Pseudo Informative Episode Construction for Few-Shot Class-Incremental Learning

Chaofan Chen[1,2], Xiaoshan Yang[1,2,3], Changsheng Xu[1,2,3*]
[1]State Key Laboratory of Multimodal Artificial Intelligence Systems (MAIS),
Institute of Automation, Chinese Academy of Sciences (CASIA)
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences (UCAS)
[3]Peng Cheng Laboratory, China
chencfbupt@gmail.com, xiaoshan.yang@nlpr.ia.ac.cn, csxu@nlpr.ia.ac.cn

Abstract

Few-Shot Class-Incremental Learning (FSCIL) studies how to empower the machine learning system to continually learn novel classes with only a few annotated examples. To tackle the FSCIL task, recent state-of-the-art methods propose to employ the meta-learning mechanism, which constructs the pseudo incremental episodes/tasks in the training phase. However, these methods only select part of the base classes to construct the pseudo novel classes in the feature space of the base classes, which cannot mimic the real novel classes of the testing scenario. To deal with this problem, we propose a new Pseudo Informative Episode Construction (PIEC) framework. Specifically, we first perform distribution-level mixing to generate a set of pseudo novel classes in the feature space of the novel class. Then, we propose two diversity criteria to select the informative pseudo novel classes that have large discrepancies with each other and high information gain over the base classes to construct the pseudo incremental session. In this way, we can allow the model to learn rich new concepts beyond the base classes as in the real incremental session during the episodic training procedure, thus improving its generalization ability. Extensive experiments on three popular classification benchmarks (i.e., CUB200, miniImageNet, and CIFAR100) show that the proposed framework can outperform other state-of-the-art methods.

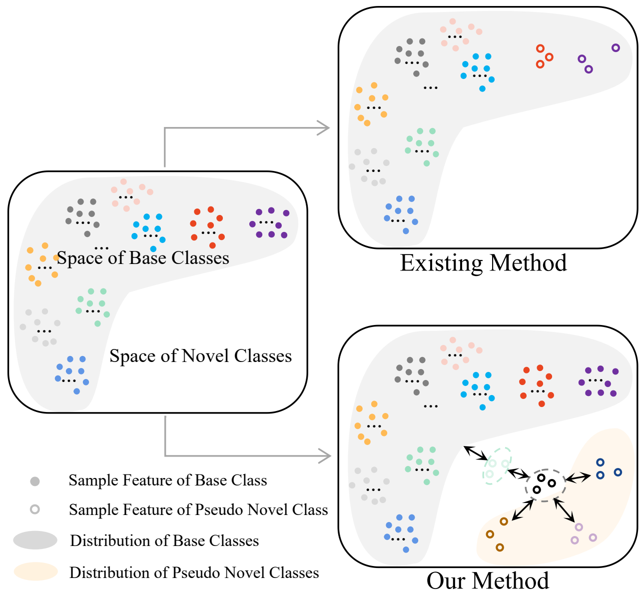


Figure 1: Comparison between existing meta learning-based method and this work on pseudo incremental episode construction. Note that different colors represent different classes. Existing methods simply construct the pseudo novel classes in the feature space of the base classes. In this work, we generate informative pseudo novel classes that are highly different from each other and have high information gain relative to the base classes.

Introduction

Deep learning has achieved significant success in image recognition tasks, largely due to the availability of large-scale datasets. However, collecting such large datasets requires intensive human labor and is not feasible in some real-world scenarios where the examples are scarce. In contrast, humans can learn new concepts from very few examples, which has inspired the development of few-shot learning (FSL), where a model tries to recognize the novel classes correctly given only a few labeled examples. While few-shot learning is promising, its practical usability is a concern because it only focuses on learning novel classes, ignoring estimating the recognition ability of the updated model in the base classes. To address this limitation, a straightforward solution is to employ class-incremental learning (Castro et al. 2018a; Chaudhry et al. 2019; Kirkpatrick et al. 2017) (CIL), which aims to continually update the model after new classes or environments are encountered while maintaining the recognition performance of the base classes. Few-shot class-incremental learning (FSCIL), which incorporates incremental learning and few-shot learning, requires the model to learn from large datasets for base classes while still being flexible enough to learn from a few examples for novel classes.

A naive approach for FSCIL is to finetune the old model (learned from the base classes) on the training set of the novel classes. However, this approach often causes the problem of catastrophic forgetting, which refers to the signifi-

*Corresponding author.

cant decline in the model's performance on old classes. To alleviate the forgetting problem, existing FSCIL methods usually employ the regularization (e.g., TOPIC (Tao et al. 2020) and IDLVQC (Chen and Lee 2021)) and replay (e.g., ERDFR (Liu et al. 2022), ERG (Dong et al. 2021) and C-FSCIL (Hersche et al. 2022)) strategies commonly used in CIL tasks. Despite their success, the incremental learning procedure of these methods is not engaged in the training phase of the base classes. Specifically, the training objective on the dataset of the base classes is to learn a general classification model that can recognize the unseen samples of the base classes correctly, which is not consistent with the testing scenario where the model is expected to learn new concepts over the base knowledge continuously. In conclusion, the learning strategy of these existing methods cannot help guarantee the generalization ability of the recognition model, thereby limiting its performance in the testing phase.

In order to address the above problem, recent state-of-the-art methods (e.g., CEC (Zhang et al. 2021), Meta-FSCIL (Chi et al. 2022) and LIMIT (Zhou et al. 2023)) propose to employ the meta learning strategy to construct the pseudo incremental tasks/episodes to simulate the real-world testing scenario. As shown in the top branch of Figure 1, for a pseudo incremental episode, existing meta learning-based methods select a set of base classes and treat their rotated or original samples as pseudo novel classes. However, these pseudo novel classes still lie in the feature space of the base classes, which **cannot mimic the distribution of the real novel classes.** Since **the backbone pretrained on the base task has learned the knowledge of the base classes**, these methods can hardly constrain the model to learn new concepts in the pseudo incremental episode. To deal with this limitation, we need to generate/synthesize pseudo novel classes that contain sufficient new knowledge to allow the model to learn incrementally in the feature space of the novel classes (i.e., the feature space outside the base classes). To generate informative pseudo novel classes for effective pseudo incremental task construction, we believe that two aspects should be considered as shown in the bottom branch of Figure 1: **(1) The pseudo novel classes should have large discrepancies with each other. (2) The pseudo novel classes should provide high information gain beyond the base classes.** In this way, we can incentivize the model to learn rich new concepts in each pseudo incremental session as in the real incremental session, thereby ensuring the generalization ability of the model after episodic training.

Based on the above insights, we propose a new meta learning paradigm, i.e., **pseudo informative episode construction (PIEC)**, to solve the FSCIL task. In the pretraining stage, we rely on the supervised learning of the whole base classes to optimize the general classification model (e.g., a CNN backbone followed by a fully connected layer), which can help learn a robust backbone. In the pseudo incremental learning stage, we design a three-step paradigm to construct the informative pseudo incremental tasks and conduct the episodic training. Specifically, for a pseudo incremental session in an episode, **(1)** we first perform the distribution-level pairwise mixing on the Gaussian distributions of the base classes to generate a set of pseudo novel classes. This is inspired by the fact that humans have the ability to integrate old concepts to generate new ones. However, not all the pseudo novel classes are suitable for constructing the pseudo incremental session because some pseudo novel classes cannot provide sufficient new information to facilitate the pseudo learning procedure. Therefore, **(2)** we propose two diversity criteria strategies (i.e., *novel-novel diversity criteria* and *base-novel diversity criteria*) to select the informative pseudo novel classes, which can help ensure that the recognition model is constrained to continually learn new knowledge. The *novel-novel diversity criteria* aims to reserve the pseudo novel classes with large diversity, i.e., a pseudo novel class should be significantly different from the other novel classes. In this work, we measure the class similarities and discard the classes that have large similarity scores with the others. The *base-novel diversity criteria* aims to select pseudo novel classes that have high information gain over the base classes. To reflect the information gain, we measure the K-L divergence between the set of the base classes and the novel class. Based on the Gaussian distributions of the selected pseudo novel classes, we randomly sample a few examples to simulate the real-world few-shot scenario. In this way, a pseudo informative session is constructed for the pseudo incremental learning. Finally, **(3)** we feed the query samples, the base class prototypes, and the novel class prototypes into a metric-based inference module (Ye et al. 2020) to conduct few-shot classification.

To sum up, the main contributions of this paper are threefold:

- We propose a new meta learning paradigm for the FSCIL task, which allows the recognition model to continuously learn new concepts beyond the base classes for each pseudo incremental session, thereby improving its generalization ability. To the best of our knowledge, this research line has not been studied in the existing literatures.
- We propose two diversity criteria strategies to help select the informative pseudo novel classes to construct the pseudo incremental session, which can provide rich new knowledge to facilitate the pseudo incremental learning procedure.
- We conduct extensive experiments on three classification benchmarks for the FSCIL task, and the results demonstrate that the proposed PIEC obtains the state-of-the-art performance.

Related Work

Few-shot Learning

Few-shot Learning (FSL) aims to recognize the unseen samples (i.e., query samples) of the novel classes successfully when only a few labeled samples (i.e., support samples) are available. Existing FSL methods usually employ the meta-learning strategy to guarantee the generalization ability of the model to the testing environments. We can roughly divide these meta-learning based methods into two groups, i.e., metric-based method and gradient-based method. (1) The metric-based methods first embed the support and query samples into the same feature space, and then employ a metric to compare the sample to sample similarities (e.g., MatchingNet (Vinyals et al. 2016), RelationNet (Sung et al. 2018)) or sample to prototype similarities (e.g., ProtoNet (Snell, Swersky, and Zemel 2017)) for inference. As an important branch of the metric-based method, the GNN-based methods (Satorras and Estrach 2018; Liu et al. 2019; Chen et al. 2021) treat the samples as nodes and leverage the graph neural networks to evolve the relation information among the samples for representation enhancement and metric-based inference. (2) The gradient-based methods (Ravi and Larochelle 2017; Finn, Abbeel, and Levine 2017; Rusu et al. 2019) target at learning the model parameters that can be adapted to the new environment quickly within just a few optimization steps.

Few-Shot Class-Incremental Learning

The goal of Few-Shot Class-Incremental Learning (FSCIL) is to incrementally learn new classes with only a few annotated samples, while preserving its ability to recognize previously learned classes. In the early stages of the research, the regularization (Tao et al. 2020; Chen and Lee 2021; Akyürek et al. 2022) and replay (Dong et al. 2021; Hersche et al. 2022; Liu et al. 2022) strategies are employed to conduct the FSCIL task. Recently, an increasing popular research line is to employ the meta learning mechanism to conduct the FSCIL task. For example, CEC (Zhang et al. 2021) formulates a pseudo incremental learning paradigm and employs a graph model to combine classifiers learned on individual sessions. In this way, it can bring the knowledge learned in the early session to the current session to avoid the forgetting problem. SPPR (Zhu et al. 2021) randomly selects part of the base classes to construct the pseudo novel classes. In each episode, it considers the dependencies among different classes to help strengthen the representation ability of the novel classes. LIMIT (Zhou et al. 2023) splits the base classes to construct multi-phase fake-incremental tasks and utilizes the transformer architecture to constrain the old class classifiers and new class prototypes to the same scale. MetaFSCIL (Chi et al. 2022) employs a similar fake-incremental task generation procedure with LIMIT. It proposes the bi-directional guided modulation and bi-level meta-learning-based optimization to help realize forgetting alleviation and adaptation.

Methodology

Problem Statement

We follow the settings widely used in (Tao et al. 2020; Zhang et al. 2021; Chi et al. 2022) to establish the few-shot class-incremental learning (FSCIL) task. There is a stream of training datasets $\mathcal{D} = \{\mathcal{D}^0, \mathcal{D}^1, ..., \mathcal{D}^T\}$ with the corresponding label set of $\mathcal{C} = \{\mathcal{C}^0, \mathcal{C}^1, ..., \mathcal{C}^T\}$, where T denotes the number of the incremental sessions. Note that the label set of different training datasets are disjoint, i.e., $\forall i \neq j$, $\mathcal{C}^i \cap \mathcal{C}^j = \varnothing$. $\mathcal{D}^0$ denotes the training set of the base session, which contains sufficient labeled samples for each class. We leverage $L = |\mathcal{C}^0|$ to represent the number of the base classes in the base session. $\mathcal{D}^t = \{x_i, y_i\}_{i=1}^{NK}$ denotes the few-shot training/support set of the t^{th} ($t > 0$) incremental session, where $N = |\mathcal{C}^t|$ is the number of incremental class and K is the number of samples for each class. $y_i \in \mathcal{C}^t$ denotes the class label of the sample x_i. Take the 5-way 5-shot FSCIL task as an example, $\mathcal{D}^t$ contains 5 (i.e., $K = 5$) samples for each of the 5 (i.e., $N = 5$) incremental classes. After adapting the recognition model to $\mathcal{D}^t$, we are supposed to evaluate the performance of the model on the testing/query samples ($\mathcal{D}^t_{test}$) of all encountered classes (i.e., $\widetilde{\mathcal{C}}^t = \mathcal{C}^0 \cup \mathcal{C}^1 ... \cup \mathcal{C}^t$). In order to avoid violating the privacy concern, we only have access to the training set of the current session as in (Tao et al. 2020; Zhang et al. 2021; Chi et al. 2022).

In this work, we employ the two-step training paradigm to learn the recognition model. **(1) In the pre-training step**, we leverage all the training data ($\mathcal{D}^0$) of the base session to optimize the general classification model, which is composed of a CNN backbone and a fully connected layer. **(2) In the pseudo incremental learning step**, we leverage the base dataset $\mathcal{D}^0$ to help construct the pseudo incremental task to simulate the real testing scenario, thereby ensuring the generalization ability of the recognition model.

Pseudo Informative Episode Construction

Figure 2 shows a pseudo incremental episode/task of the proposed new meta learning paradigm. To match the real-world incremental scenario, each pseudo incremental episode contains G pseudo incremental sessions.

Pseudo Novel Class Generation In the beginning of each episode, we feed the image x_i (support sample or query sample) of the base classes into the CNN backbone (i.e., ResNet (He et al. 2016)) pre-trained on the base session to learn its feature embeddings $\boldsymbol{v_i} \in \mathbb{R}^d$, where d denotes the dimension of the embedding. In this work, we assume that each dimension in the feature vectors follows a Gaussian distribution and estimate the mean vector $\boldsymbol{\mu}_c \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma}_c \in \mathbb{R}^{d\times d}$ for each base class $c \in \mathcal{C}^0$ as follows:

$$\boldsymbol{\mu}_c = \frac{\sum_{j=1}^{n_c} \boldsymbol{v}_j}{n_c}, \tag{1}$$

$$\boldsymbol{\Sigma}_c = \frac{1}{n_c - 1} \sum_{j=1}^{n_c} (\boldsymbol{v}_j - \boldsymbol{\mu}_c)^\intercal (\boldsymbol{v}_j - \boldsymbol{\mu}_c), \tag{2}$$

where n_c denotes the number of (support) samples with label c on the training dataset $\mathcal{D}^0$ and $^\intercal$ denotes the transpose operation. In this way, we can leverage the Gaussian distribution $X_c \sim N(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ to represent the information of class c.

To mimic the real incremental learning scenario, we need to construct the pseudo incremental tasks for episodic training. Different from existing methods that simply treat part of the base classes as the pseudo novel classes, we propose to employ the distribution-level pairwise mixing to generate a set of pseudo novel classes in the feature space of the novel classes. Specifically, for a pseudo incremental session in an

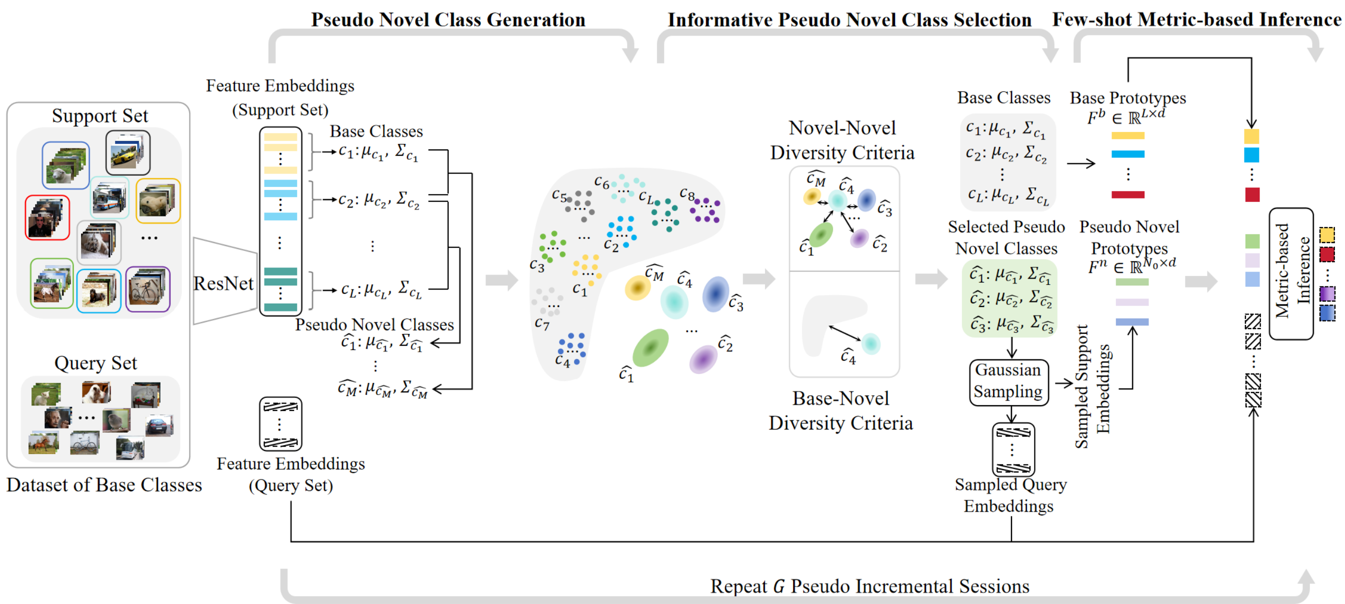


Figure 2: The overall framework of the proposed pseudo informative episode construction.

episode, we randomly select M base class pairs and perform the pairwise mixing on their distributions at first. For the class pair of c_m and c_n, the distribution of the generated pseudo novel class $\hat{c}$ (i.e., $X_{\hat{c}} \sim N(\boldsymbol{\mu}_{\hat{c}}, \boldsymbol{\Sigma}_{\hat{c}})$) is obtained as follows:

$$\boldsymbol{\mu}_{\hat{c}} = \lambda \boldsymbol{\mu}_{c_m} + (1-\lambda)\boldsymbol{\mu}_{c_n}, \tag{3}$$

$$\boldsymbol{\Sigma}_{\hat{c}} = \lambda \boldsymbol{\Sigma}_{c_m} + (1-\lambda)\boldsymbol{\Sigma}_{c_n}, \tag{4}$$

where $\lambda \in [0,1]$ follows the $Beta$ distribution (Johnson, Kotz, and Balakrishnan 1995), i.e., $\lambda \sim Beta(\beta, \beta)$. β denotes the hyper-parameter that controls the probability distribution. In this way, we can obtain a set of pseudo novel classes. Note that we define the mean vector set of the pseudo novel classes as $\mathcal{P} = \{\boldsymbol{\mu}_{\hat{c}}\} \in \mathbb{R}^{M \times d}$.

Informative Pseudo Novel Class Selection Although we have generated a set of pseudo novel classes, not all of them are suitable for the current pseudo incremental session construction. Because some pseudo novel classes may be similar to the other novel classes and the base classes, which cannot provide rich new knowledge to facilitate the incremental learning. As mentioned above, the pseudo novel class of an informative pseudo incremental task should differ significantly from the other novel classes and have high information gain over the base classes. In this work, we propose two diversity criteria strategies (i.e., *novel-novel diversity criteria* and *base-novel diversity criteria*) to select the informative pseudo novel classes.

The *novel-novel diversity criteria* aims to select a set of pseudo novel classes that have diverse characteristics based on the mean vector set $\mathcal{P} \in \mathbb{R}^{M \times d}$. We think that the pseudo novel class with great diversity should be dissimilar to the others in the feature representation. Therefore, we first estimate the similarities of the pseudo novel class $\hat{c}$ and the others as follows:

$$\mathcal{P}' = \mathcal{P} W_p, \quad R = \mathcal{P}' \mathcal{P}'^{\intercal}, \tag{5}$$

$$\tilde{R} = R(J - I), \quad \mathrm{u}_{\hat{c}} = sum(\tilde{R}_{\hat{c}}), \tag{6}$$

where $W_p \in \mathbb{R}^{d \times d}$ denotes the transformation weight matrix and sum denotes the summation operation. $R \in \mathbb{R}^{M \times M}$ is the similarity matrix, $J \in \mathbb{R}^{M \times M}$ is matrix with the element of 1 and $I \in \mathbb{R}^{M \times M}$ is the identity matrix. $\tilde{R}$ is generated by removing the diagonal elements of matrix R. $\mathrm{u}_{\hat{c}}$ represents the summation of similarities between the pseudo novel class $\hat{c}$ and the other novel classes. The class with smaller similarity score $\mathrm{u}_{\hat{c}}$ has greater diversity. In this work, we discard the classes with large scores to reserve rN_0 pseudo novel classes.

The *base-novel diversity criteria* aims to select the classes that can provide rich new information beyond the base classes, i.e., having high information gain relative to the base classes. Specifically, we first estimate the information divergence between the distribution of the base classes ($\{X_c\}_{c=1}^{L}$) and the pseudo novel class ($X_{\hat{c}}$) as follows:

$$\mathrm{h}_{\hat{c}} = \sum_{c=1}^{L} D(X_c || X_{\hat{c}}) = \frac{1}{2} \sum_{c=1}^{L} log \frac{|\boldsymbol{\Sigma}_{\hat{c}}|}{|\boldsymbol{\Sigma}_c|} - d + \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\hat{c}}^{-1} \Sigma_c\right) + (\boldsymbol{\mu}_c - \boldsymbol{\mu}_{\hat{c}}) \boldsymbol{\Sigma}_{\hat{c}}^{-1} (\boldsymbol{\mu}_c - \boldsymbol{\mu}_{\hat{c}})^{\intercal}, \tag{7}$$

where $D(*||*)$ denotes the K-L divergence measurement function and Tr denotes the trace of the matrix. The class with larger value has larger divergence with the base classes. Then, we reserve the top N_0 pseudo novel classes that have the largest values to help construct the pseudo informative incremental session. Note that we denote the mean vector set of the selected pseudo novel classes as $\tilde{P} \in \mathbb{R}^{N_0 \times d}$.

After obtaining the N_0 pseudo informative novel classes, we randomly generate K_0 sampled support embeddings for each pseudo novel class from the corresponding Gaussian distribution to simulate the few-shot scenario (i.e., N_0-way K_0-shot setting). We also randomly generate some sampled query embeddings for the pseudo novel classes. In the current pseudo incremental session, the features of the many-shot samples in the base class and the few-shot samples in the pseudo novel class are averaged respectively to generate their corresponding prototypes $F^b \in \mathbb{R}^{L\times d}$ and $F^n \in \mathbb{R}^{N_0\times d}$.

Few-Shot Metric-based Inference In this work, we feed the query samples, the base class prototypes and the enriched novel class prototypes to a metric-based inference network to conduct the few-shot classification. Following the existing methods (Zhang et al. 2021; Ye et al. 2020), we employ a representative metric-based method FEAT (Ye et al. 2020) in this module.

Specifically, for each query sample q, we first combine its features $\boldsymbol{v_q}$ with the prototypes of all classes (i.e., $F = \{F^b, F^n\}$). Next, we feed the combined features into the self attention layer to obtain the updated features of the query sample $\hat{v}_q$ and the class prototypes $\hat{F}$. Finally, we estimate the cosine similarities between the updated query sample and the class prototypes to predict the class label of q, i.e., $\hat{y}_q$:

$$\hat{y}_q = \underset{i}{argmax}(cos(\hat{v}_q,\ \hat{F}_i)), \tag{8}$$

where cos denotes the cosine similarity measurement, i.e., $cos(a, b) = \frac{a\cdot b}{\|a\|\|b\|}$. $\|*\|$ denotes the L_2-$norm$. More details about the architecture can be found in (Zhang et al. 2021; Ye et al. 2020).

Optimization and Evaluation

In each pseudo incremental task, we propose to optimize the proposed framework with the following classification loss function:

$$\mathcal{L}_{cls} = \sum_{q=1}^{O} \mathcal{L}_{ce}(\hat{y}_q, y_q), \tag{9}$$

where $\mathcal{L}_{ce}$ denotes the cross-entropy loss function and O denotes the number of the query samples.

After finishing the pseudo incremental learning stage, we output the recognition model for real-world few-shot class-incremental learning. For each incremental session t, we first leverage the support samples of the novel classes (i.e., $\mathcal{D}^t$) to generate their corresponding prototypes. Then, we combine the features of the query samples with the prototypes of the novel classes and old classes to perform the self attention mechanism. Finally, we employ Eq. (8) to conduct the metric-based inference for the current incremental session.

Experiments

Datasets

In this work, we perform extensive experiments to evaluate our approach and compare it with the state-of-the-art methods on three widely-used classification benchmarks (CUB200 (Wah et al. 2011), miniImageNet (Vinyals et al. 2016) and CIFAR100 (Krizhevsky and Hinton 2009))) for few-shot class-incremental learning (FSCIL) task.

We follow recent work (Zhang et al. 2021; Zhou et al. 2023) to establish the incremental setting. Specifically, for CUB200 dataset, the first 100 classes are treated as the base classes and the remaining 100 classes are divided to construct 10 incremental sessions. In each incremental session, we sample 5 images for each of the 10 novel classes, i.e., 10-way 5-shot setting. For CIFAR100 and miniImageNet datasets, 60 classes are selected to construct the base session and the remaining 40 classes are used to form 8 incremental sessions. Each incremental session contains 5 classes with 5 images per class, i.e., 5-way 5-shot setting.

Experimental Setup

Implementation details. Following the existing methods (Zhang et al. 2021; Zhou et al. 2023), we adopt the ResNet-18 as the backbone for CUB200 and miniImageNet datasets and utilize the ResNet-20 to embed the images of the CIFAR100 dataset. The dimension of the visual embedding is set to 64 for CIFAR dataset and 512 for CUB200 and miniImageNet datasets. The proposed PIEC is trained using SGD optimizer with momentum. In the pre-training procedure, we optimize the backbone followed by the fully connected layer for 100 epochs with the batch size of 128. The initial learning rate is 0.1, which is decayed by 0.1 at epoch 60 and 70. In the pseudo incremental learning stage, we construct two pseudo incremental sessions in each episode, i.e., $G = 2$. In this work, we set β to 1.0 and M to 100. The recognition model is trained for 100 epochs with the learning rate of 0.0002. We decay the learning rate by 0.5 every 20 epochs.

For each base class on CIFAR100 and miniImageNet datasets, we split the corresponding training images into 280 support samples and 20 query samples. For each base class on CUB200 dataset, we select 10 and 15 images to build the support set and query set, respectively. For the pseudo novel class, we employ the 5-way 5-shot setting (i.e., $N_0 = 5$, $K_0 = 5$) for the CIFAR100 and miniImageNet datasets, and the 10-way 5-shot setting (i.e., $N_0 = 10$, $K_0 = 5$) for the CUB200 dataset. The number of the query samples in each pseudo novel class is the same as that in the base class. All models are deployed with PyTorch on the A100. We train the pseudo incremental tasks for 50 episodes and set the random seed to 3.

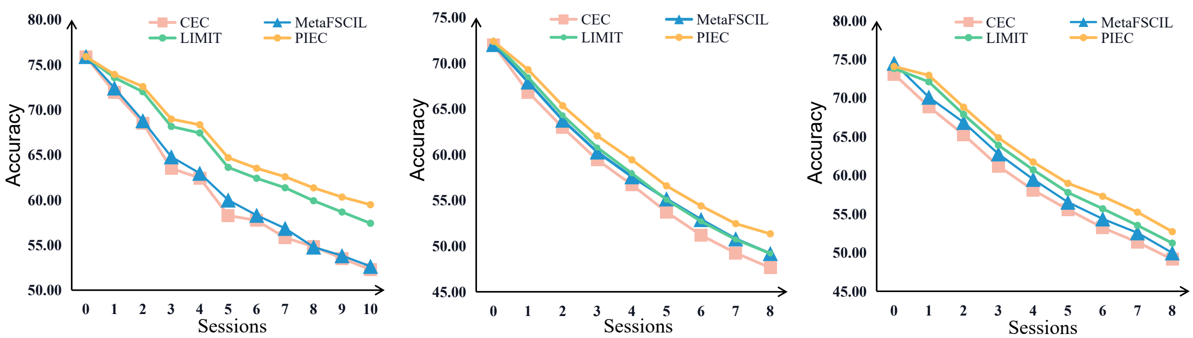
Evaluation protocol. Following the existing methods (Zhang et al. 2021; Zhou et al. 2023), we evaluate the classification accuracy $\mathcal{A}_t$ of the model after the t^{th} session on the testing set $\mathcal{D}^t_{test}$ to estimate its recognition ability. In order to quantitatively estimate the ability of resisting forgetting, we further measure the performance dropping (PD) between the last session and the base session, i.e., $PD = \mathcal{A}_0 - \mathcal{A}_T$. Smaller PD and larger $\mathcal{A}_t$ mean that the model obtains better FSCIL performance.

Main Results

In this work, we compare the proposed PIEC with the recent state-of-the-art methods on three public datasets.

Method	Acc. in each session (%) ↑											PD ↓	Improvement of ours in PD
	0	1	2	3	4	5	6	7	8	9	10		
Finetune-CNN	68.68	43.7	25.05	17.72	18.08	16.95	15.1	10.6	8.93	8.93	8.47	60.21	**+43.80**
iCaRL (Rebuffi et al. 2017)	68.68	52.65	48.61	44.16	36.62	29.52	27.83	26.26	24.01	23.89	21.16	47.52	**+31.11**
EEIL (Castro et al. 2018b)	68.68	53.63	47.91	44.2	36.3	27.46	25.93	24.7	23.95	24.13	22.11	46.57	**+30.16**
NCM (Hou et al. 2019)	68.68	57.12	44.21	28.78	26.71	25.66	24.62	21.52	20.12	20.06	19.87	48.81	**+32.40**
TOPIC (Tao et al. 2020)	68.68	62.49	54.81	49.99	45.25	41.4	38.35	35.36	32.22	28.31	26.28	42.40	**+25.99**
SF-FSCIL (Cheraghian et al. 2021)	68.78	59.37	59.32	54.96	52.58	49.81	48.09	46.32	44.33	43.43	43.23	25.55	**+9.14**
ERDFR (Liu et al. 2022)	75.90	72.14	68.64	63.76	62.58	59.11	57.82	55.89	54.92	53.58	52.39	23.51	**+7.10**
FACT (Zhou et al. 2022)	75.90	73.23	70.84	66.13	65.56	62.15	61.74	59.83	58.41	57.89	56.94	18.96	**+2.55**
DSN (Yang et al. 2023a)	76.06	72.18	69.57	66.68	64.42	62.12	60.16	58.94	56.99	55.10	54.21	21.85	**+5.44**
SoftNet(c=0.9) (Kang et al. 2023)	78.07	74.58	71.37	67.54	65.37	62.60	61.07	59.37	57.53	57.21	56.75	21.32	**+4.91**
WaRP (Kim et al. 2023)	77.74	74.15	70.82	66.90	65.01	62.64	61.40	59.86	57.95	57.77	57.01	20.73	**+4.32**
MCNet (Ji et al. 2023)	77.57	73.96	70.47	65.81	66.16	63.81	62.09	61.82	60.41	60.09	59.08	18.49	**+2.08**
MgSvF (Zhao et al. 2024)	72.29	70.53	67.00	64.92	62.67	61.89	59.63	59.15	57.73	55.92	54.33	17.96	**+1.55**
NC-FSCIL (Yang et al. 2023b)	80.45	75.98	72.30	70.28	68.17	65.16	64.43	63.25	60.66	60.01	59.44	21.01	**+4.60**
SAVC (Song et al. 2023)	81.85	77.92	74.95	70.21	69.96	67.02	66.16	65.30	63.84	63.15	62.50	19.35	**+2.94**
H^2G (Cui et al. 2024)	63.20	62.61	59.83	56.82	55.07	53.06	51.56	50.05	47.50	46.82	45.87	17.33	**+0.92**
MICS (Kim et al. 2024)	78.77	75.37	72.30	68.72	67.45	65.40	64.72	63.39	61.89	61.89	61.37	17.40	**+0.99**
SPPR* (Zhu et al. 2021)	68.68	61.85	57.43	52.68	50.19	46.88	44.65	43.07	40.17	39.63	37.33	31.35	**+14.94**
CEC* (Zhang et al. 2021)	75.85	71.94	68.50	63.50	62.43	58.27	57.73	55.81	54.83	53.50	52.28	23.57	**+7.16**
MetaFSCIL* (Chi et al. 2022)	75.90	72.41	68.78	64.78	62.96	59.99	58.30	56.85	54.78	53.82	52.64	23.26	**+6.85**
LIMIT* (Zhou et al. 2023)	75.89	73.55	71.99	68.14	67.42	63.61	62.40	61.35	59.91	58.66	57.41	18.48	**+2.07**
PIEC*	75.88	73.91	72.57	68.96	68.33	64.68	63.51	62.57	61.34	60.32	59.47	16.41	

Table 1: Comparison with the state-of-the-art methods on CUB200 dataset with 10-way 5-shot FSCIL setting. * indicates the meta-learning methods that construct pseudo incremental episodes.



(a) CUB200 (b) miniImageNet (c) CIFAR100

Figure 3: Classification accuracy of each few-shot incremental session on the three benchmarks.

To illustrate the accuracy degradation of the classification model during the incremental learning procedure, we show the classification results on the three benchmarks in Figure 3. **Note that, to make a fair comparison with existing meta-learning methods (e.g., CEC, MetaFSCIL, and LIMIT), we use the same model pre-trained on the base session (i.e., session = 0) with the accuracy around 75.90%.** From these experimental results, we have the following observations: The proposed PIEC consistently outperforms the SOTA methods on the three datasets in terms of the performance dropping (PD), which verifies its superiority in conducting the FSCIL tasks. As shown, compared with the recent meta learning-based methods CEC/MetaFSCIL/LIMIT, our method achieves the relative improvement of **(1) 7.16%** (16.41% vs. 23.57%)/**6.85%** (16.41% vs. 23.26%)/**2.07%** (16.41% vs. 18.48%) on the CUB200 dataset, **(2) 3.33%** (21.04% vs. 24.37%)/**1.81%** (21.04% vs. 22.85%)/**2.09%** (21.04% vs. 23.13%) on the miniImageNet dataset and **(3) 2.57%** (21.36% vs. 23.93%)/**3.17%** (21.36% vs. 24.53%)/**1.22%** (21.36% vs. 22.58%) on the CIFAR100 dataset, respectively.

Ablation Studies

In this section, we conduct ablation studies on the CUB200 dataset with ResNet-18 using the 10-way 5-shot FSCIL setting. In this work, we propose two diversity criteria to help select the informative pseudo novel classes. To reveal their importance, we design three variants of the proposed PIEC, i.e., PIEC w/o NN, PIEC w/o BN and PIEC w/o NN-BN, which corresponds to the PIEC without the *novel-novel diversity criteria*, the *base-novel diversity criteria* and both of them. PIEC w/o NN-BN means that we randomly select N_0 pseudo novel classes to construct the pseudo incremental task. As shown in Table 2, the performance dropping (PD) of PIEC w/o NN and PIEC w/o BN increase by 1.02% and 1.36% respectively compared with PIEC. Without these two criteria strategies, the PD increases by 2.04%. These results verify the effectiveness of the proposed selection criteria strategies.

Hyperparameter Analysis

Impact of the Hyperparameter β**.** In this work, we leverage the $Beta$ distribution to help generate the mixing coef-

Method	Acc. in each session (%) ↑											PD ↓	Improvement in PD
	0	1	2	3	4	5	6	7	8	9	10		
PIEC w/o NN	75.80	73.78	72.42	68.86	68.01	64.32	63.25	61.94	60.64	59.37	58.37	17.43	**+1.02**
PIEC w/o BN	75.74	73.63	72.26	68.70	67.85	64.18	63.03	61.76	60.40	59.02	57.97	17.77	**+1.36**
PIEC w/o NN-BN	75.66	73.10	71.68	68.08	67.39	63.62	62.38	61.21	59.83	58.46	57.21	18.45	**+2.04**
PIEC	75.88	73.91	72.57	68.96	68.33	64.68	63.51	62.57	61.34	60.32	59.47	16.41	

Table 2: Ablation studies on CUB200 dataset with 10-way 5-shot FSCIL setting.

ficient λ. To reveal its impact, we vary the degree of β and draw their corresponding PD in Figure 4. As shown, when β is too large or small, the performance change significant, which verifies the effect of distribution-level pairwise mixing for the pseudo-novel class initialization.

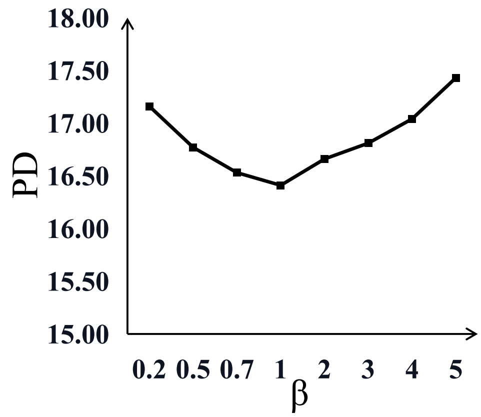


Figure 4: Analysis for the hyperparameter β.

Impact of the Selection Ratio. The selection number of pseudo novel classes (i.e., rN_0) influences the final performance. To reveal the impact of the selection ratio r, we range it from 1.0 to 5.0 and show the PD in Figure 5. From the figure, we can observe that the proposed method obtains the best results when the ratio $r = 2.0$.

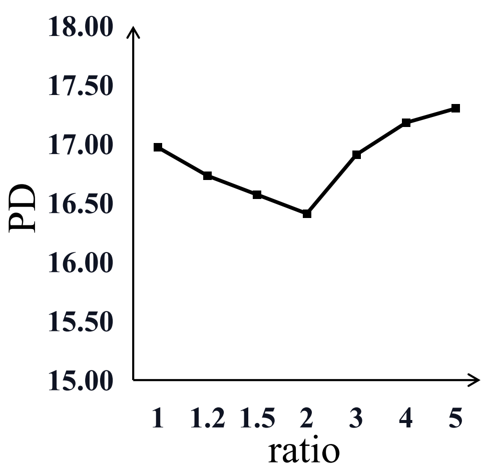


Figure 5: Analysis for the hyperparameter ratio r.

Impact of the Number of the Pseudo Incremental Sessions. The real-world testing phase contains multiple incremental sessions instead of one session. In this part, we vary the number of the pseudo incremental sessions from 1 to 9 to study its impact. As illustrated in Figure 6, we can observe that the classification results of the last session improved slightly when the number of the pseudo incremental sessions is greater than 2. In this work, we set the number of the pseudo incremental sessions to 2 (i.e., $G = 2$).

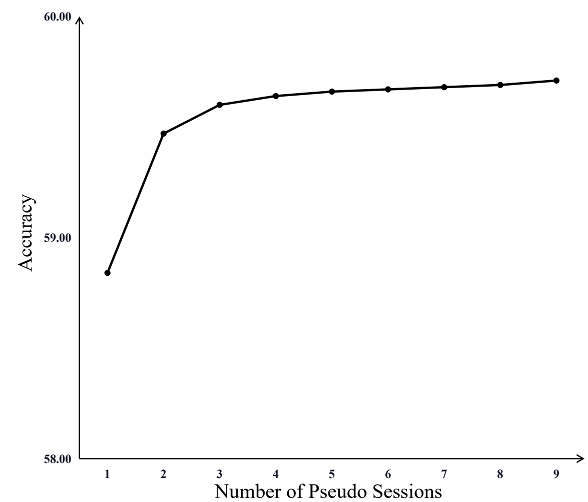


Figure 6: Analysis for the number of the pseudo sessions.

Conclusion

In this work, we propose a new pseudo informative episode construction framework to conduct the FSCIL task. First of all, we perform the distribution-level pairwise mixing on the base classes to generate a set of pseudo novel classes. Then, we propose two diversity criteria to select the informative ones to construct the pseudo incremental session. Finally, we feed the query samples and the prototypes of all classes into a transformer architecture for metric-based inference. We conduct comprehensive experiments on three classification benchmarks and the results show that the proposed method can obtain the state-of-the-art performance. In the future, we would like to deploy the proposed PIEC to a more realistic incremental learning task, where the new samples may belong to the classes encountered before. This task requires the model to be able to leverage the new samples to improve the performance of the few-shot old classes, which is a challenging research line that has not been widely explored.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grants 62036012, U23A20387, 62322212, 62072455, in part by Pengcheng Laboratory Research Project under Grant PCL2023A08, in part by Alibaba Innovative Research Program, and also in part by CAS Project for Young Scientists in Basic Research(YSBR-116).

References

Akyürek, A. F.; Akyürek, E.; Wijaya, D.; and Andreas, J. 2022. Subspace Regularizers for Few-Shot Class Incremental Learning. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*.

Castro, F. M.; Marín-Jiménez, M. J.; Guil, N.; Schmid, C.; and Alahari, K. 2018a. End-to-End Incremental Learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 11216, 241–257.

Castro, F. M.; Marín-Jiménez, M. J.; Guil, N.; Schmid, C.; and Alahari, K. 2018b. End-to-End Incremental Learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 11216, 241–257.

Chaudhry, A.; Ranzato, M.; Rohrbach, M.; and Elhoseiny, M. 2019. Efficient Lifelong Learning with A-GEM. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*.

Chen, C.; Yang, X.; Xu, C.; Huang, X.; and Ma, Z. 2021. ECKPN: Explicit Class Knowledge Propagation Network for Transductive Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6596–6605.

Chen, K.; and Lee, C. 2021. Incremental few-shot learning via vector quantization in deep embedded space. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*.

Cheraghian, A.; Rahman, S.; Ramasinghe, S.; Fang, P.; Simon, C.; Petersson, L.; and Harandi, M. 2021. Synthesized Feature based Few-Shot Class-Incremental Learning on a Mixture of Subspaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 8641–8650.

Chi, Z.; Gu, L.; Liu, H.; Wang, Y.; Yu, Y.; and Tang, J. 2022. MetaFSCIL: A Meta-Learning Approach for Few-Shot Class Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14146–14155.

Cui, Y.; Yu, Z.; Peng, W.; Tian, Q.; and Liu, L. 2024. Rethinking Few-Shot Class-Incremental Learning With Open-Set Hypothesis in Hyperbolic Geometry. *IEEE Transactions on Multimedia (TMM)*, 26: 5897–5910.

Dong, S.; Hong, X.; Tao, X.; Chang, X.; Wei, X.; and Gong, Y. 2021. Few-Shot Class-Incremental Learning via Relation Knowledge Distillation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 1255–1263.

Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 1126–1135.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hersche, M.; Karunaratne, G.; Cherubini, G.; Benini, L.; Sebastian, A.; and Rahimi, A. 2022. Constrained Few-shot Class-incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9047–9057.

Hou, S.; Pan, X.; Loy, C. C.; Wang, Z.; and Lin, D. 2019. Learning a Unified Classifier Incrementally via Rebalancing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 831–839.

Ji, Z.; Hou, Z.; Liu, X.; Pang, Y.; and Li, X. 2023. Memorizing Complementation Network for Few-Shot Class-Incremental Learning. *IEEE Transactions on Image Processing*, 32: 937–948.

Johnson, N. L.; Kotz, S.; and Balakrishnan, N. 1995. Chapter 25:Beta Distributions. *Continuous Univariate Distributions*, 2.

Kang, H.; Yoon, J.; Madjid, S. R. H.; Hwang, S. J.; and Yoo, C. D. 2023. On the Soft-Subnetwork for Few-Shot Class Incremental Learning. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 1–23. OpenReview.net.

Kim, D.; Han, D.; Seo, J.; and Moon, J. 2023. Warping the Space: Weight Space Rotation for Class-Incremental Few-Shot Learning. In *Proceedings of the International Conference on Learning Representations (ICLR))*, 1–19. OpenReview.net.

Kim, S.; Jeong, Y.; Park, J. S.; and Yoon, S. W. 2024. MICS: Midpoint Interpolation to Learn Compact and Separated Representations for Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2225–2234.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences (PNAS)*, 114(13): 3521–3526.

Krizhevsky, A.; and Hinton, G. 2009. Learning multiple layers of features from tiny images. Technical report.

Liu, H.; Gu, L.; Chi, Z.; Wang, Y.; Yu, Y.; Chen, J.; and Tang, J. 2022. Few-Shot Class-Incremental Learning via Entropy-Regularized Data-Free Replay. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 13684, 146–162.

Liu, Y.; Lee, J.; Park, M.; Kim, S.; Yang, E.; Hwang, S. J.; and Yang, Y. 2019. Learning to Propagate Labels: Transductive Propagation Network for Few-Shot Learning. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*.

Ravi, S.; and Larochelle, H. 2017. Optimization as a Model for Few-Shot Learning. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*.

Rebuffi, S.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. iCaRL: Incremental Classifier and Representation Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5533–5542.

Rusu, A. A.; Rao, D.; Sygnowski, J.; Vinyals, O.; Pascanu, R.; Osindero, S.; and Hadsell, R. 2019. Meta-Learning with Latent Embedding Optimization. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*.

Satorras, V. G.; and Estrach, J. B. 2018. Few-Shot Learning with Graph Neural Networks. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*.

Snell, J.; Swersky, K.; and Zemel, R. S. 2017. Prototypical Networks for Few-shot Learning. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 4077–4087.

Song, Z.; Zhao, Y.; Shi, Y.; Peng, P.; Yuan, L.; and Tian, Y. 2023. Learning with Fantasy: Semantic-Aware Virtual Contrastive Constraint for Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 24183–24192. IEEE.

Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P. H. S.; and Hospedales, T. M. 2018. Learning to Compare: Relation Network for Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1199–1208.

Tao, X.; Hong, X.; Chang, X.; Dong, S.; Wei, X.; and Gong, Y. 2020. Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12180–12189.

Vinyals, O.; Blundell, C.; Lillicrap, T.; Kavukcuoglu, K.; and Wierstra, D. 2016. Matching Networks for One Shot Learning. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 3630–3638.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The Caltech-UCSD Birds-200-2011 Dataset. Technical report.

Yang, B.; Lin, M.; Zhang, Y.; Liu, B.; Liang, X.; Ji, R.; and Ye, Q. 2023a. Dynamic Support Network for Few-Shot Class Incremental Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(3): 2945–2951.

Yang, Y.; Yuan, H.; Li, X.; Lin, Z.; Torr, P. H. S.; and Tao, D. 2023b. Neural Collapse Inspired Feature-Classifier Alignment for Few-Shot Class-Incremental Learning. In *ICLR*.

Ye, H.; Hu, H.; Zhan, D.; and Sha, F. 2020. Few-Shot Learning via Embedding Adaptation With Set-to-Set Functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8805–8814.

Zhang, C.; Song, N.; Lin, G.; Zheng, Y.; Pan, P.; and Xu, Y. 2021. Few-Shot Incremental Learning With Continually Evolved Classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12455–12464.

Zhao, H.; Fu, Y.; Kang, M.; Tian, Q.; Wu, F.; and Li, X. 2024. MgSvF: Multi-Grained Slow versus Fast Framework for Few-Shot Class-Incremental Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 46(3): 1576–1588.

Zhou, D.; Wang, F.; Ye, H.; Ma, L.; Pu, S.; and Zhan, D. 2022. Forward Compatible Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9036–9046.

Zhou, D.; Ye, H.; Ma, L.; Xie, D.; Pu, S.; and Zhan, D. 2023. Few-Shot Class-Incremental Learning by Sampling Multi-Phase Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(11): 12816–12831.

Zhu, K.; Cao, Y.; Zhai, W.; Cheng, J.; and Zha, Z. 2021. Self-Promoted Prototype Refinement for Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6801–6810.