# CVPR

## 2024

### OrCo: Towards Better Generalization via Orthogonality and Contrast for Few-Shot Class-Incremental Learning

Few-Shot Class-Incremental Learning (FSCIL) intro-duces a paradigm in which the problem space expands with limited data. FSCIL methods inherently face the chal-lenge of catastrophic forgetting as data arrives incremen-tally, making models susceptible to overwriting previously acquired knowledge. Moreover, given the scarcity of la-beled samples available at any given time, models may be prone to overfitting and find it challenging to strike a bal-ance between extensive pretraining and the limited incre-mental data. To address these challenges, we propose the OrCo framework built on two core principles: features' orthogonality in the representation space, and contrastive learning. In particular, we improve the generalization of the embedding space by employing a combination of su-pervised and self-supervised contrastive losses during the pretraining phase. Additionally, we introduce OrCo loss to address challenges arising from data limitations during in-cremental sessions. Through feature space perturbations and orthogonality between classes, the OrCo loss maxi-mizes margins and reserves space for the following incre-mental data. This, in turn, ensures the accommodation of incoming classes in the feature space without compromising previously acquired knowledge. Our experimental results showcase state-of-the-art performance across three bench-mark datasets, including mini-ImageNet, CIFAR100, and CUB datasets. Code is available at: https://github.com/noorahmedds/OrCo.

### Calibrating Higher-Order Statistics for Few-Shot Class-Incremental Learning with Pre-trained Vision Transformers

Few-shot class-incremental learning (FSCIL) aims to adapt the model to new classes from very few data (5 samples) without forgetting the previously learned classes. Recent works in many-shot CIL (MSCIL) (using all available training data) exploited pre-trained models to reduce forgetting and achieve better plasticity. In a similar fashion, we use ViT models pre-trained on large-scale datasets for few-shot settings, which face the critical issue of low plasticity. FSCIL methods start with a many-shot first task to learn a very good feature extractor and then move to the few-shot setting from the second task onwards. While the focus of most recent studies is on how to learn the many-shot first task so that the model generalizes to all future few-shot tasks, we explore in this work how to better model the few-shot data using pre-trained models, irrespective of how the first task is trained. Inspired by recent works in MSCIL, we explore how using higher-order feature statistics can influence the classification of few-shot classes. We identify the main challenge of obtaining a good covariance matrix from few-shot data and propose to calibrate the covariance matrix for new classes based on semantic similarity to the many-shot base classes. Using the calibrated feature statistics in combination with existing methods significantly improves few-shot continual classification on several FSCIL benchmarks. Code is available at https://github.com/dipamgoswami/FSCIL-Calibration.

## 2025

# AAAI

## 2024

### M2SD:Multiple Mixing Self-Distillation for Few-Shot Class-Incremental Learning

Few-shot Class-incremental learning (FSCIL) is a challenging task in machine learning that aims to recognize new classes from a limited number of instances while preserving the ability to classify previously learned classes without retraining the entire model. This presents challenges in updating the model with new classes using limited training data, particularly in balancing acquiring new knowledge while retaining the old. We propose a novel method named Multiple Mxing Self-Distillation (M2SD) during the training phase to address these issues. Specifically, we propose a dual-branch structure that facilitates the expansion of the entire feature space to accommodate new classes. Furthermore, we introduce a feature enhancement component that can pass additional enhanced information back to the base network by self-distillation, resulting in improved classification performance upon adding new classes. After training, we discard both structures, leaving only the primary network to classify new class instances. Extensive experiments demonstrate that our approach achieves superior performance over previous state-of-the-art methods.

## 2025

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

Few-Shot Class-Incremental Learning (FSCIL) studies how to empower the machine learning system to learn novel classes with only a few annotated examples continually. 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

### Few-Shot Audio-Visual Class-Incremental Learning with Temporal Prompting and Regularization

Audio-Visual Learning (AVL) aims at the audio-visual perception with both audio and vision modalities. AVL also suffers from data insufficiency in many applications as with other unimodal tasks. Concurrently, AVL often needs to continuously learn over time rather than all knowledge simultaneously. Considering the above two perspectives, our work mainly focuses on benchmarking the unexplored Few-Shot Audio-Visual Class-Incremental Learning (FS-AVCIL), i.e., continually perceiving novel categories described by a limited number of labeled examples with audio and visual modalities. Firstly, we provide the detailed task configuration together with a thorough analysis of the challenges in FS-AVCIL: (1) how to efficiently learn and fuse multimodal information with limited labeled examples; and (2) how to alleviate catastrophic forgetting cross-modal semantic correlations with limited data. Then, we propose an efficient framework based on Vision Transformer to solve FS-AVCIL. This framework contains two parts: temporal-residual prompting for audio-visual synergy adapter and temporal prompt regularization. Specifically, temporal-residual prompting is incorporated into the audio-visual adapter to efficiently finetune the pre-trained foundation model with limited data and capture audio-visual correlation by learning temporal-relevant prompts. Besides, we regularize temporal-relevant prompts to memorize previous knowledge by fully using the temporal knowledge from various perspectives. This framework is validated in audio-visual classification tasks under the FS-AVCIL scenario, and extensive experiments demonstrate its superior performance.

### AnchorInv: Few-Shot Class-Incremental Learning of Physiological Signals via Feature Space-Guided Inversion

Deep learning models have demonstrated exceptional performance in a variety of real-world applications. These successes are often attributed to strong base models that can generalize to novel tasks with limited supporting data while keeping prior knowledge intact. However, these impressive results are based on the availability of a large amount of high-quality data, which is often lacking in specialized biomedical applications. In such fields, models are often developed with limited data that arrive incrementally with novel categories. This requires the model to adapt to new information while preserving existing knowledge. Few-Shot Class-Incremental Learning (FSCIL) methods offer a promising approach to addressing these challenges, but they also depend on strong base models that face the same aforementioned limitations. To overcome these constraints, we propose AnchorInv following the straightforward and efficient buffer-replay strategy. Instead of selecting and storing raw data, AnchorInv generates synthetic samples guided by anchor points in the feature space. This approach protects privacy and regularizes the model for adaptation. When evaluated on three public physiological time series datasets, AnchorInv exhibits efficient knowledge forgetting prevention and improved adaptation to novel classes, surpassing state-of-the-art baselines.

### Adaptive Decision Boundary for Few-Shot Class-Incremental Learning

Few-Shot Class-Incremental Learning (FSCIL) aims to continuously learn new classes from a limited set of training samples without forgetting knowledge of previously learned classes. Conventional FSCIL methods typically build a robust feature extractor during the base training session with abundant training samples and subsequently freeze this extractor, only fine-tuning the classifier in subsequent incremental phases. However, current strategies primarily focus on preventing catastrophic forgetting, considering only the relationship between novel and base classes, without paying attention to the specific decision spaces of each class. To address this challenge, we propose a plug-and-play Adaptive Decision Boundary Strategy (ADBS), which is compatible with most FSCIL methods. Specifically, we assign a specific decision boundary to each class and adaptively adjust these boundaries during training to optimally refine the decision

### Few-Shot Incremental Learning via Foreground Aggregation and Knowledge Transfer for Audio-Visual Semantic Segmentation

Audio-Visual Semantic Segmentation (AVSS) has gained significant attention in the multi-modal domain, aiming to segment video objects that produce specific sounds in the corresponding audio. Despite notable progress, existing methods still struggle to handle new classes not included in the original training set. To this end, we introduce Few-Shot Incremental Learning (FSIL) to the AVSS task, which seeks to seamlessly integrate new classes with limited incremental samples while preserving the knowledge of old classes. Two challenges arise in this new setting: (1) To reduce labeling costs, old classes within the incremental samples are treated as background, similar to silent objects. Training the model directly with background annotations may worsen the loss of distinctive knowledge about old classes, such as their outlines and sounds. (2) Most existing models adopt early cross-modal fusion with a single-tower design, incorporating more characteristics into class representations, which impedes knowledge transfer between classes based on similarity. To address these issues, we propose a Few-shot Incremental learning framework via class-centric foregrouNd aggreGation and dual-tower knowlEdge tRansfer (FINGER) for the AVSS task, which comprises two targeted modules: (1) The class-centric foreground aggregation gathers class-specific features for each foreground class while disregarding background features. The background class is excluded during training and inferred from the foreground predictions. (2) The dual-tower knowledge transfer postpones cross-modal fusion to separately conduct knowledge transfer for each modality. Extensive experiments validate the effectiveness of the FINGER model, significantly surpassing state-of-the-art methods.

# IJCAI

## 2024

### FineFMPL: Fine-grained Feature Mining Prompt Learning for Few-Shot Class Incremental Learning

Few-Shot Class Incremental Learning (FSCIL) aims to continually learn new classes with few training samples without forgetting already learned old classes. Existing FSCIL methods generally fix the backbone network in incremental sessions to achieve a balance between suppressing forgetting old classes and learning new classes. However, the fixed backbone network causes insufficient learning of new classes from a few samples. Benefiting from the powerful visual and textual understanding ability of Vision-Language (VL) pre-training models, we propose a Fine-grained Feature Mining Prompt Learning (FineFMPL) approach to adapt the VL model to FSCIL, which comprehensively learns and memorizes fine-grained discriminative information of emerging classes. Concretely, the visual probe prompt is firstly proposed to guide the image encoder of VL model to extract global-level coarse-grained features and object-level fine-grained features, and visual prototypes are preserved based on image patch significance, which contains the discriminative characteristics exclusive to the class. Secondly, the textual context prompt is constructed by cross-modal mapping of visual prototypes, feeding into the text encoder of VL model to memorize the class information as textual prototypes. Finally, integrating visual and textual prototypes based on fine-grained feature mining into the model improves the recognition performance of all classes in FSCIL. Extensive experiments on three benchmark datasets demonstrate that our FineFMPL achieves new state-of-the-art. The code is available at https://github.com/PKU-ICST-MIPL/FineFMPL\_IJCAI2024.

### Delve into Base-Novel Confusion: Redundancy Exploration for Few-Shot Class-Incremental Learning

Few-shot class-incremental learning (FSCIL) aims to acquire knowledge from novel classes with limited samples while retaining information about base classes. Existing methods address catastrophic forgetting and overfitting by freezing the feature extractor during novel-class learning. However, these methods usually tend to cause the confusion between base and novel classes, i.e., classifying novel-class samples into base classes.In this paper, we delve into this phenomenon to study its cause and solution. We first interpret the confusion as the collision between the novel-class and the base-class region in the feature space.Then, we find the collision is caused by the label-irrelevant redundancies within the base-class feature and pixel space. Through qualitative and quantitative experiments, we identify this redundancy as the shortcut in the base-class training, which can be decoupled to alleviate the collision. Based on this analysis, to alleviate the collision between base and novel classes, we propose a method for FSCIL named Redundancy Decoupling and Integration (RDI). RDI first decouples redundancies from base-class space to shrink the intra-base-class feature space. Then, it integrates the redundancies as a dummy class to enlarge the inter-base-class feature space. This process effectively compresses the base-class feature space, creating buffer space for novel classes and alleviating the model's confusion between the base and novel classes. Extensive experiments across benchmark datasets, including CIFAR-100, miniImageNet, and CUB-200-2011 demonstrate that our method achieves state-of-the-art performance.

## 2025

# ACM

## 2024

## 2025

# PMLR

## 2024

## 2025