective thresholds using base session data.

Following existing approaches on FSCIL, for supervised training on the base session, ResNet-18 is used for both CUB200 and miniImageNet datasets. For our feature fusion approach, we use a model from CEC [104] trained on base session data ($\mathcal{D}^0_{train}$) to provide the supervised features. For self-supervised features, we use models trained on ImageNet-2012 [85] and OpenImages-v6 [59] respectively for experiments on CUB200, and miniImageNet. Specifically, we use ResNet-50 models trained by DeepCluster-v2 [18], and Moco-v2 [28] respectively for miniImageNet, and CUB200. In a separate evaluation [2], DeepCluster-v2 and Moco-v2 performed best for FSCIL evaluation compared to other explored self-supervised approaches, including SwAV [19], and SeLa-v2 [4], and hence used here for VFSCIL and VF-SOWL. The mismatch of the disjoint dataset ($\mathcal{D}^u_{train}$) is imposed to enforce no overlap between the datasets used for supervised and self-supervised models.

To enhance the training data for the classification module, we extract features from both supervised and self-supervised models using various augmented versions of each image in each incremental session. For each dataset, we use the same augmentations as originally employed by CEC [104].

The classification module is trained for 500 epochs at a learning rate of 0.1 for the base session. We use the same number of epochs for each incremental session but a lower learning rate of 0.001. A batch size of 256 is used for both base and incremental training. We further employ class balancing to emphasize the importance of old class centroids. We choose the model saved with the last epoch for each incremental session, not the best performing one on the test set. This is because there is no held-out validation set because there are so few samples, and we did not want to tweak the test set that might result in marginal improvement. On the other hand, CEC pipeline is set by respective authors to save the best performing checkpoint on the validation data. For experiments on CUB200, images are resized to 256 maintaining aspect ratio, and then 224×224 random crops or horizontally flipped random crops are used for training. For miniImageNet, we follow CEC [104] and resize images

Table 1. Comparison of FeSSSS with the state-of-the-art FFSCIL on CUB200 dataset. ‡ indicates results reported in [104], * identifies the few-shot approaches adapted by [104] for FSCIL, and † shows the results for approaches taken from their respective papers. Our relative performance gain with respect to each approach in terms of average incremental accuracy is noted in the last column. Using a two-sided t-test with each iteration as the data, our approach is statistically significantly better than the state-of-the-art with $p < .0001$

FFSCIL Performance on CUB200

Method	Acc. in each session (%) ↑											Avg. ↑	our relative improvement
	0	1	2	3	4	5	6	7	8	9	10		
TOPIC‡ [88]	68.68	62.49	54.81	49.99	45.25	41.4	38.35	35.36	32.22	28.31	26.28	43.92	**+18.93**
LEC-Net† [100]	70.86	58.15	54.83	49.34	45.85	40.55	39.70	34.59	36.58	33.56	31.96	45.08	**+17.77**
SS-iCaRL† [31]	69.89	61.24	55.81	50.99	48.18	46.91	43.99	39.78	37.50	34.54	31.33	47.28	**+15.57**
SS-NCM† [31]	69.89	61.91	55.51	51.71	49.68	46.11	42.19	39.03	37.96	34.05	32.65	47.33	**+15.52**
SPPR† [109]	68.68	61.85	57.43	52.68	50.19	46.88	44.65	43.07	40.17	39.63	37.33	49.32	**+13.53**
SS-NCM-CNN† [31]	69.89	64.87	59.82	55.14	52.48	49.60	47.87	45.10	40.47	38.10	35.25	50.78	**+12.07**
Decoupled-DeepEMD‡ [103]*	75.35	70.69	66.68	62.34	59.76	56.54	54.61	52.52	50.73	49.20	47.60	58.73	**+4.12**
Decoupled-Cosine‡ [90]*	75.52	70.95	66.46	61.20	60.86	56.88	55.40	53.49	51.94	50.93	49.31	59.36	**+3.49**
ERL† [35]	73.52	70.12	65.12	62.01	58.56	57.99	56.77	56.52	55.01	53.68	50.01	59.93	**+2.92**
ERL++† [35]	73.52	71.09	66.13	63.25	59.49	59.89	58.64	57.72	56.15	54.75	52.28	61.18	**+1.67**
CEC‡ [104]	75.85	71.94	68.50	63.5	62.43	58.27	57.73	55.81	54.83	53.52	52.28	61.33	**+1.52**
FeSSSS (Ours)	**79.60**	**73.46**	**70.32**	**66.38**	**63.97**	**59.63**	**58.19**	**57.56**	**55.01**	**54.31**	**52.98**	**62.85**	

Table 2. Comparison of FeSSSS with the state-of-the-art on miniImageNet data set. ‡ indicates results copied from CEC [104], * identifies the few-shot approaches adapted by [104] for FSCIL, and † shows the results for approaches taken from their respective papers. Further ◇ identifies that the results have been approximated from graphs since tabular results are unavailable from respective papers.

FFSCIL Performance on miniImageNet

Method	Acc. in each session (%) ↑									Avg. ↑	our relative improvement
	0	1	2	3	4	5	6	7	8		
LEC-Net† [100]	61.31	35.37	36.66	38.59	33.90	35.89	36.12	32.97	30.55	37.92	**+30.31**
TOPIC‡ [88]	61.31	50.09	45.17	41.16	37.48	35.52	32.19	29.46	24.42	39.64	**+28.59**
ERL†◇ [35]	61.67	56.19	54.70	51.19	47.61	45.23	44.0	40.95	39.8	49.03	**+19.20**
ERL++†◇ [35]	61.67	57.61	54.76	51.67	48.57	46.42	44.04	42.85	40.71	49.81	**+18.42**
SS-NCM-CNN†◇ [31]	62.88	60.66	57.55	52.66	50.44	48.44	45.11	41.55	40.88	51.13	**+17.10**
Decoupled-DeepEMD‡ [103]*	69.77	64.59	60.21	56.63	53.16	50.13	47.49	45.42	43.41	54.53	**+13.70**
Decoupled-Cosine‡ [90]*	70.37	65.45	61.41	58.00	54.81	51.89	49.10	47.27	45.63	55.99	**+12.24**
CEC‡ [104]	72.00	66.83	62.97	59.43	56.70	53.73	51.19	49.24	47.63	57.74	**+10.49**
SPPR†◇ [109]	80.0	74.0	68.66	64.33	61.0	57.33	54.66	51.66	49.0	62.29	**+5.94**
FeSSSS (Ours)	**81.5**	**77.04**	**72.92**	**69.56**	**67.27**	**64.34**	**62.07**	**60.55**	**58.87**	**68.23**	

to 92 and then 84×84 random or horizontally flipped random crops are used. During an evaluation, only central crop-based features are concatenated and forward passed through the trained classification module for each image.

Table 3. Incremental learning comparison of FeSSSS against extended CEC on CUB200 dataset across all experimental settings. Mean (μ) and standard deviation (σ) for 5 experiments per each experimental setting is provided.

VFSCIL Performance

experimental setting	FeSSSS		CEC	
	μ	σ	μ	σ
Up-to 10-Ways, Up-to 10-Shots	62.05	1.46	59.12	1.69
Up-to 10-Ways, Up-to 5-Shots	60.67	0.53	57.82	0.94
Up-to 5-Ways, Up-to 5-Shots	60.70	0.81	58.02	0.79
Up-to 5-Ways, Up-to 10-Shots	61.60	1.36	58.67	1.14
Avg ↑	61.26	1.18	58.41	1.21
Gain over CEC	**+2.85**	-	-	-

5.3 Evaluation Metrics

For VFSCIL experiments, we adopt the conventional metric used for FFSCIL *i.e.*, classification accuracy for each session and the *average incremental accuracy*. For each session, all validation samples belonging to the enrolled classes are used for evaluation. To establish a fair comparison between VFSOWL methods, a fixed true positive rate (TPR) is decided which is used to determine the system threshold using the data belonging to the base session. The determined threshold stays fixed throughout the incremental sessions. We have used a 95% TPR for comparing baseline and our approach. An ablation on varying this TPR is also provided. To measure the performance of VFSOWL approaches, we report *average incremental accuracy*, *average incremental unknown detection accuracy (UDA)*, *average incremental known detection accuracy (KDA)*, and *average incremental total detection accuracy (TDA)* for the validation data, and *average incremental unknown detection accuracy (UDA Tr.)* for the training data in the incremental sessions. The recognition accuracy is reported for the current session, whereas the binary detection accuracies are reported for the subsequent incremental sessions.

5.4 Results

We focus on the results and ablations for the CUB200 dataset in the main paper and provide results for miniImageNet in the supplemental. CUB200 is more challenging as there are fewer samples (3K) and more classes (100) in the base session compared to miniImageNet (30K samples, 60 classes).

Comparison Against SOTA FSCIL For completeness, we document our FSCIL results on CUB200, and miniImageNet in Tabs. 1, and 2 respectively, comparing our feature fusion approach against latest state-of-the-art FSCIL methods [31, 35, 88, 100, 104, 109] and outperform each one of them by a significant margin. To emphasize the relative performance gain, we report the average incremental accuracy in second-to-last column and percentage improvement due

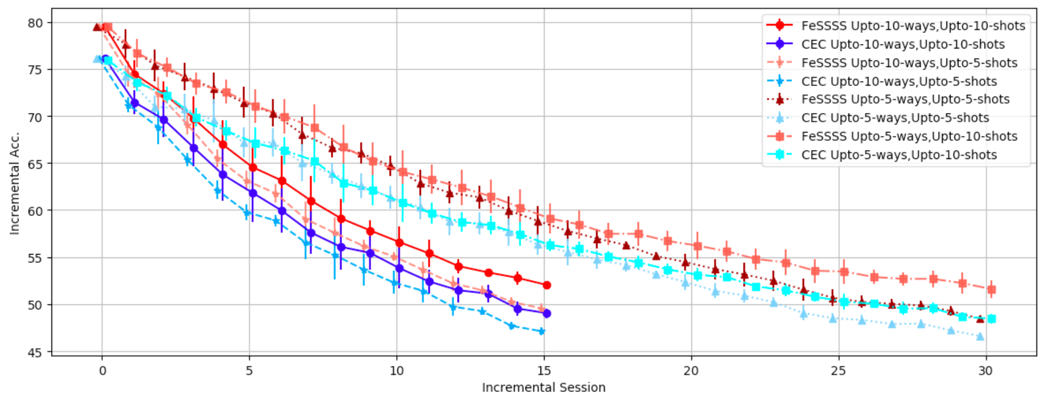

Figure 2. Incremental recognition accuracy for CUB200 for 4 experimental settings. For each setting, average over the five experiments along with the standard deviation in terms of error bars are plotted for both extended baseline (CEC) and our proposed approach FeSSSS. In each experimental setting, FeSSSS outperforms CEC baseline by a large margin.

to our approach in the last column. Since FSCIL is not the main contribution of the paper, we focus on VFSCIL, and VFSOWL evaluation.

VFSCIL Fig. 2 depicts the comparison of FeSSSS against extended CEC where incremental recognition accuracy for each of the 4 experimental settings is plotted as a session-wise average along with the error bars reflecting the standard deviations for each set of experiments. We can note that FeSSSS outperforms extended CEC in each experimental setting and in each incremental session. For both methods, performance drops more rapidly in experimental settings where more ways are introduced in fewer incremental sessions *i.e.*, (i) Up-to 10-Ways, Up-to 10-Shots, and (ii) Up-to 10-Ways, Up-to 5-Shots. Comparatively, for the other two settings, the drop in average incremental accuracy is more gradual. It should also be noted that even for the base session, FeSSSS has higher accuracy than that of extended CEC. Tab. 3 demonstrates the experimental-setting-wise average incremental accuracy and overall average incremental accuracy with respective standard deviations. Overall, FeSSSS outperforms CEC by about 3.0% performance gain averaged across all experiments.

VFSOWL Tab. 4 provides the results for VFSOWL for FeSSSS and extended CEC baseline where thresholds for both approaches are determined at a 95% TPR on the base session. Averages of all the metrics mentioned above are reported. Each metric is first averaged across all incremental sessions and then across all experiments in each experimental setting. We should note that recognition accuracy is reported for each incremental session, whereas detection performances are reported for data from subsequent sessions. Tab. 4 demonstrates that FeSSSS outperforms CEC on 4 out of 5 metrics. Understanding the relation between UDA, KDA, and TDA is very important for open-world learning. Although CEC performs better than FeSSSS in terms of UDA, that is at the expense of declined KDA and consequently dropped TDA. It is interesting to note that unknown detection accuracy for incremental training data of both methods is below 50%, which means they could only enroll about half of the training data in each session.

5.5 Ablations

Varying TPR The value of true positive rate (TPR%) plays a critical role as a trade-off between incremental recognition performance and unknown detection accuracy. Tab. 5 provides the comparison between FeSSSS and CEC on CUB200 dataset when a 90% TPR is used to determine the threshold. Comparing Tabs. 4, and 5 we can see that although the UDA has increased for both training and validation data, KDA and TDA have declined, and resultantly the recognition performance (Acc) has also dropped. We should note that FeSSSS still outperforms CEC on 90% TPR as well.

Importance of Feature Fusion For our proposed approach, we relied on the fusion of supervised and self-supervised features. In an ablation, we studied the role of these features independently. Training two-layer MLP on either of the feature representations alone results in lower performance than using fused features. Tab. 6 documents the results comparing the performance of classification module trained on independent feature representation; either of these resultant systems underperforms the extended CEC baseline.

Updated Supervised Features for Feature Fusion For FeSSSS, the supervised features are extracted from the model trained on data belonging to the base session. It is natural to question if extracting supervised features from an updated model with each incremental session would help improve the performance. Tab. 7 addresses this question where we can see extracting supervised features from updated incremental models results in marginal gain for FeSSSS.

6 Limitations

Unlike the real world, where incremental sessions are not predetermined, the current evaluation assumes a fixed number

Table 4. Comparison of FeSSSS with the extended baseline (CEC) on CUB200 dataset for open-world learning at a 95% TPR. For both methods, we report averages of the *average incremental accuracy* (Acc), average *unknown detection accuracy* (UDA), average *known detection accuracy* (KDA), average *total detection accuracy* (TDA) for 4 experimental settings. We further report the average of average *unknown detection accuracy* (UDA-Tr.) for the training data belonging to incremental sessions. The average across all experiments are documented in second-to-last row, and performance gain over CEC is noted in the last row.

VFSOWL Performance @ 95% TPR

experimental setting	FeSSSS					CEC				
	Acc	UDA	KDA	TDA	UDA-Tr.	Acc	UDA	KDA	TDA	UDA-Tr.
Up-to 10-Ways, Up-to 10-Shots	57.83	37.12	78.20	76.19	48.05	52.45	50.18	68.46	67.37	48.24
Up-to 10-Ways, Up-to 5-Shots	56.48	37.50	77.04	75.10	47.75	52.18	48.12	70.25	68.96	46.96
Up-to 5-Ways, Up-to 5-Shots	56.60	34.55	81.86	80.73	46.11	53.13	46.27	74.67	73.95	29.77
Up-to 5-Ways, Up-to 10-Shots	57.69	33.75	82.23	81.05	42.18	51.85	53.04	70.17	69.61	41.15
Avg ↑	57.15	35.73	79.83	78.27	46.02	52.40	49.40	70.89	69.97	41.53
Gain over CEC	**+4.75**	**-13.67**	**+8.94**	**+8.30**	**+4.49**	-	-	-	-	-

Table 5. Comparison of FeSSSS with the extended baseline (CEC) on CUB200 dataset for open-world learning at a 90% TPR. For both methods, we report averages of the *average incremental accuracy* (Acc), average *unknown detection accuracy* (UDA), average *knowns detection accuracy* (KDA), average *total detection accuracy* (TDA) for 4 experimental settings. We further report the average of average *unknown detection accuracy* (UDA-Tr.) for the training data belonging to incremental session.

VFSOWL Performance @ 90% TPR

experimental setting	FeSSSS					CEC				
	Acc	UDA	KDA	TDA	UDA-Tr.	Acc	UDA	KDA	TDA	UDA-Tr.
Up-to 10-Ways, Up-to 10-Shots	55.35	51.01	69.97	68.96	62.09	49.20	64.65	59.98	59.97	58.90
Up-to 10-Ways, Up-to 5-Shots	54.16	50.63	69.29	68.28	62.90	49.10	62.71	61.37	61.17	55.38
Up-to 5-Ways, Up-to 5-Shots	54.98	51.65	73.02	72.49	61.76	50.30	63.53	64.69	64.61	40.37
Up-to 5-Ways, Up-to 10-Shots	55.18	48.44	73.82	73.16	58.51	48.47	68.73	60.57	60.61	52.30
Avg ↑	54.92	50.43	71.53	70.72	61.31	49.27	64.91	61.65	61.59	51.74
Gain over CEC	**+5.65**	**-14.48**	**+9.88**	**+9.13**	**+9.57**	-	-	-	-	-

Table 6. Ablation demonstrating the importance of fused features. Classification module trained on either of the independent representations performs poorly than CEC. Mean of average incremental accuracy for 5 experiments per experimental setting is provided.

VFSCIL Performance

experimental setting	FeSSSS	Self-Supervised	Supervised
	μ	μ	μ
Up-to 10-Ways, Up-to 10-Shots	62.05	54.49	57.40
Up-to 10-Ways, Up-to 5-Shots	60.67	52.78	56.41
Up-to 5-Ways, Up-to 5-Shots	60.70	53.21	56.63
Up-to 5-Ways, Up-to 10-Shots	61.60	53.70	56.99
Avg ↑	61.26	53.54	56.86
Gain over CEC	**+2.85**	**-4.87**	**-1.55**

Table 7. Ablation exploring performance gain if supervised features are extracted from updated model adapted with each incremental session instead of base-session model. The performance gain is understandably minimal as base classes still dominate.

VFSCIL Performance

experimental setting	FeSSSS – base	FeSSSS – incremental
	μ	μ
Up-to 10-Ways, Up-to 10-Shots	62.05	62.36
Up-to 10-Ways, Up-to 5-Shots	60.67	60.81

that is essential for comparing approaches. The current up-to N-ways, up-to K-shots evaluation applies to a few-shot setting; its applicability to general class incremental and open-world learning is yet to be explored. FeSSSS and extended CEC baseline assume the presence of an oracle that provides ground truth labels for correctly detected novel samples; this assumption can be relaxed as a future work in an unsupervised setting where clustering can be used to provide pseudo labels and becomes closer to [52]. In the current evaluation of VFSCIL/VFSOWL, the lower bound of N and K is by default fixed to 1; in an extension, a more constrained setting of M-to-N-ways, J-to-K-shots can further be explored where both upper and lower ends are bounded. The current evaluation is conducted in a curricular fashion where classes in incremental sessions do not overlap.

7 Conclusions

In this work, we generalized beyond existing fixed few-shot class incremental learning, proposing the more realistic Variable Few Shot Class Incremental Learning (VFSCIL) where a learning agent can expect up-to N-ways, up-to K-shots per incremental session. We also introduced the variable few-shot variant of open-world learning *i.e.*, VFSOWL. State-of-the-art fixed FSCIL is extended as a baseline to operate in both VFSCIL and VFSOWL. Our proposed approach is demonstrated to outperform existing SOTA FFSCIL methods and the extended baseline on both introduced problems using extensive experiments. Ablations were conducted to shed light on the importance of feature fusion and the trade-off between recognition performance and unknown detection accuracy. Next, we plan to investigate variable class incremental and open-world learning where data in incremental sessions is not scarce.

Acknowledgement

This research is supported by the Defense Advanced Research Projects Agency (DARPA), specifically SAIL-ON program under contract number HR001120C0055.

References

[1] Mohamed Abdelsalam, Mojtaba Faramarzi, Shagun Sodhani, and Sarath Chandar. Iirc: Incremental implicitly-refined classification. In *CVPR*, pages 11038–11047, 2021. 2

[2] Touqeer Ahmad, Akshay Raj Dhamija, Steve Cruz, Ryan Rabinowitz, Chunchun Li, Mohsen Jafarzadeh, and Terrance E. Boult. Few-shot class incremental learning leveraging self-supervised features. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, 2022. 2, 3, 5

[3] Hongjoon Ahn, Jihwan Kwak, Subin Lim, Hyeonsu Bang, Hyojun Kim, and Taesup Moon. Ss-il: Separated softmax for incremental learning. In *ICCV*, pages 844–853, 2021. 2

[4] Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. In *Int. Conf. Learn. Represent.*, 2020. 5

[5] Sara Atito, Muhammad Awais, and Josef Kittler. Sit: Self-supervised vision transformer. *arXiv preprint arXiv:2104.03602*, 2021. 3

[6] Ali Ayub and Alan Wagner. Eec: Learning to encode and regenerate images for continual learning. In *ICLR*, 2020. 2

[7] Ali Ayub and Alan R Wagner. Cognitively-inspired model for incremental learning using a few examples. In *CVPRW*, pages 222–223, 2020. 3

[8] Hangbo Bao, Li Dong, and Furu Wei. Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*, 2021. 3

[9] Eden Belouadah and Adrian Popescu. Deesil: Deep-shallow incremental learning. In *ECCV*, pages 0–0, 2018. 2

[10] Eden Belouadah and Adrian Popescu. Scail: Classifier weights scaling for class incremental learning. In *WACV*, pages 1266–1275, 2020. 2

[11] Eden Belouadah, Adrian Popescu, and Ioannis Kanellos. A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Networks*, 2020. 2

[12] Abhijit Bendale and Terrance Boult. Towards open world recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2015. 1, 3

[13] Abhijit Bendale and Terrance E. Boult. Towards Open Set Deep Networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016. 1, 2

[14] Terrance E Boult, Steve Cruz, Akshay Raj Dhamija, Manuel Gunther, James Henrydoss, and Walter J Scheirer. Learning and the unknown: Surveying steps toward open world recognition. In *AAAI*, pages 9801–9807, 2019. 3

[15] Zhipeng Cai, Ozan Sener, and Vladlen Koltun. Online continual learning with natural distribution shifts: An empirical study with visual data. In *ICCV*, pages 8281–8290, 2021. 3

[16] Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-World Semi-Supervised Learning. In *arXiv:2102.03526v2*, 2021. 2, 3

[17] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep Clustering for Unsupervised Learning of Visual Features. In *ECCV*, 2018. 3

[18] Mathilde Caron, Piotr Bojanowski, Julien Mairal, and Armand Joulin. Unsupervised Pre-Training of Image Features on Non-Curated Data. In *ICCV*, 2019. 3, 5

[19] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Adv. Neural Inform. Process. Syst.*, 2020. 5

[20] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. *arXiv preprint arXiv:2104.14294*, 2021. 3

[21] Francisco M. Castro, Manuel J. Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-End Incremental Learning. In *ECCV*, 2018. 1

[22] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Deep metric learning for open world semantic segmentation. In *ICCV*, pages 15333–15342, 2021. 3

[23] Hyuntak Cha, Jaeho Lee, and Jinwoo Shin. Co2l: Contrastive continual learning. In *ICCV*, pages 9516–9525, 2021. 2

[24] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *CVPR*, pages 12299–12310, 2021. 3

[25] Kuilin Chen and Chi-Guhn Lee. Incremental few-shot learning via vector quantization in deep embedded space. In *ICLR*, 2020. 3

[26] Pei-Hung Chen, Wei Wei, Cho-Jui Hsieh, and Bo Dai. Overcoming catastrophic forgetting by bayesian generative regularization. In *ICML*, pages 1760–1770. PMLR, 2021. 2

[27] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*, 2020. 3

[28] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *ArXiv*, 2010.15277, 2020. 3, 5

[29] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. *arXiv preprint arXiv:2104.02057*, 2021. 3

[30] Ali Cheraghian, Shafin Rahman, Pengfei Fang, Soumava Kumar Roy, Lars Petersson, and Mehrtash Harandi. Semantic-aware knowledge distillation for few-shot class-incremental learning. In *CVPR*, pages 2534–2543, 2021. 3

[31] Yawen Cui, Wuti Xiong, Mohammad Tavakolian, and Li Liu. Semi-Supervised Few-Shot Class-Incremental Learning. In *International Conference on Image Processing*, 2021. 3, 6

[32] Matthias Delange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Greg Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *PAMI*, 2021. 2

[33] Mohammad Mahdi Derakhshani, Xiantong Zhen, Ling Shao, and Cees Snoek. Kernel continual learning. In *ICML*, pages 2621–2631. PMLR, 2021. 2

[34] Akshay Raj Dhamija, Touqeer Ahmad, Jonathan Schwan, Mohsen Jafarzadeh, Chunchun Li, and Terrance E. Boult. Self-Supervised Features Improve Open-World Learning. In *arXiv:2102.07848v2*, 2021. 1, 2, 3

[35] Songlin Dong, Xiaopeng Hong, Xiaoyu Tao, Xinyuan Chang, Xing Wei, and Yihong Gong. Few-Shot Class-Incremental Learning via Relation Knowledge Distillation. In *AAAI*, 2021. 3, 6

[36] Arthur Douillard, Matthieu Cord, Charles Ollion, Thomas Robert, and Eduardo Valle. Podnet: Pooled outputs distillation for small-tasks incremental learning. In *Eur. Conf. Comput. Vis.*, 2020. 1, 2

[37] Jhair Gallardo, Tyler L. Hayes, and Christopher Kanan. Self-Supervised Training Enhances Online Continual Learning. In *arXiv:2103.14010v2*, 2021. 2, 3

[38] Wouter Van Gansbeke, Simon Vandenhende, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. SCAN: Learning to Classify Images without Labels. In *ECCV*, 2020. 3

[39] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised Representation Learning by Predicting Image Rotations. In *ICLR*, 2018. 3

[40] Priya Goyal, Dhruv Mahajan, Abhinav Gupta, and Ishan Misra. Scaling and Benchmarking Self-Supervised Visual Representation Learning. In *ICCV*, 2019. 3

[41] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically Discovering and Learning New Visual Categories with Ranking Statistics. In *ICLR*, 2020. 2, 3

[42] Chen He, Ruiping Wang, and Xilin Chen. A tale of two cils: The connections between class incremental learning and class imbalanced learning, and beyond. In *CVPR*, pages 3559–3569, 2021. 2

[43] Jiangpeng He, Runyu Mao, Zeman Shao, and Fengqing Zhu. Incremental learning in online scenario. In *CVPR*, pages 13926–13935, 2020. 3

[44] Jiangpeng He and Fengqing Zhu. Online continual learning for visual food classification. In *ICCV*, pages 2337–2346, 2021. 2

[45] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum Contrast for Unsupervised Visual Representation Learning. In *CVPR*, 2020. 3

[46] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *Int. Conf. Learn. Represent.*, 2017. 5

[47] James Henrydoss, Steve Cruz, Ethan M Rudd, Manuel Gunther, and Terrance E Boult. Incremental open set intrusion recognition using extreme value machine. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1089–1093. IEEE, 2017. 3

[48] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a Unified Classifier Incrementally via Rebalancing. In *CVPR*, 2019. 1

[49] Xinting Hu, Kaihua Tang, Chunyan Miao, Xian-Sheng Hua, and Hanwang Zhang. Distilling causal effect of data in class-incremental learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3957–3966, 2021. 2

[50] Ahmet Iscen, Jeffrey Zhang, Svetlana Lazebnik, and Cordelia Schmid. Memory-efficient incremental learning through feature adaptation. In *ECCV*, pages 699–715. Springer, 2020. 2

[51] Mohsen Jafarzadeh, Touqeer Ahmad, Akshay Raj Dhamija, Chunchun Li, Steve Cruz, and Terrance E Boult. Automatic open-world reliability assessment. In *WACV*, pages 1984–1993, 2021. 3

[52] Mohsen Jafarzadeh, Akshay Raj Dhamija, Steve Cruz, Chunchun Li, Touqeer Ahmad, and Terrance E Boult. Open-world learning without labels. *arXiv preprint arXiv:2011.12906v2*, 2020. 8

[53] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A survey on contrastive self-supervised learning. *Technologies*, 9(1):2, 2021. 3

[54] Simon Jenni, Hailin Jin, and Paolo Favaro. Steering Self-Supervised Feature Learning Beyond Local Pixel Statistics. In *CVPR*, 2020. 3

[55] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *CVPR*, pages 5830–5840, 2021. 3

[56] Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto, 2009. 5

[57] Anna Kukleva, Hilde Kuehne, and Bernt Schiele. Generalized and incremental few-shot learning by explicit learning and calibration without forgetting. In *ICCV*, pages 9020–9029, 2021. 3

[58] Vinod K Kurmi, Badri N Patro, Venkatesh K Subramanian, and Vinay P Namboodiri. Do not forget to attend to uncertainty while mitigating catastrophic forgetting. In *WACV*, pages 736–745, 2021. 2

[59] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, Tom Duerig, and Vittorio Ferrari. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *Int. J. Comput. Vis.*, 128:1956–1981, 2020. 2, 5

[60] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Eric Horvitz. Identifying unknown unknowns in the open world: Representations and policies for guided exploration. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. 3

[61] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021. 1

[62] Justin Leo and Jugal Kalita. Incremental deep neural network learning using classification confidence thresholding. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 2

[63] Timothée Lesort, Vincenzo Lomonaco, Andrei Stoian, Davide Maltoni, David Filliat, and Natalia Díaz-Rodríguez. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information Fusion*, 58:52–68, 2020. 2

[64] Chunyuan Li, Jianwei Yang, Pengchuan Zhang, Mei Gao, Bin Xiao, Xiyang Dai, Lu Yuan, and Jianfeng Gao. Efficient self-supervised vision transformers for representation learning. *arXiv preprint arXiv:2106.09785*, 2021. 3

[65] Zhaowen Li, Zhiyang Chen, Fan Yang, Wei Li, Yousong Zhu, Chaoyang Zhao, Rui Deng, Liwei Wu, Rui Zhao, Ming

Tang, et al. Mst: Masked self-supervised transformer for visual representation. *arXiv preprint arXiv:2106.05656*, 2021. 3
[66] Yahui Liu, Enver Sangineto, Wei Bi, Nicu Sebe, Bruno Lepri, and Marco De Nadai. Efficient training of visual transformers with small-size datasets. *arXiv preprint arXiv:2106.03746*, 2021. 3
[67] Yaoyao Liu, Bernt Schiele, and Qianru Sun. Adaptive aggregation networks for class-incremental learning. In *CVPR*, pages 2544–2553, 2021. 2
[68] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030*, 2021. 3
[69] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open-world. In *CVPR*, pages 2537–2546, 2019. 3
[70] Vincenzo Lomonaco, Lorenzo Pellegrini, Andrea Cossu, Antonio Carta, Gabriele Graffieti, Tyler L Hayes, Matthias De Lange, Marc Masana, Jary Pomponi, Gido M van de Ven, et al. Avalanche: an end-to-end library for continual learning. In *CVPR*, pages 3600–3610, 2021. 2
[71] Zheda Mai, Ruiwen Li, Hyunwoo Kim, and Scott Sanner. Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning. In *CVPR*, pages 3589–3599, 2021. 2
[72] M Mancini, MF Naeem, Y Xian, and Zeynep Akata. Open world compositional zero-shot learning. In *CVPR*. IEEE, 2021. 3
[73] Fei Mi, Lingjing Kong, Tao Lin, Kaicheng Yu, and Boi Faltings. Generalized class incremental learning. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 240–241, 2020. 1
[74] Umberto Michieli and Pietro Zanuttigh. Knowledge distillation for incremental learning in semantic segmentation. *Computer Vision and Image Understanding*, 205:103167, 2021. 2
[75] Ishan Misra and Laurens van der Maaten. Self-Supervised Learning of Pretext-Invariant Representations. In *CVPR*, 2020. 3
[76] Sudhanshu Mittal, Silvio Galesso, and Thomas Brox. Essentials for class incremental learning. In *CVPR*, pages 3513–3522, 2021. 2
[77] Martin Mundt, Yong Won Hong, Iuliia Pliushch, and Visvanathan Ramesh. A wholistic view of continual learning with deep neural networks: Forgotten lessons and the bridge to active and open world learning. *arXiv preprint arXiv:2009.01797*, 2020. 3
[78] Mehdi Noroozi and Paolo Favaro. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. In *ECCV*, 2016. 3
[79] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. Representation Learning by Learning to Count. In *ICCV*, 2017. 3
[80] Jaeyoo Park, Minsoo Kang, and Bohyung Han. Class-incremental learning for action recognition in videos. In *ICCV*, pages 13698–13707, 2021. 2
[81] Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Mubarak Shah. itaml: An incremental task-agnostic meta-learning approach. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 1
[82] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. iCaRL: Incremental Classifier and Representation Learning. In *CVPR*, 2017. 1
[83] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. In *NIPS*, pages 350–360, 2019. 2
[84] Ethan M Rudd, Lalit P Jain, Walter J Scheirer, and Terrance E Boult. The extreme value machine. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(3):762–768, 2017. 1, 3
[85] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, 2015. 2, 3, 5
[86] Vikash Sehwag, Arjun Nitin Bhagoji, Liwei Song, Chawin Sitawarin, Daniel Cullina, Mung Chiang, and Prateek Mittal. Analyzing the robustness of open-world machine learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, pages 105–116, 2019. 3
[87] Wenju Sun, Jing Zhang, Danyu Wang, Yangli-ao Geng, and Qingyong Li. Ilcoc: An incremental learning framework based on contrastive one-class classifiers. In *CVPR*, pages 3580–3588, 2021. 2
[88] Xiaoyu Tao, Xiaopeng Hong, Xinyuan Chang, Songlin Dong, Xing Wei, and Yihong Gong. Few-Shot Class-Incremental Learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 1, 3, 4, 5, 6
[89] Gido M van de Ven, Zhe Li, and Andreas S Tolias. Class-incremental learning with generative classifiers. In *CVPR*, pages 3611–3620, 2021. 2
[90] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching Networks for One Shot Learning. In *NIPS*, 2016. 6
[91] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 5
[92] Feng Wang, Tao Kong, Rufeng Zhang, Huaping Liu, and Hang Li. Self-supervised learning by estimating twin class distributions. *arXiv preprint arXiv:2110.07402*, 2021. 3
[93] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020. 1
[94] Guile Wu, Shaogang Gong, and Pan Li. Striking a balance between stability and plasticity for class-incremental learning. In *ICCV*, pages 1124–1133, 2021. 2
[95] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers. *arXiv preprint arXiv:2103.15808*, 2021. 3
[96] Ziyang Wu, Christina Baek, Chong You, and Yi Ma. Incremental learning via rate reduction. In *CVPR*, pages 1125–1133, 2021. 2

[97] Shudong Xie, Yiqun Li, Dongyun Lin, Tin Lay Nwe, and Sheng Dong. Meta module generation for fast few-shot incremental learning. In *ICCV*, pages 0–0, 2019. 3

[98] Zhenda Xie, Yutong Lin, Zhuliang Yao, Zheng Zhang, Qi Dai, Yue Cao, and Han Hu. Self-supervised learning with swin transformers. *arXiv preprint arXiv:2105.04553*, 2021. 3

[99] Shipeng Yan, Jiangwei Xie, and Xuming He. Der: Dynamically expandable representation for class incremental learning. In *CVPR*, pages 3014–3023, 2021. 2

[100] Boyu Yang, Mingbao Lin, Binghao Liu, Mengying Fu, Chang Liu, Rongrong Ji, and Qixiang Ye. Learnable Expansion-and-Compression Network for Few-shot Class-Incremental Learning. In *arXiv:2104.02281*, 2021. 6

[101] Zhe Yang, Sameer Al-Dahidi, Piero Baraldi, Enrico Zio, and Lorenzo Montelatici. A novel concept drift detection method for incremental learning in nonstationary environments. *IEEE transactions on neural networks and learning systems*, 31(1):309–320, 2019. 3

[102] Lu Yu, Bartlomiej Twardowski, Xialei Liu, Luis Herranz, Kai Wang, Yongmei Cheng, Shangling Jui, and Joost van de Weijer. Semantic drift compensation for class-incremental learning. In *CVPR*, pages 6982–6991, 2020. 3

[103] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. DeepEMD: Differentiable Earth Mover's Distance for Few-Shot Learning. In *CVPR*, 2020. 1, 6

[104] Chi Zhang, Nan Song, Guosheng Lin, Yun Zheng, Pan Pan, and Yinghui Xu. Few-Shot Incremental Learning with Continually Evolved Classifiers. In *CVPR*, 2021. 2, 3, 5, 6

[105] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful Image Colorization. In *ECCV*, 2016. 3

[106] Richard Zhang, Phillip Isola, and Alexei A. Efros. Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction. In *CVPR*, 2017. 3

[107] Bowen Zhao, Xi Xiao, Guojun Gan, Bin Zhang, and Shu-Tao Xia. Maintaining discrimination and fairness in class incremental learning. In *CVPR*, pages 13208–13217, 2020. 2

[108] Fan Zhou and Chengtai Cao. Overcoming catastrophic forgetting in graph neural networks with experience replay. In *AAAI*, pages 4714–4722, 2021. 2

[109] Kai Zhu, Yang Cao, Wei Zhai, Jie Cheng, and Zheng-Jun Zha. Self-Promoted Prototype Refinement for Few-Shot Class-Incremental Learning. In *CVPR*, 2021. 1, 2, 3, 6