Review

A survey on few-shot class-incremental learning

Songsong Tian [a,b,d], Lusi Li [e], Weijun Li [a,c,d], Hang Ran [a,d], Xin Ning [a,c,d,*], Prayag Tiwari [f,*]

[a] Institute of Semiconductors, Chinese Academy of Sciences, Beijing, 100083, China
[b] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, 100049, China
[c] School of Integrated Circuits, University of Chinese Academy of Sciences, Beijing, 100083, China
[d] Beijing Key Laboratory of Semiconductor Neural Network Intelligent Sensing and Computing Technology, Beijing, 100083, China
[e] Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA
[f] School of Information Technology, Halmstad University, Halmstad, 30118, Sweden

ARTICLE INFO

ABSTRACT

Large deep learning models are impressive, but they struggle when real-time data is not available. Few-shot class-incremental learning (FSCIL) poses a significant challenge for deep neural networks to learn new tasks from just a few labeled samples without forgetting the previously learned ones. This setup can easily leads to catastrophic forgetting and overfitting problems, severely affecting model performance. Studying FSCIL helps overcome deep learning model limitations on data volume and acquisition time, while improving practicality and adaptability of machine learning models. This paper provides a comprehensive survey on FSCIL. Unlike previous surveys, we aim to synthesize few-shot learning and incremental learning, focusing on introducing FSCIL from two perspectives, while reviewing over 30 theoretical research studies and more than 20 applied research studies. From the theoretical perspective, we provide a novel categorization approach that divides the field into five subcategories, including traditional machine learning methods, meta learning-based methods, feature and feature space-based methods, replay-based methods, and dynamic network structure-based methods. We also evaluate the performance of recent theoretical research on benchmark datasets of FSCIL. From the application perspective, FSCIL has achieved impressive achievements in various fields of computer vision such as image classification, object detection, and image segmentation, as well as in natural language processing and graph. We summarize the important applications. Finally, we point out potential future research directions, including applications, problem setups, and theory development. Overall, this paper offers a comprehensive analysis of the latest advances in FSCIL from a methodological, performance, and application perspective.

1. Introduction

In recent years, significant advancements in computing technology and the widespread availability of large-scale datasets have enabled deep neural networks (DNNs) to make remarkable progresses in various computer vision tasks (He, Zhang, Ren, & Sun, 2016; Krizhevsky, Sutskever, & Hinton, 2017). However, many of these successes rely on idealized assumptions and massive amounts of available training data, which may not accurately reflect the real-world scenarios where high-quality data is often scarce. For instance, in scenarios where data arrives incrementally in batches and newly added categories contain very few samples, many existing methods prove to be ineffective.

The goal of few-shot class-incremental learning (FSCIL) is to endow AI with the capability to address the aforementioned challenges. This requires DNN models to learn new tasks incrementally from a small number of labeled samples, without forgetting the previously learned ones (Tao et al., 2020). Since Tao first proposed the concept of FSCIL in Tao et al. (2020), many scholars have extended it to various application scenarios beyond visual tasks because it conforms to human learning patterns and is suitable for real-world applications.

An intuitive method for FSCIL is to fine-tune a base model on a new training set. However, it would lead to catastrophic forgetting (McCloskey & Cohen, 1989) and overfitting, corresponding to two core challenges: the stability-plasticity dilemma and unreliable empirical risk minimization.

- **Stability-plasticity dilemma**

  The stability-plasticity dilemma reflects the contradiction between stability and plasticity. Stability means that a neural network should maintain its learned knowledge and resist changes caused

---

* Corresponding authors.
E-mail addresses: tiansongsong@semi.ac.cn (S. Tian), lusili@cs.odu.edu (L. Li), wjli@semi.ac.cn (W. Li), ranhang@semi.ac.cn (H. Ran), ningxin@semi.ac.cn (X. Ning), prayag.tiwari@ieee.org (P. Tiwari).
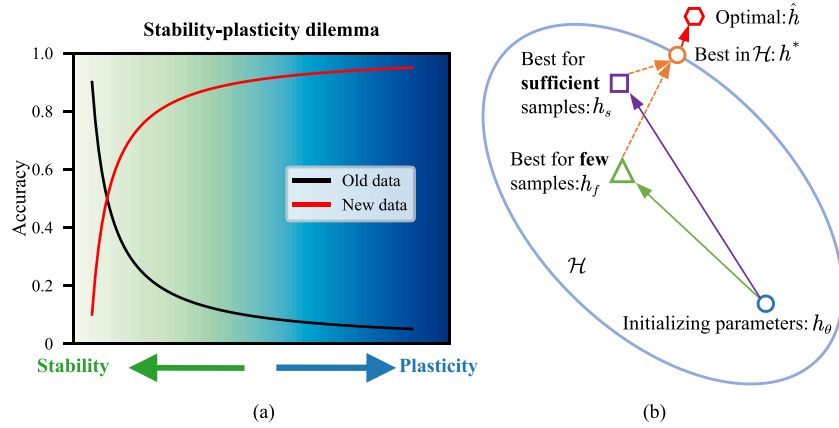
Fig. 1. (a) Stability and plasticity cannot be achieved simultaneously. When a model has high stability, it performs well on old data but struggles with new data. As plasticity increases, the model demonstrates enhanced generalization on new data while gradually forgetting old data; (b) Given a hypothesis space $\mathcal{H}$ and initial parameters h_θ, $\hat{h}$ is the function that minimizes the expected risk, h^* is the function in $\mathcal{H}$ that minimizes the expected risk. h_f and h_s correspond to the functions of minimizing the empirical risk when data samples are few and sufficient, respectively. When the data is sufficient, ERM yields results closer to h^*.

by new inputs. Conversely, plasticity means that the network should have the ability to adapt to new inputs or tasks. Catastrophic forgetting can be seen as a manifestation of the stability-plasticity dilemma. In incremental learning (IL), an overly stable model might fail to learn new tasks or data effectively. In contrast, an exceedingly plastic model might rapidly lose information about previously learned tasks or data. See Fig. 1(a) for more details.

- **Unreliable empirical risk minimization**
  In traditional machine learning frameworks, empirical risk minimization (ERM) aims to optimize the average loss on training data. This strategy works well in large-scale data environments where there are enough samples to ensure statistical consistency during training. However, in the context of few shot learning (FSL), this strategy faces a challenge known as the unreliable empirical risk minimizer problem (Wang, Yao, Kwok, & Ni, 2020). The core of this problem lies in the fact that when the number of training samples is limited or when there is noise in the samples, the ERM strategy may lead to overfitting. Overfitting means that the model performs well on the training data but has poor generalization performance on new, unseen data. This shortfall arises because limited data may not fully represent the true distribution of the entire data generation process, causing the model to capture random noise in the data rather than the underlying true patterns. Fig. 1(b) shows that when training samples are insufficient, the ERM function cannot accurately approximate the optimal expected risk minimization function.

FSCIL, needs to overcome these two challenges, is even more difficult. In addition to the challenges mentioned above, due to the large difference in the number of samples between old and new categories, the model tends to bias towards the larger set of old-class training samples during training or prediction, and the imbalance between base and novel class samples also makes it difficult for the model to learn new categories (Chen & Lee, 2021; Hou, Pan, Loy, Wang, & Lin, 2019; Tao et al., 2020).

Although FSCIL has great potential in real-world applications and has gained significant attentions from researchers, it remains a relatively underexplored area, with a lack of comprehensive reviews. Existing reviews primarily focus on either FSL or IL separately, rather than their combination in FSCIL. For example, Parisi, Kemker, Part, Kanan, and Wermter (2019) focus on continual lifelong learning, though much of the content may not reflect recent advancements. Wang, Yao, et al. (2020) introduced the theoretical foundation of FSL and classified FSL methods from different perspectives. Belouadah, Popescu, and Kanellos (2021) provide a summary of Class-IL in visual tasks only. Zhou et al.

(2023) summarized the latest progress in deep Class-IL from three aspects: data, model, and algorithm.

Our contributions to the field of FSCIL can be summarized as follows:

1. We conducted an in-depth analysis of fundamental and applied research of FSCIL. Our comprehensive review explores various FSCIL approaches, highlighting their advantages, limitations, and performance on benchmark datasets.
2. We revisited the theoretical foundations and practical implementations of various FSCIL approaches and proposed a taxonomy of methods based on the underlying approach or technique. This framework provides a useful guide for researchers and practitioners working on FSCIL.
3. We evaluated the performance of various FSCIL approaches on benchmark datasets, providing insights into the strengths and weaknesses of different methods.
4. We discussed the potential applications of FSCIL in various domains, such as computer vision, natural language processing, and graph analysis. This analysis highlights the broad range of applications for FSCIL and its potential impact on these fields.
5. We identified open research challenges and opportunities for future work in the field of FSCIL. This provides a roadmap for future research in the area and helps to guide the direction of future work.

The remainder of this paper is organized as follows. Section 2 introduces the problem definition of FSCIL and the relevant research background. Section 3 reviews the approaches and notable architectures used in FSL. Section 4 summarizes the existing FSCIL approaches, including traditional machine learning methods, meta learning-based methods, feature and feature space-based methods, replay-based methods, and dynamic network structure-based methods. Section 5 presents the performance of different FSCIL approaches on benchmark datasets. Section 6 discusses the applications of FSCIL in different domains. Section 7 outlines the future research directions in the FSCIL field. Finally, Section 8 concludes the paper.

2. Problem definition

In supervised learning, we want to learn a function $f \in \mathcal{F}: \mathcal{X} \to \mathcal{Y}$ that is able to predict the target vector $y \in \mathcal{Y}$, for a given input sample $x \in \mathcal{X}$. To do so, a model is fed with the training data with sufficient instances: $D = \left\{ \left(x_i, y_i \right) \right\}_{i=1}^{N}$, which contains independent and identically distributed samples from the distribution $P(\mathcal{X}, \mathcal{Y})$. $x_i \in \mathbb{R}^n$ is a training instance from class $y_i \in \mathcal{Y}$ and $\mathcal{Y}$ is the corresponding label
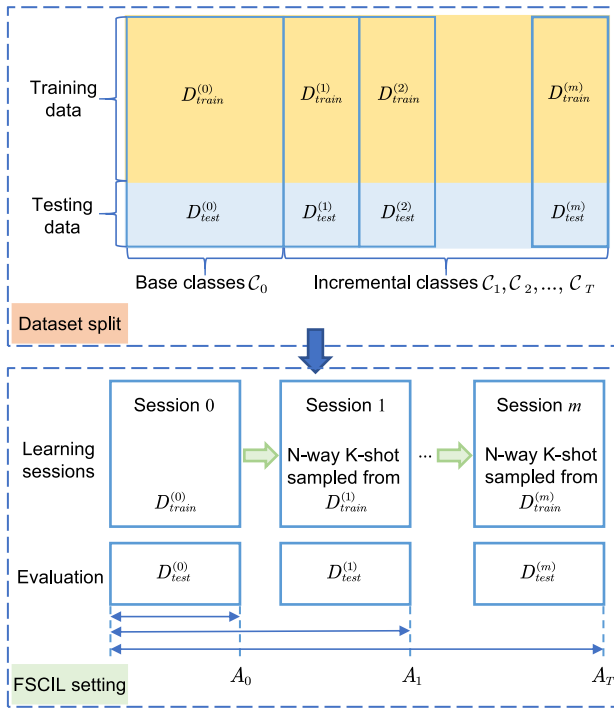
Fig. 2. Dataset setting.
Source: Figure adapted from [Zheng and Zhang (2021)](#).

space. In order to train this function f, we minimize the expected risk over the instance distribution P:

$$\varepsilon_{ex} = \mathbb{E}_{(x,y) \sim P(\mathcal{X}, \mathcal{Y})} \left[\ell \left(f(x), y \right) \right], \tag{1}$$

where $\ell(\cdot, \cdot)$ captures the discrepancy between prediction and ground-truth label. However, the joint distribution P in unknown, therefore the learning algorithm actually aims at minimizing the empirical risk:

$$\varepsilon_{em} = \mathbb{E}_{(x,y) \sim D} \left[\ell \left(f(x), y \right) \right], \tag{2}$$

2.1. Problem formalization

Fig. 2 shows the form of dataset split and the way of FSCIL experiment setup. FSCIL task comprises a base session with sufficient training data and multiple incremental sessions with limited training data. The learning process within each session involves only the data relevant to the current task, while the model is also required to preserve the knowledge of previous tasks when acquiring new ones. The task is to train the model from a continuous data stream in a class-incremental form.

The FSCIL problem is defined as follows. Here we assume an m-step FSCIL task. Let $\left\{ D_{train}^{(0)}, D_{train}^{(1)}, \ldots, D_{train}^{(m)} \right\}$ and $\left\{ D_{test}^{(0)}, D_{test}^{(1)}, \ldots, D_{test}^{(m)} \right\}$ denote the training and testing data for sessions $\{0, 1, \ldots, m\}$, respectively. For session j, it has training data $D_{train}^{(j)}$ with the corresponding label space of $\mathcal{Y}_j$. Training data from different sessions are disjoint, that is, $\mathcal{Y}_a \cap \mathcal{Y}_b = \varnothing (a \neq b)$. The limited instances in $D_{train}^{(j)}$ can be organized as N-way K-shot data format, i.e., there are N classes in the dataset, and each class has K training images. Facing a new dataset $D_{train}^{(j)}$, a model should learn new classes and meanwhile maintain performance over old classes, i.e., minimize the expected risk R over all the seen classes:

$$\mathbb{E}_{(x,y) \sim D_{train}^{(0)} \cup D_{train}^{(1)} \cup \cdots \cup D_{train}^{(j)}} \left[\ell \left(f \left(x; D_{train}^{(j)}, W^{j-1} \right), y \right) \right], \tag{3}$$

In Eq. (3), the learning algorithm f should build the new model based on new dataset $D_{train}^{(j)}$ and current old model W^{j-1}, and minimize the loss over all seen classes. During testing, the model will be evaluated on all seen classes so far. For session j, its testing data $D_{test}^{(j)}$ has the corresponding label space of $\mathcal{Y}_0 \cup \mathcal{Y}_1 \ldots \cup \mathcal{Y}_i$.

2.2. Relevant learning problems

Few-shot Learning. Humans are very skilled at identifying a new object with very few samples. For example, a child can recognize what a "zebra" or "rhinoceros" is with just a few pictures from a book. Inspired by human's rapid learning ability, researchers hope that machine learning models can quickly learn new categories with only a small number of samples after learning a large amount of data for a certain number of categories. This is the problem that FSL aims to solve. In recent years, the concept of FSL has received widespread attention, and there have been many outstanding algorithm models in the field of image classification (Finn, Abbeel, & Levine, 2017; Snell, Swersky, & Zemel, 2017; Zhang, Sung, Qiang, Yang, & Hospedales, 2018). There are mainly three categories of FSL methods: fine-tune based, data augmentation based, and transfer learning based.

Considering a learning task T, FSL deals with a dataset $D = \left\{ D_{train}, D_{test} \right\}$. It consists of a training set $D_{train} = \left\{ (x_i, y_i) \right\}_{i=1}^{I}$, where I is small, and a testing set $D_{test} = \{ x_{test} \}$. Usually, one considers the N-way K-shot classification in which D_{train} contains $I = KN$ examples from N classes each with K examples. FSL is mainly a supervised learning problem (Wang, Yao, et al., 2020). Due to the small size of D_{train}, the model bias, $\varepsilon = |\varepsilon_{ex} - \varepsilon_{em}|$, is too large, making it hard to learn a high-quality prediction function $f \in \mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$.

One-shot Learning. In the late 1980s and 1990s, some researchers already noticed the problem of one-shot learning. It was not until 2003 that Fe-Fei et al. (2003) formally introduced the concept. They believed that when there is only one or a few labeled samples for a new category, the previously learned old categories can help predict the new category (Fei-Fei, Fergus, & Perona, 2006). In the N-way K-shot paradigm, when $N = 1$, FSL is called one-shot learning problem. Since the settings are similar, it is not necessary to distinguish between the two concepts in most cases.

Zero-shot Learning. In the N-way K-shot paradigm, FSL becomes a zero-shot learning problem (ZSL) when $N = 0$. ZSL was first introduced by Palatucci, Pomerleau, Hinton, and Mitchell (2009). Since ZSL does not contain examples with supervised information, it recognizes new sample categories by utilizing semantic label attribute information in the absence of training samples. This approach is inspired by human learning and reasoning capabilities, allowing computers to possess transfer and reasoning abilities. Specifically, a training data for ZSL is formulated as $S = \left\{ (x, y, a(y)) \mid x \in \mathcal{X}_S, y \in \mathcal{Y}_S, a(y) \in \mathcal{A} \right\}$, where $\mathcal{X}_S$ is set of image/features from seen classes, $\mathcal{Y}_S$ is set of seen class labels, $a(y)$ is semantic embedding for class y. The test set is formulated as $U = \left\{ (x, y, a(y)) \mid x \in \mathcal{X}_U, y \in \mathcal{Y}_U, a(y) \in \mathcal{A} \right\}$, where $\mathcal{X}_U$ is set of unseen class image/features, $\mathcal{Y}_U$ is set of unseen class labels, $\mathcal{Y}_U \cap \mathcal{Y}_C = \varnothing$.

Meta Learning. Meta learning is often understood as learning to learn. It is the process of extracting the experience of multiple learning episodes and using this experience to improve future learning performance (Hospedales, Antoniou, Micaelli, & Storkey, 2022). Meta learning is usually divided into two stages. In the meta-training stage, the model is trained using multiple source (or training) tasks to obtain initial network parameters with strong generalization ability. In the meta-testing stage, the settings of the new tasks are the same as those of the source tasks, but these samples have not been seen during the training process. Each task in the training tasks or testing tasks is divided into a support set and a query set. Meta learning has wide applications in the fields of computer vision, reinforcement learning, and architecture search. Meta learning is naturally suitable for FSL, and many studies have used meta-learning as a means of FSL, enabling

the model to learn from a small number of new task samples (Elsken, Staffler, Metzen, & Hutter, 2020; Jamal & Qi, 2019; Ren et al., 2018).

Transfer Learning. Transfer Learning (Zhuang et al., 2020) focuses on the transfer of knowledge across different domains, enabling the transfer of knowledge from domains/tasks with abundant training data to novel domains/tasks with scarce training data. Its definition is as follows.

Definition 1 (*Transfer Learning*). *Given a source domain D_S and a corresponding task T_S, a target domain D_T and a corresponding task T_T. The primary aim of transfer learning is to leverage the knowledge obtained from D_S and T_S to enhance the learning performance of D_T and T_T, where $D_S \neq D_T$ or $T_S \neq T_T$ (Pan & Yang, 2010).*

The key to successful knowledge transfer is the presence of a connection between the two learning activities. If there are few commonalities between domains, knowledge transfer may fail and have a negative impact on the new task. In everyday life, people engage in many instances of transfer learning, such as learning to ride a bike, which makes it easier to learn how to ride a motorcycle. Transfer learning can reduce the reliance on large amounts of target domain data when constructing learning machines. As a result, it has broad applications in zero-shot and few-shot domains, including style transfer, feature space transfer for data augmentation, and label-efficient learning of transferable representations across domains (Azadi et al., 2018; Liu, Wang, Dixit, Kwitt, & Vasconcelos, 2018; Luo, Zou, Hoffman, & Fei-Fei, 2017).

Incremental Learning. The definition of IL can also be expressed using Eq. (3), but the difference from FSCIL is that there are plenty of samples for each incremental category. IL is also known as continuous learning, lifelong learning, or never-ending learning, is a field of machine learning that is gaining increasing attention. It is typically used to address the problem of catastrophic forgetting, where performance on previously learned tasks deteriorates sharply after learning new tasks. The ability of IL is to continuously process a stream of information from the real world while retaining, integrating, and optimizing old knowledge at the same time. The methods proposed in IL are broadly categorized into three categories: replay-based methods, regularization-based methods, and parameter isolation methods (De Lange et al., 2021). Van de Ven and Tolias (2019) proposed three scenarios for IL, including Task-IL, Domain-IL, and Class-IL. And Class-IL is considered the most difficult one since the newly added classes often exhibit high similarity with the already learned classes. Currently, only replay-based methods produce acceptable results for Class-IL.

2.3. Variants of few-shot class incremental learning

Generalized few-shot incremental learning. Before the emergence of FSCIL, similar settings had been proposed in previous research, such as those presented by Gidaris and Komodakis (2018), Qi, Brown, and Lowe (2018), Xie, Li, Lin, Lay Nwe, and Dong (2019), Yoon, Kim, Seo, and Moon (2020). These studies introduced Generalized Few-Shot Incremental Learning (GFSIL). Specifically, a pre-trained model will learn new classes with limited instances. The goal of GFSIL is to maintain classification performance for both old and new classes. However, GFSIL only has one incremental phase, and its data partitioning format is different from FSCIL. For example, CIFAR-100 is randomly divided into 40, 10, and 50 categories, which serve as the meta-training, meta-validation, and meta-testing sets respectively. GFSIL is considered less challenging than FSCIL. To address the challenge of GFSIL, Qi et al. (2018) proposes a solution that utilizes the average feature initialization method with few shots to initialize new class representations. Meanwhile, Gidaris and Komodakis (2018) introduces dynamic few-shot learning to avoid forgetting, which employs a novel attention-based weight generator for few-shot classification. The dot-product calculation method is replaced with the cosine-similarity

function to incorporate the few-shot classification weight generator into the recognition system. Ren, Liao, Fetaya, and Zemel (2019) proposes an Attention Attractor Network to regulate the learning of novel classes. Additionally, Yoon et al. (2020) suggests a method for fusing base features, while Ye, Hu, and Zhan (2021) puts forward the idea of synthesizing few-shot classifiers with a shared neural dictionary. Xie et al. (2019) introduces Meta Module Generation (MetaMG) which utilizes meta-learning to learn a set of meta-modules, which are small neural networks that can be quickly adapted to new tasks. During the IL process, the MetaMG approach uses the learned meta-modules to generate task-specific modules for new classes.

Few-shot incremental learning. Similar to FSCIL, Ayub and Wagner (2020a) examines the problem of few-shot incremental learning (FSIL) and proposes a cognitively-inspired approach. They represent each image class as a centroid. In the experimental setting of FSIL, the number of classes for both base and incremental is the same, which differs from the rich base data setting in FSCIL. Additionally, in order to tackle the issue of the inability to learn from data streams in ZSL, Wei, Deng, Yang, and Tao (2021), Wei et al. (2020) have proposed the concept of incremental zero-shot learning (IZSL). Unlike traditional ZSL, IZSL involves multiple learning phases for new classes.

Incremental few-shot object detection. In the setting of incremental few-shot object detection (iFSD) Perez-Rua, Zhu, Hospedales, and Xiang (2020), abundant base-class samples and a few novel-class samples are available. The model can use all the base-class samples for bootstrapping as prior knowledge is required for the model to learn in the few-shot way. Equipped with the prior knowledge of base-class data, the model cannot visit base-class samples again when learning knowledge of novel classes. In other words, the model with prior knowledge should be able to learn from the few samples of unseen categories without relearning basic knowledge, which is aligned with the practical application scenes where the pre-trained model should be competent to adapt to unseen information incrementally.

Despite many studies sharing similar settings to FSCIL, the current mainstream in academia still focuses on FSCIL. Therefore, this review primarily focuses on the more challenging FSCIL research.

3. Methods for few-shot learning

For FSL tasks, specialized network architectures or tricks are typically required to handle limited annotated data. In FSCIL research, many methods build upon advancements in FSL. In this section, we focus on providing a brief overview of commonly used network architectures in FSL, without discussing the novelty or effectiveness of the methods. And they might not represent the latest research.

Numerous surveys have been conducted on the topic of FSL, proposing various classification approaches (Jadon, 2020; Song, Wang, Cai, Mondal, & Sahoo, 2023; Wang, Yao, et al., 2020). One straightforward approach is to categorize FSL into four categories: data augmentation methods, metric-based methods, model-based methods, and optimization-based methods (Jadon, 2020). Hereafter, we will provide a brief introduction to the commonly used network architectures within these four categories.

3.1. Data augmentation methods

In FSL, data augmentation is an important strategy. It alleviates the problem of data scarcity by increasing the diversity of existing data, rather than collecting new data. Data augmentation significantly reduces the risk of overfitting and effectively enhances the model's generalization ability. Data augmentation can be categorized by its source: transforming samples from the training set, transforming samples from a weakly labeled or unlabeled data set, or transforming samples from similar data sets (Wang, Yao, et al., 2020). Besides directly augmenting the data, one can also train a model to generate new samples or features (Kong, Kim, Han, & Kwak, 2022), such as VAEs or GANs, to achieve the goal of data augmentation.
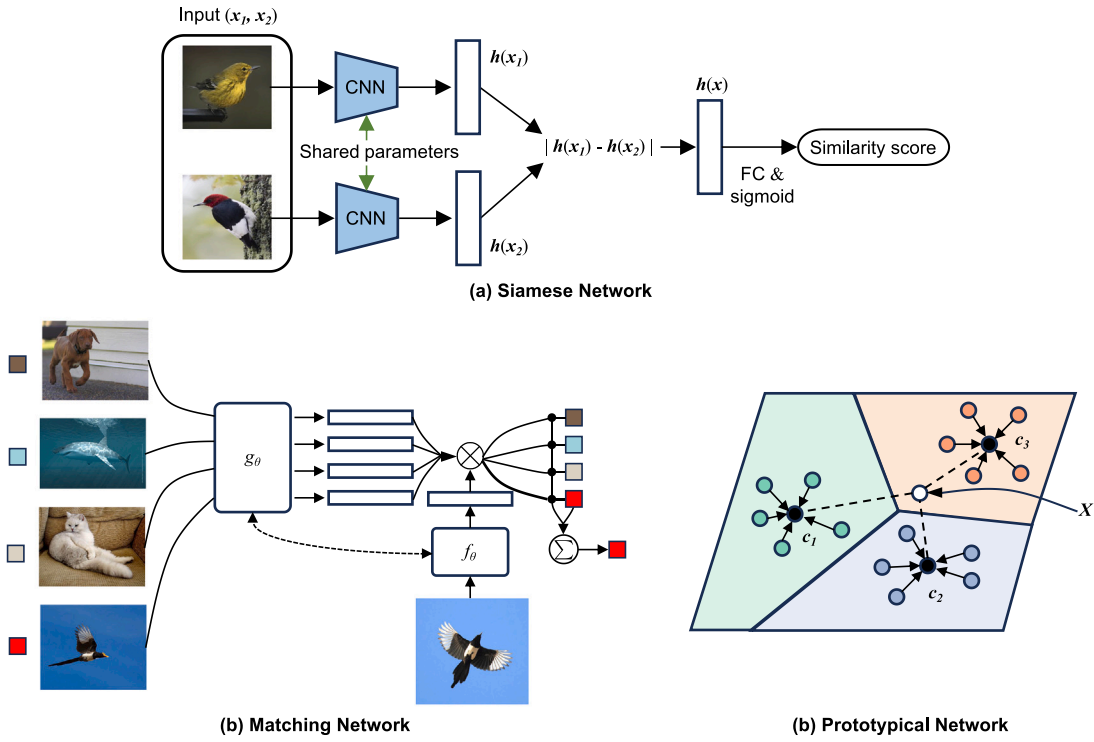
Input (x_1, x_2)

$h(x_1)$

Shared parameters

CNN

$h(x_2)$

$|h(x_1) - h(x_2)|$

$h(x)$

FC & sigmoid

Similarity score

(a) Siamese Network

g_θ

f_θ

(b) Matching Network

c_3

X

c_1

c_2

(b) Prototypical Network

Fig. 3. Common network architectures in metric-based methods: **(a) Siamese Networks**: Utilize twin subnetworks to extract features from two input samples and compute the distance between these features; **(b) Matching Networks**: By using the attention mechanism to dynamically match and aggregate the support set and query set examples, Matching Networks can generate class-related feature representations for query samples; **(c) Prototypical Networks**: Represent each class by the mean of their features. Thus, in the embedding space, closer features are more likely to belong to the same class.

3.2. Metric-based methods

Methods based on metrics classify objects in the embedded space by computing the similarity or distance between samples in the support set and the query set. For instance, by calculating the Euclidean distance between a test sample and each class in the support set, the test sample is assigned to the category of the nearest support set sample. In FSL, commonly used metric learning methods include Siamese Network (Koch et al., 2015), Matching Network (Vinyals et al., 2016), and Prototypical Network (Snell et al., 2017). Fig. 3 illustrates the network structure differences among these three methods. These methods do not require extensive data but optimize metrics to ensure similar samples are close, while samples of different classes are distant.

3.3. Model-based methods

Model-based methods primarily refer to designing or using specific network architectures to address FSL challenges. For instance, Memory-Augmented Neural Networks (MANN) (Santoro, Bartunov, Botvinick, Wierstra, & Lillicrap, 2016) use external memory spaces to explicitly store class information, thus leveraging the long-term memory capabilities inherent in neural networks for FSL tasks. Meta Networks (Munkhdalai & Yu, 2017) learn meta-level knowledge across tasks and adjust their inductive biases through quick parameterization for swift generalization. These network structures efficiently utilize a limited number of labeled samples for rapid learning and adaptation.

3.4. Optimization-based methods

Optimization-based methods focus on adjusting the training strategy of models to adapt to situations with limited annotated data. It typically involves modifying the loss function, regularization terms, or the optimization algorithm itself to ensure that the models can quickly converge on few-shot data without overfitting. For example,

Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) is a common optimization technique that quickly learns knowledge from limited new data. It trains the model's initial parameters using various datasets to ensure peak performance when tackling new tasks. Building on MAML, Reptile (Nichol & Schulman, 2018) simplifies computational complexity by reducing gradient calculations from two steps to one, thereby increasing computational speed.

4. Few-shot class-incremental learning: taxonomy

For fundamental research on FSCIL, there is currently no unified classification standard. Zou, Zhang, Li, and Li (2022) divided FSCIL into metric-based and fine-tuning-based methods. The metric-based method is similar to the concept of FSL (Snell et al., 2017; Vinyals et al., 2016), and its key issue lies in the prototype representation and similarity metric. In FSCIL, the fine-tuning-based approaches are widely used, and we refer to this method as Base Classes Pretraining and Novel Classes Fine-tuning (BPNF).

Definition 2. Base Classes Pretraining and Novel Classes Fine-tuning (BPNF) *is a common approach used in FSCIL, which involves pre-training a model on data-rich base data and fine-tuning the model to better fit the novel classes in the incremental phase. This approach leverages the knowledge learned from the base classes to improve the model's performance on novel, unseen classes.*

However, the above classification method is too broad and not suitable for many FSCIL studies. In this paper, we have summarized 33 advanced studies and categorized them into five families based on the key point or technique used in FSCIL:

- Traditional machine learning methods
- Meta learning-based methods
- Feature and feature space-based methods

Fig. 4. Chronological overview of key FSCIL research developments.

- Replay-based methods
- Dynamic network structure-based methods

Fig. 4 displays an approach classification chart for corresponding years chronologically. It is worth noting that although the experimental settings in FSCIL often involve the idea of meta-learning, these methods are not classified as meta learning-based methods because the key points of the methods used are not based on meta-learning techniques.

4.1. Traditional machine learning methods

4.1.1. Supervised learning strategies

The capacity of a model that has undergone fine-tuning through an incremental process is limited by the amount of new class sample data available. To alleviate this constraint, certain studies have introduced additional semi-supervised or unsupervised data, in addition to relying solely on labeled supervised data, to refine the supervision method.

In Cui, Xiong, Tavakolian, and Liu (2021), semi-supervised learning was introduced to FSCIL and, based on the setting in Tao et al. (2020), 50 unlabeled data were introduced in each incremental session. During the training process, the unlabeled data were combined with labeled data to enhance the performance of FSCIL. In Ahmad, Dhamija, Cruz, et al. (2022), leveraging self-supervised learning was proposed to alleviate overfitting and catastrophic forgetting. Specifically, in addition to training the ResNet-18 model with base-class data, a deeper ResNet-50 network was trained using self-supervised methods on a large dataset. These two networks were then frozen to possess two powerful feature extractors. Two sets of feature vectors were input into a Gaussian Generator to learn models for new classes while passing their features. Subsequently, through feature fusion plus classifier, the forgetting can be effectively countered, and adaptation to the emergence of new classes can be achieved. For the first time, Kalla and Biswas (2022) proposed the self-supervised stochastic classifier (S3C) to solve FS-CIL. The stochasticity of the classifier avoids overfitting to few-shot novel classes, while combining self-supervised training enables better preservation of base-class knowledge.

4.1.2. Statistical distribution

From the statistical distribution perspective, solving the FSCIL problem involves fitting models to existing datasets and predicting the data distribution of the classes, which has excellent model interpretability. To address the limitations of common Gaussian process classification in large-scale class classification tasks, Achituve, Navon, Yemini, Chechik, and Fetaya (2021) proposed GP-Tree. GP-Tree is a tree-based hierarchical model that uses Polya-Gamma data augmentation to fit data to a Gaussian process, which can adapt well to the number of classes and data size. Liu, Yang, et al. (2022) proposed the learnable distribution calibration (LDC) approach, which is rooted in a parameterized calibration unit (PCU). PCU initializes the feature distribution of each class by using a Gaussian sampler defined by the mean vector and stored

covariance matrix to generate a set of feature samples. Specifically, the Gaussian sampler generates enough feature samples during IL to form biased distributions for old and new classes. The PCU cyclically updates the generated feature samples, thereby restoring the old class distribution and calibrating the new class distribution. Due to the fixed size of the covariance matrix, this method has low memory consumption. Both methods achieve good results in FSCIL, but the drawback is that the modeling process is complex.

4.1.3. Function optimization

Existing methods focus on overcoming catastrophic forgetting when learning new tasks, while Shi, Chen, Zhang, Zhan, and Wu (2021) have analyzed this issue from the perspective of function optimization and found that flat local minima obtained during training on base classes have better generalization ability than sharp minima. Flat minima is a crucial concept in machine learning and optimization theory (Hochreiter & Schmidhuber, 1997). In the vicinity of flat minima, minor parameter alterations do not significantly impact the loss function, leading to models with robustness. Furthermore, flat minima serve as a natural form of regularization, typically preventing models from overfitting and enhancing their generalization capabilities. Specifically, Shi et al. (2021) suggest searching for flat local minima of the base training objective function and then fine-tune the model parameters within the flat region on new tasks, substantially reducing catastrophic forgetting.

4.2. Meta learning-based methods

In the realm of FSL or IL, meta-learning can leverage existing knowledge to address current learning problems, and improve the stability and reliability of the system through continuous knowledge accumulation. In FSL, meta-learning enhances the learning effect of the current task by utilizing data from other related tasks (Finn et al., 2017; Liu, Song, & Qin, 2020; Rusu et al., 2019; Snell et al., 2017). In IL, meta-learning can be used to reduce dependence on new data, thereby avoiding overfitting (Riemer et al., 2019). It is natural to apply meta-learning to FSCIL.

Here, we divide the meta learning-based FSCIL method into two categories: prototype learning-based, and meta process-based method.

4.2.1. Prototype learning

Prototype learning aims to identify a small set of exemplars that accurately represent a given dataset, and then use the similarity between the data points and the prototypes to classify new data points or complete other visual tasks. Commonly used class prototypes are defined as follows:

$$\boldsymbol{\mu}_c = \frac{1}{|S_c|} \sum_{\boldsymbol{x} \in S_c} f_\theta(\boldsymbol{x}), \tag{4}$$

where S_c is the set of all samples from class c; f_θ is the embedding network parameterized by θ. Compared to traditional supervised learning

methods, prototype learning requires less labeled data and has stronger generalization ability.

However, simply aggregating all learned class prototypes using traditional prototype-based methods may render some prototypes indistinguishable from one another. To address this problem, Zheng and Zhang (2021) introduced the class structure regularizer to regulate the distribution of the learned classes in the embedding space of FSCIL. By using class distribution as prior knowledge to regularize the learning of new classes, this approach ensures that classes from the same or different sessions are distinguishable from one another.

In FSL, prototype-based methods face challenges in IL scenarios, primarily due to two issues: (i) With the increase in data volume, sample features or label distributions change because of potential concept drift or data distribution drift, making prototype samples fail to accurately represent the latest data distribution; (ii) Newly introduced later-task classes might differ conceptually from earlier classes, causing conflicts within the prototype space, thereby affecting the efficacy of prototype distance measurement and consequently influencing classification accuracy. To address these issues, Zhu, Cao, Zhai, Cheng, and Zha (2021) proposes an incremental prototype learning scheme consisting of random episode selection and dynamic relation projection. Random episode selection improves the extensibility of the feature representation by adapting gradients to different simulated incremental processes generated randomly. Dynamic relation projection utilizes the relationship matrix between new class samples and old class prototypes to update existing prototypes.

Learning Vector Quantization (LVQ) is a prototype clustering method that selects vector points as prototypes based on distance as the clustering criterion. Chen and Lee (2021) uses a non-parametric method based on LVQ in deep embedding space. They compress the information of the learning task into a few quantized reference vectors. These include within-class variation, less forgetting regularization, and calibrated reference vectors to alleviate catastrophic forgetting. Based on the idea of the CIL algorithm, Mazumder, Singh, and Rai (2021) proposes few-shot lifelong learning (FSLL). This algorithm selects some parameters to update in each incremental session to resist overfitting. At the same time, it minimizes the cosine similarity between the new class prototypes and old class prototypes to maximize their separation, thereby improving classification performance.

According to Hersche et al. (2022), the input images are mapped to quasi-orthogonal prototypes from the perspective of hyperdimensional computing. The proposed C-FSIL comprises a frozen meta-learned feature extractor, a trainable fixed-size fully connected layer, and a rewritable dynamically growing memory. The three parameter update forms provided effectively balance accuracy and compute-memory cost. In Yao, Zhu, Zhou, and Li (2022), a human cognition-inspired prototype representation enhancement scheme is proposed for FSCIL. This method uses prototype representations and iteratively learns the knowledge of novel classes by exploring similarity correlations with previously learned classes. Yang et al. (2023) argue that misalignment between the feature and classifier of old classes caused by fine-tuning the backbone or previous classifier prototypes is the reason for forgetting. Inspired by the neural collapse theory, they align a set of prototypes during neural collapse with prototypes required for FSL, which improves the classifier's performance.

The aforementioned methods exhibit conciseness in their algorithms, but the semantic gap between the few-shot class prototypes and the real data distribution is a major obstacle to improving the accuracy of prototype-based methods.

4.2.2. Meta process

Inspired by the multi-task optimization method MAXL (Liu, Davison, & Johns, 2019), Chi et al. (2022) proposed MetaFSCIL, which directly transforms adapting to new knowledge and retaining old knowledge into a meta-objective. They mimicked the scenario during meta-testing

by sampling a sequence of incremental tasks from base classes. Furthermore, they proposed a bi-directional guided modulation based on meta-learning to automatically adapt to new knowledge. Drawing on metric learning within the context of meta-learning, Zou et al. (2022) discovered that using large margin classification improves the performance of the base classes but leads to a decrease in performance when learning new classes, a phenomenon termed class-level overfitting. The authors explain that this is due to the easily satisfied constraint of learning shared or class-specific patterns. Subsequently, they propose the boundary-based CLOM framework, which introduces an additional constraint that effectively addresses the aforementioned issue.

4.3. Feature and feature space-based methods

4.3.1. Feature decoupling

Feature decoupling, which entails dividing features into distinct representations, allows models to concentrate on more pertinent information. According to Zhao et al. (2021), the disentanglement of features results in low-frequency components playing a more significant role in preserving old knowledge. Specifically, they employed discrete cosine transform to disentangle features and proposed a frequency-aware regularization method to enhance inter-space learning performance. Moreover, the proposed feature space composition operation further improves the inter-space learning performance.

4.3.2. Feature space

The representation of subspaces increases the efficiency of algorithms by mapping the original data to a low-dimensional space while preserving its useful features. Based on subspace representation, FSCIL projects new-class data into the subspace composed of base or old-class features, thereby enabling the model to better adapt to new classes. In Cheraghian, Rahman, Ramasinghe, et al. (2021), a mixture of subspaces is proposed to describe the visual and semantic domain distribution of the data, which helps to avoid forgetting old classes. Additionally, a variational autoencoder is utilized to generate synthesized visual samples that enhance the performance of pseudo-features and prevent overfitting during IL of new classes. In Akyürek, Akyürek, Wijaya, and Andreas (2022), the authors propose a subspace regularization scheme that encourages the weight representation of new classes to be close to the subspace spanned by the weights of existing old classes. This regularization term is straightforward and user-friendly, and can incorporate more prior knowledge. From the perspective of parameter feature space, Kim, Han, Seo, and Moon (2023) proposed WaRP by fusing the advantages of F2M (Shi et al., 2021) for finding flat minimums of the loss function and FSLL (Mazumder et al., 2021) for parameter fine-tuning. They seek directions in the parameter space that are flat with respect to the loss function, and use the method of singular value decomposition to represent the parameter space. In each incremental session, they fine-tune unimportant parameters in the parameter space to learn novel classes.

Recently, Song, Zhao, et al. (2023) presents the concept of fantasy space to enhance semantic knowledge. The core idea is to introduce placeholders for unseen classes within the fantasy space. These placeholders derive from the original classes using discrete transformation. By learning to recognize and contrast in the fantasy space fostered by virtual classes, it boosts base classes separation and novel classes generalization.

4.3.3. Prospective learning

Backward compatibility is an issue that requires special consideration in the process of software updates. It demands that newer versions of software be able to accept data from previous versions. Conversely, forward compatibility requires that older versions of software be able to accept data from newer versions. From this perspective, the ability of the FSCIL model to overcome forgetting represents its backward compatibility (Zhou, Wang, et al., 2022). This means that a model trained

on a new session should not forget old class samples. Few studies have addressed the model's forward compatibility, which involves preparing for possible novel classes and updates during current training sessions. Here, we define:

Definition 3. **Prospective Learning** *refers to a certain method or technique in FSCIL, where the model is trained on base dataset to have forward compatibility performance, thus enabling the model to better handle incremental few shot novel classes.*

In order to enable the model to handle new classes, Zhou, Wang, et al. (2022) proposed forward compatible training (FACT), which allocates multiple virtual prototypes as a reserved space in the feature space to make the model scalable. FACT optimizes virtual prototypes to minimize intra-class distances and reserves more space for upcoming new classes. The model is made prospective through instance mixing to generate virtual instances. In subsequent research, Zhou, Ye, et al. (2022) proposed LIMIT, which creates fake FSCIL tasks from the base dataset and obtains generalizable features through meta-learning from different fake tasks to prepare the model for real FSCIL tasks. Additionally, an instance-specific embedding is generated by a transformer-based meta-calibration module to further improve performance. From the perspective of open-set recognition, Peng, Zhao, Wang, Li, and Lovell (2022) linked FSCIL with open-set tasks to prepare the model for new classes. Specifically, they proposed using angular penalty loss in face recognition to obtain good clustering features instead of cross-entropy loss. They combined class enhancement and data augmentation to improve the feature extractor's generalization ability for future incremental classes.

4.4. Replay-based methods

Based on the rehearsal technique, FSCIL approaches replay previously learned information for the task solver when presented with a new task. Replay-based methods employ episodic memory $\mathcal{M}$ to replay the examples from previous tasks while updating the model with the current task t. There are two types: direct replay involves saving examples from old tasks to $\mathcal{M}$, while generative replay involves using a generative model to remember the distribution of data from old tasks and generate examples to $\mathcal{M}$. When fine-tuning the model with data D^t, the loss function can be expressed as:

$$\mathcal{L} = \frac{1}{|D^{(t)} \cup \mathcal{M}^{(t)}|} \sum_{(x,y) \in (D^{(t)} \cup \mathcal{M}^{(t)})} \ell(f(x), y), \tag{5}$$

4.4.1. Direct replay

Kukleva, Kuehne, and Schiele (2021) proposed a three-stage framework, wherein the first two stages train the network on base and novel classes separately and employ a model parameter constraint method to prevent forgetting of old classes. In the third stage, a small set of stored samples are used for replay and calibration of the classifier's performance across all classes (both base and novel classes). IL methods based on knowledge distillation usually store a set of old class exemplars and add additional distillation loss to transfer and preserve old knowledge (Castro, Marín-Jiménez, Guil, Schmid, & Alahari, 2018; Hou, Pan, Loy, Wang, & Lin, 2018; Rebuffi, Kolesnikov, Sperl, & Lampert, 2017; Wu et al., 2019). However, due to class imbalance in few-shot scenarios and performance trade-offs between novel and base classes (Hou et al., 2019), knowledge distillation is not the preferred method for FSCIL. Cheraghian, Rahman, Fang, et al. (2021) proposed the semantic-aware knowledge distillation method by storing a small number of samples for the previous classes. By incorporating word embeddings as auxiliary information and mapping images to vector space, the effectiveness of knowledge distillation for FSCIL has been demonstrated. Unlike CIL based on individual knowledge distillation (Park,

Kim, Lu, & Cho, 2019), Dong et al. (2021) applied graph distillation techniques to FSCIL for the first time. They proposed a scheme for exemplar relation distillation incremental learning (ERDIL) based on graph relation knowledge distillation for knowledge extraction and representation. It effectively transfers old knowledge to the model for learning new tasks by maintaining a graph that represents the relationship between classes.

4.4.2. Generative replay

In light of the privacy issues caused by storing real old data, Liu, Gu, et al. (2022) proposes a data-free replay scheme for synthesizing old samples. By imposing entropy regularization, the generator is encouraged to produce uncertain examples that are closer to the decision boundary. Since the traditional generative replay paradigm in CIL cannot be applied to FSCIL, Agarwal, Banerjee, Cuzzolin, and Chaudhuri (2022) proposes few-shot incremental learning GAN (FSIL-GAN), which consists of a pre-trained feature extractor, a generator, a discriminator, and a semantic projection module. This is used to address the problem of approximating the real data distribution with a small amount of data. They first match class-specific synthesized visual features with their respective latent semantic vectors, and then ensure the diversity and distinguishability of the synthetic features through an anti mode-collapse regularizer. However, this method's performance cannot be guaranteed for multi-domain data.

4.5. Dynamic network structure-based methods

Dynamic network structures (Chen et al., 2020; Sabour, Frosst, & Hinton, 2017) enable automatic adjustment of network architecture during runtime, based on input data features, thereby possessing strong generalization capabilities and reduced risks of overfitting. Due to their flexibility and robust scalability, dynamic architectures have been extensively researched for their applications in IL (Aljundi, Chakravarty, & Tuytelaars, 2017; Rosenfeld & Tsotsos, 2018; Rusu et al., 2016). Leveraging these advancements, researchers have recently applied dynamic network structures in the context of FSCIL. Depending on the initial network structure employed, these methods can be categorized into three distinct groups.

4.5.1. Neural gas network

Tao et al. (2020) proposed the TOPIC framework, which utilizes a neural gas (NG) network to learn the topological structure of the feature space formed by different categories for knowledge representation. The stability of the NG's topology is maintained to prevent forgetting of old categories. With the dynamic growth of NG to accommodate new samples, the representation of few-shot new classes is improved. Fig. 5 (left) displays the stabilization and adaptation of TOPIC.

4.5.2. Graph attention network

The Graph attention network can dynamically process different types of graph data and make dynamic decisions based on the importance of nodes and edges learned on the graph. Zhang et al. (2021) have pointed out that decoupling the training process into embedding learning and classifier learning can effectively prevent knowledge forgetting in the backbone. They proposed the Continually Evolved Classifier (CEC), which first trains the backbone with base data to give the network strong feature extraction capabilities. Then, the graph attention model is introduced, and the graph attention network is used in the classifier layer to adapt to the changes of incremental tasks. With the arrival of incremental tasks, the nodes and weights of the Graph model dynamically increase. Fig. 5 (middle) illustrates the continual evolution of classifier.
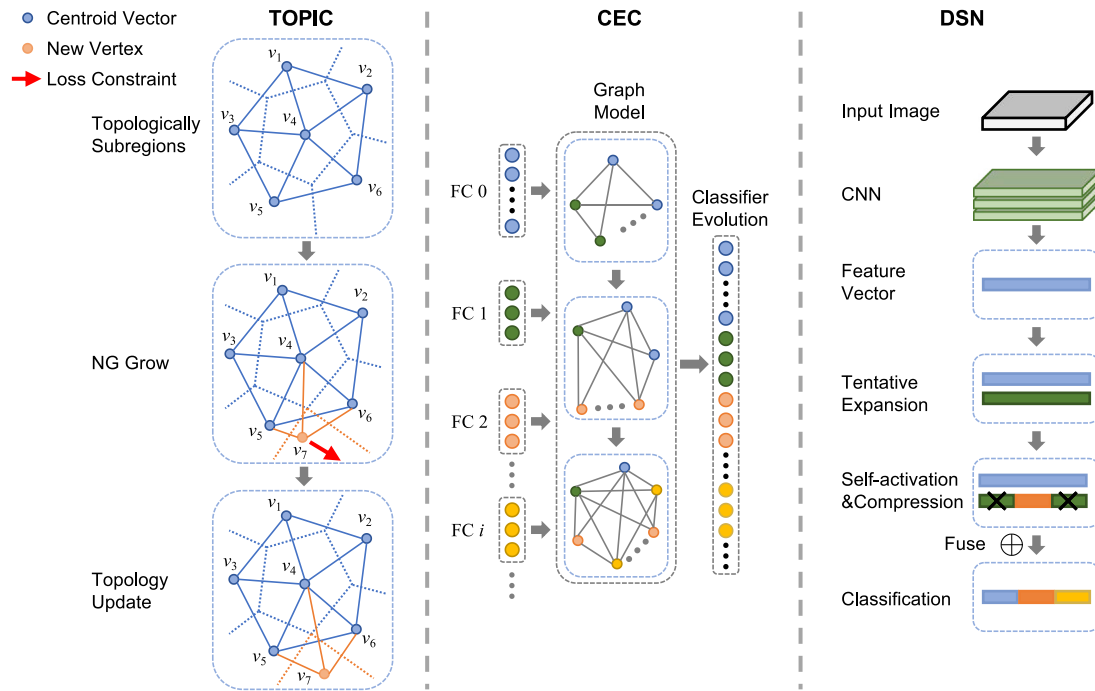
Fig. 5. During training, the network structure dynamically adjusts. **Left**: Sample features form the neural graph's topology. With new nodes added, TOPIC (Tao et al., 2020) uses loss constraints for topology updates. **Middle**: To make the classifier suitable for all categories, CEC (Zhang et al., 2021) applies graph models to the classifier. As new tasks emerge and categories increase, the classifier's topology continuously evolves. **Right**: When training on new classes, DSN (Yang et al., 2022) temporarily expands network nodes to learn new class features, and then compresses redundant nodes to provide a compact feature representation.

4.5.3. Dynamic neural networks

Yang et al. (2021) proposed a learnable expansion-and-compression network (LEC-Net) which enhances the feature representation capability by selectively expanding the network nodes and reduces feature drift from a model regularization perspective. Furthermore, they introduce the dynamic support network (DSN) (Yang et al., 2022) which can adaptively expand the network. DSN leverages compressive network expansion to enrich feature representation in each incremental task and dynamically adjusts the feature space by invoking the old class distribution. During each training, DSN selectively expands the network nodes to enhance the feature representation capability of incremental classes. Then, it dynamically compresses and expands the network through node self-activation to pursue a compact feature representation, thereby alleviating overfitting. Fig. 5 (right) shows the expansion and compression of DSN.

In the latest study, Yoon et al. (2023) explores a masking-based method in network structure. They utilize non-binary masks to construct soft-subnetworks from the original network, effectively balancing forgetting and overfitting. In the base classes session, soft-subnetwork parameters and weight score are learned. In the incremental learning session, minor parameters of the subnetwork are updated.

4.6. Methods summary

This section reviews recent advancements in FSCIL. The following critically examines the strengths and weaknesses of various families.

Traditional machine learning methods offer promising research prospects. By carefully designing the supervised approach of the model, introducing additional data proves effective. Studying FSCIL from a statistical distribution or function optimization perspective enhances model interpretability. However, the complexity of statistical distribution modeling still presents difficulties.

Meta learning-based methods aim to make machine learning models more flexible and adaptive. But meta-learning typically assumes all tasks are from the same or similar data distributions and has high dependence on the meta-training set. When incremental tasks have

different distributions from the base classes, model performance can be affected.

Feature and feature space-based methods leverage the core idea of learning more robust and efficient feature representations. In particular, prospective learning methods are worth exploring for their natural capability in handling unseen samples.

Replay-based methods directly address catastrophic forgetting in FSCIL. However, direct replay faces constraints in storage space, sample selection, and privacy. In contrast, generative replay partially alleviates these issues and offers a more flexible approach. Nevertheless, the challenges of training complexity and subpar data quality persist in generative replay methods.

Dynamic network structure-based methods serve as vital solutions to FSCIL challenges. They adapt to continuously changing data streams by adjusting model structures or inter-class relationships, thereby learning new knowledge while retaining old knowledge. Dynamic Networks have gained traction in IL (Wang, Zhang, et al., 2022; Wang, Zhou, Ye, & Zhan, 2022), and exploring their application in FSCIL is encouraged.

Overall, there remains an open research challenge to develop methods that harmoniously balance performance, scalability, efficiency, and complexity.

5. Model performance

In this section, we will present the performance of typical FS-CIL methods on three different datasets. Firstly, we will outline the methodology for model selection, followed by an introduction to the classical datasets and evaluation metrics. Finally, we will summarize the performance results of various models.

5.1. Model selection

Comparing the performance of different methods is necessary, but currently, many of these methods' codes are not publicly available. As most studies follow the standards set forth by Tao et al. (2020) (see

Table 1
Experimental setup for the three datasets.

Dataset	Base classes	Incremental sessions setup	Sessions
CIFAR-100 (Krizhevsky et al., 2009)	60	5-way, 5-shot	8
MiniImageNet (Vinyals et al., 2016)	60	5-way, 5-shot	8
CUB-200 (Wah et al., 2011)	100	10-way, 5-shot	10

Table 2
AA (%) and PD (%) in CIFAR-100, MiniImageNet and CUB-200.

Families	Methods	Venue	CIFAR-100			MiniImageNet			CUB-200		
			AA↑	PD↓	ResNet	AA↑	PD↓	ResNet	AA↑	PD↓	ResNet
Traditional machine learning methods	SSFSCIL (Cui et al., 2021)	ICIP 2021	–	–	–	–	–	–	50.78	34.64	18
	GP-Tree (Achituve et al., 2021)	ICML 2021	–	–	–	–	–	–	54.26	30.12	18
	F2M (Shi et al., 2021)	NeurIPS 2021	53.65	20.04	18	54.89	22.63	18	69.49	20.81	18
	LDC (Liu, Yang, et al., 2022)	arXiv 2022	–	–	–	–	–	–	68.32	16.31	18
	FeSSSS (Ahmad, Dhamija, Cruz, et al., 2022)	CVPR 2022	–	–	–	68.24	22.63	18	62.86	26.62	18
Meta learning-based methods	FSLL (Mazumder et al., 2021)	AAAI 2021	–	–	–	–	–	–	62.62	19.81	18
	SPPR (Zhu et al., 2021)	CVPR 2021	54.51	20.85	18	52.75	19.53	18	49.32	31.35	18
	CSR (Zheng & Zhang, 2021)	ICDMW 2021	59.07	23.02	20	54.11	23.15	18	62.32	19.60	18
	C-FSCIL (Hersche et al., 2022)	CVPR 2022	61.64	27.00	12	61.61	24.99	12	–	–	–
	MetaFSCIL (Chi et al., 2022)	CVPR 2022	60.79	24.53	20	58.85	22.85	20	61.93	23.26	18
	CLOM (Zou et al., 2022)	NeurIPS 2022	60.57	23.95	20	58.48	25.08	18	67.17	19.99	18
	NC-FSCIL (Yang et al., 2023)	ICLR 2023	67.50	26.41	12	67.82	25.71	12	67.28	21.01	18
Feature and feature space-based methods	VAE-based[a] (Cheraghian, Rahman, Ramasinghe, et al., 2021)	ICCV 2021	50.86	20.36	18	50.63	19.30	18	51.84	25.55	18
	FACT (Zhou, Wang, et al., 2022)	CVPR 2022	–	–	–	–	–	–	64.42	18.96	18
	ALICE (Peng et al., 2022)	ECCV 2022	63.21	24.90	18	63.99	24.90	18	65.75	17.30	18
	LIMIT (Zhou, Ye, et al., 2022)	PAMI 2022	61.84	22.58	20	59.06	23.13	18	65.48	18.48	18
	MgSvF (Zhao et al., 2021)	PAMI 2022	–	–	–	–	–	–	62.37	17.96	18
Replay-based methods	ERDR (Liu, Gu, et al., 2022)	ECCV 2022	60.77	24.26	20	58.02	23.63	18	61.52	23.51	18
Dynamic network structure-based methods	TOPIC (Tao et al., 2020)	CVPR 2020	42.62	34.73	18	39.64	36.89	18	43.92	42.40	18
	CEC (Zhang et al., 2021)	CVPR 2021	59.53	23.93	20	57.75	24.37	18	61.33	23.57	18
	LEC-Net (Yang et al., 2021)	arXiv 2022	43.14	29.37	18	–	–	–	45.09	38.90	18
	DSN (Yang et al., 2022)	PAMI 2022	60.14	23.00	18	54.39	21.06	18	71.02	17.65	18

[a] The method name "VAE-based" is defined by us.

Section 2.1), it is feasible to use the data reported in the original papers of the methods being compared, and we have adhered to this principle. Thus, the results reported in this section are based on the original paper's reported data or the data processed from these original data. We have selected and compared the performance of 22 methods from five different families.

5.2. Datasets

At present, there is no specific dataset for FSCIL, and most of them are made from existing datasets for new tasks. In the majority of FSCIL experiments (Cheraghian, Rahman, Ramasinghe, et al., 2021; Chi et al., 2022; Liu, Gu, et al., 2022; Peng et al., 2022; Shi et al., 2021; Tao et al., 2020; Yang et al., 2022; Zhang et al., 2021; Zheng & Zhang, 2021; Zhou, Ye, et al., 2022; Zhu et al., 2021), the three image classification datasets CIFAR-100 (Krizhevsky et al., 2009), MiniImageNet (Vinyals et al., 2016) and CUB-200 (Wah, Branson, Welinder, Perona, & Belongie, 2011) are commonly used.

CIFAR-100 contains 100 classes with 600 RGB images per class, where each class has 500 training images and 100 testing images. The size of each image is 32×32 pixels.

MiniImageNet contains 60000 RGB images of size 84×84 pixels from ImageNet-1k (Deng et al., 2009). It possesses the same number of classes and samples as CIFAR-100, but its content is more complex and valuable for FSCIL research.

CUB-200 is currently the most widely used benchmark image dataset for fine-grained classification and recognition research. The dataset has a total of 11,788 bird images, including 200 bird subclasses, of which the training dataset has 5,994 images and the test set has 5,794 images. Each image has a size of 224×224 pixels. It provides more sessions and incremental classes for comparing the sensitivity of different methods.

The performance of the selected method was evaluated on the three benchmark datasets mentioned above. For detailed dataset settings refer to Table 1.

5.3. Metrics

Considering the scarcity of original data reported in the paper, we solely compared the accuracy of each session, average accuracy (AA) of all sessions and Performance dropping rate (PD) (Zhang et al., 2021). PD measures the absolute accuracy drops in the last session w.r.t. the accuracy in the base session, defined as

$$PD = \mathcal{A}_0 - \mathcal{A}_N, \qquad (6)$$

where $\mathcal{A}_0$ is the classification accuracy in the base session and $\mathcal{A}_N$ is the accuracy in the last session.

5.4. Results

5.4.1. Benchmark results

Average performance. Table 2 presents the performance of typical FSCIL methods on different datasets. In the comparative experiments, all methods utilized ResNet as the backbone. However, there were variations in the specific ResNet models used (e.g., ResNet-12, ResNet-18, ResNet-20). These differences are detailed in the table. We observe substantial performance disparities among different methods for various datasets. For the small-sized CIFAR-100 dataset, NC-FSCIL (Yang et al., 2023) exhibits outstanding performance at 67.50%, outperforming other methods by a large margin. For the more challenging MiniImageNet dataset, FeSSSS (Ahmad, Dhamija, Cruz, et al., 2022) utilizes self-supervised learning for data augmentation and achieves a performance of 68.24%, surpassing NC-FSCIL (Yang et al., 2023) while also exhibiting lower knowledge forgetting. For the fine-grained CUB-200 dataset, only DSN (Yang et al., 2022) with AA surpasses 70% with a performance of 71.02%, demonstrating a better ability to capture the differences between categories.

Performance comparison by session. The accuracy of each session during the incremental process of various models on the CUB-200 dataset is illustrated in the line chart in Fig. 6. The accuracy of the model on the base classes limits the accuracy improvement during the incremental phase. With the exception of some early methods
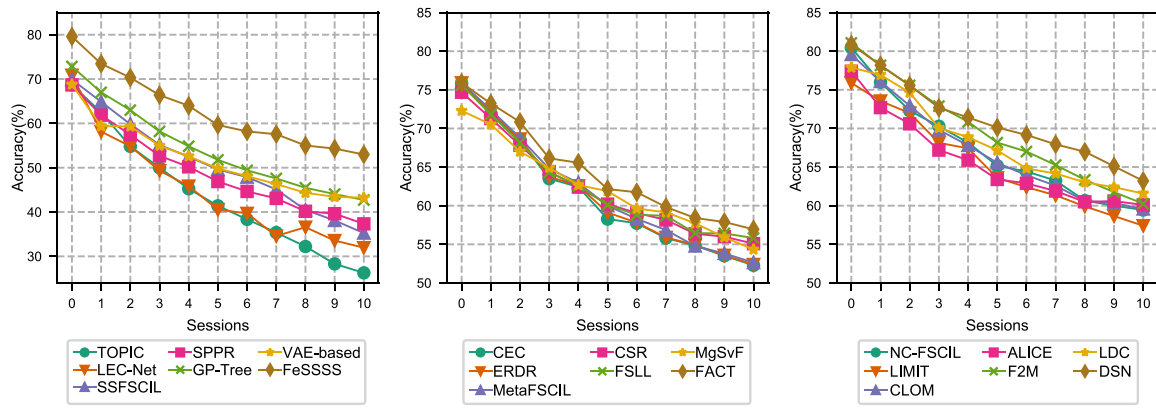
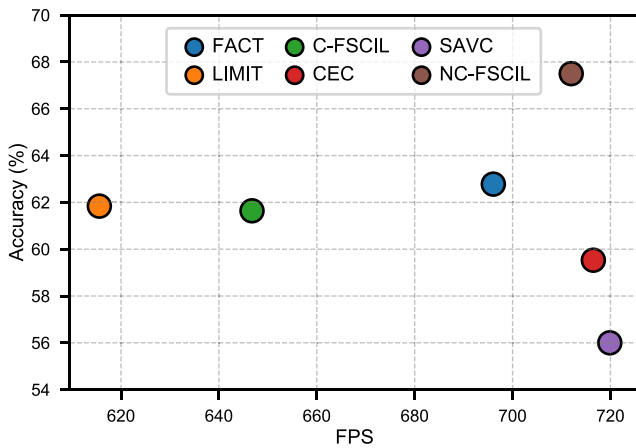**Fig. 6.** Accuracy curves of different methods on each session of CUB-200 dataset.

Fig. 7. Performance comparison of various methods on CIFAR100: FPS vs. Accuracy.

(TOPIC (Tao et al., 2020), SPPR (Zhu et al., 2021), VAE-based (Cheraghian, Rahman, Ramasinghe, et al., 2021)), most methods have an accuracy of 70% to 80% on the base dataset, and few methods have an accuracy above 80% on the base dataset (F2M (Shi et al., 2021), DSN (Yang et al., 2022), NC-FSCIL (Yang et al., 2023)). As the earliest research, TOPIC (Tao et al., 2020) was no longer competitive in each session of the training. F2M (Shi et al., 2021) based on function optimization and DSN (Yang et al., 2022) based on dynamic neural networks still demonstrate high performance advantages.

5.4.2. Performance comparison of accuracy and inference speed

Due to the unavailability of source code for most methods, in this part, we only select methods with publicly available code. We test the accuracy and inference speed of these methods on the CIFAR100 dataset. All experiments are conducted 50 times on an NVIDIA TITAN V GPU with 12 GB of memory, and the average values are reported as the final results. The experimental results are presented in Fig. 7. It is noticeable that NC-FSCIL (Yang et al., 2023), based on the neural collapse theory, leads in both accuracy and inference speed. While SAVC (Song, Zhao, et al., 2023) and CEC (Zhang et al., 2021) methods exhibit lower accuracy, they benefit from reduced model complexity, achieving the fastest inference speeds.

6. Research on few-shot incremental learning applications

In Section 3, the focus lies on fundamental research in FSCIL. In this section, we primarily introduce research that concentrates on implementing FSCIL techniques to resolve practical predicaments. We do not distinguish between the variants of FSCIL, such as FSCIL and FSIL, but rather focus on their applications. FSCIL, originating from computer vision (CV), has presently gained extensive usage in natural language processing (NLP) and Graph technology as well. Further subdivisions can be observed in Table 3.

6.1. Few-shot incremental learning in computer vision

6.1.1. Applications in image classification

To address the increasing demand for classification in hyperspectral imaging, Bai et al. (2020) proposes a linear programming IL classifier. In pedestrian attribute recognition for video surveillance, as the need for identifying new attributes increases, old models become inadequate. Based on the idea of meta-learning, Xiang et al. (2019) uses an attribute prototype generator module and attribute relationship module to generate novel classification weights from annotated data.

The FSCIL method mentioned in Section 3 is mainly used for general classification tasks and neglects the discrimination power of learned representations, making it unsuitable for fine-grained image tasks. Based on the idea of meta-learning, Wang, Liu, et al. (2020) proposes the MetaSearch model to attempt to solve the few-shot incremental product search problem in shopping and checkout processes. MetaSearch extracts different features between various novel categories to perform incremental product search. The designed multipooling-based feature extractor can capture subtle differences between fine-grained product categories, thereby improving classification accuracy. To address the fine-grained vehicle recognition problem, a compact and separable feature learning method (CSFL) is proposed in Li and Huang (2022). CSFL first decouples the feature extractor from the classifier and uses metric learning to train the feature extractor. In the class incremental stage, only the classifier is updated, and incremental LDA is introduced to learn intra-class compact and inter-class separable features, thereby giving the model fine-grained image recognition capabilities. For the even more challenging ultra-fine-grained visual categorization task, Pan et al. (2023) proposes the use of self-supervised learning and knowledge distillation to enhance the feature extraction ability of the network backbone, achieving better performance on fine-grained datasets than the classic FSCIL method.

6.1.2. Applications in object detection

Equipping computer systems with the ability to learn from few examples for object detection has strong practical significance. Inspired by meta learning, Kang (Kang et al., 2019) proposed a novel few-shot detection model. Since the model lacks the ability to incrementally learn new targets from data streams over time, it cannot be extended to real-world deployments in open environments and edge devices. There are also some researchers (Chen, Yu, & Chen, 2019; Liu, Kuang, et al., 2020; Shmelkov, Schmid, & Alahari, 2017) who

Table 3
Summary of FSCIL applied research, including CV, NLP, and graph domains.

Domain	Scenario	Method	Content	Dataset	Metrics
CV	Hyperspectral image	LPILC (Bai et al., 2020)	Propose a linear programming IL classifier	PaviaU, Indian Pines, …	AA
	Pedestrian attribute recognition	APGM (Xiang, Jin, Ding, Han, & Li, 2019)	Incremental few-shot learning for pedestrian attribute recognition	PETA (Deng, Luo, Loy, & Tang, 2014), RAP (Li, Zhang, Chen, Ling, & Huang, 2016)	mA, Precision, Recall, F1
	Fine-grained image	MetaSearch (Wang, Liu, et al., 2020)	Few-shot incremental fine-grained product search	Mini-ImageNet (Vinyals et al., 2016), RPC (Wei, Cui, Yang, Wang, & Liu, 2019)	Accuracy
		CSFL (Li & Huang, 2022)	FSCIL for fine-grained vehicle recognition	Stanford Cars (Krause, Stark, Deng, & Fei-Fei, 2013), CompCars (Yang, Luo, Change Loy, & Tang, 2015)	Accuracy, AA, PD
		SSFE-Net (Pan, Yu, Zhang, & Gao, 2023)	A self-supervised approach is proposed for ultra-fine-grained FSCIL	Cotton (Yu, Zhao, Gao, Yuan, & Xiong, 2021), SoyCultivarLocal (Yu et al., 2021), …	Accuracy
	Object detection	ONCE (Perez-Rua et al., 2020)	The first study on incremental few-shot object detection	Pascal-VOC, COCO	AP, AR
		Meta-Learning-Based* (Cheng, Wang, & Long, 2022)	Introducing meta-learning and redesigning the base framework	Pascal-VOC, COCO	AP, AR
		Sylph (Yin, Perez-Rua, & Liang, 2022)	Designing a class-conditional hypernetwork for incremental few-shot object detection	COCO, LVIS	AP
		Incremental-DETR (Dong, Zhang, Ding, & Lee, 2022)	DETR method based on fine-tuning and self-supervised learning	Pascal-VOC, COCO	AP
		MCH, BPMCH (Feng, Zhang, Yang, & Liu, 2022)	Analogous to the maintenance of new knowledge by establishing new connections in human cells	Pascal-VOC, COCO	AP, AR
	Road object detection	DualFusion (Tambwekar, Agrawal, Majee, & Subramanian, 2021)	Few-shot batch incremental road object detection	IDD (Varma, Subramanian, Namboodiri, Chandraker, & Jawahar, 2019), COCO	AP
	Surface defect detection	DKAN (Sun, Gao, Li, & Gao, 2022)	Knowledge distillation network for incremental few-shot surface defect detection	NEU-DET (Song & Yan, 2013)	AP
	Semantic segmentation	PIFS (Cermelli, Mancini, Xian, Akata, & Caputo, 2021)	Prototype-based incremental few-shot semantic segmentation	Pascal-VOC 2012 (Everingham, Van Gool, Williams, Winn, & Zisserman, 2012), COCO (Lin et al., 2014)	mIoU
		EHNet (Shi et al., 2022)	Adaptive-update and hyper-class representation for incremental few-shot semantic segmentation	PASCAL-5 (Shaban, Bansal, Liu, Essa, & Boots, 2017), COCO	mIoU
	Instance segmenter	iMTFA (Ganea, Boom, & Poppe, 2021)	The first study on incremental few-shot instance segmentation	Pascal-VOC (Everingham, Van Gool, Williams, Winn, & Zisserman, 2010), COCO	AP
		iFS-RCNN (Nguyen & Todorovic, 2022)	Incremental few-shot instance segmentation with Bayesian learning	COCO, LVIS (Gupta, Dollar, & Girshick, 2019)	AP
NLP	Intent recognition	GAL (Zhang, Jiang, Chen, Zheng, & Pan, 2022)	IL structure for few-shot intent recognition	CLINC-150 (Cavalin, Ribeiro, Appel, & Pinhanez, 2020), ATIS (Hemphill, Godfrey, & Doddington, 1990)	Accuracy
	Relation learning	ERDA (Qin & Joty, 2022)	Continual few-shot relation learning	FewRel (Han et al., 2018), TACRED (Zhang, Zhong, Chen, Angeli, & Manning, 2017)	Accuracy
	Named entity recognition	NER* (Wang, Yu, et al., 2022)	FSCIL for named entity recognition	CoNLL03 (Sang & De Meulder, 2003), Ontonote 5.0 (Weischedel et al., 2013)	F1
	Language learning	LFPT5 (Qin & Joty, 2021)	Lifelong few-shot language learning based on prompt tuning	CoNLL03, AGNews (Zhang, Zhao, & LeCun, 2015), CNNDM (Nallapati et al., 2016), …	F1, Accuracy, ROUGE scores
NLP+CV	Label-to-image translation	FILIT (Chen, Zhang, Li, & Sun, 2022)	FSIL for label-to-image translation	ADE20K (Zhou et al., 2017), COCO-Stuff (Caesar, Uijlings, & Ferrari, 2018)	mIoU, accu, …

Table 3 (*continued*).

Domain	Scenario	Method	Content	Dataset	Metrics
Graph	FSCIL in graph	HAG-Meta (Tan, Ding, Guo, & Liu, 2022)	The first GFSCIL research and solved based on prospective learning	Amazon-Clothing (McAuley, Pandey, & Leskovec, 2015), DBLP (Tang et al., 2008), Reddit (Hamilton, Ying, & Leskovec, 2017)	Accuracy, PD, RPD
		Geometer (Lu, Gan, Yang, Zhang, Fu, & Wang, 2022)	GFSCIL based on class prototype representation	Cora-ML (Bojchevski & Günnemann, 2018), Cora-Full (Bojchevski & Günnemann, 2018), Flickr (Zeng, Zhou, Srivastava, Kannan, & Prasanna, 2020), …	Accuracy

[1] * The method name "Meta-Learning-Based" is defined by us.
[2] Below are explanations of the evaluation metrics used in the table:
  AA: Average accuracy.
  PD: Performance dropping rate, see Eq. (6).
  RPD: Relative performance dropping rate, which is the PD normalized by the initial accuracy, $RPD = PD/A_0$.
  mIoU: Mean Intersection-over-Union.
  accu: Pixel accuracy.

study the problem of incremental target detection from the perspective of IL. But none of these methods can cope with the situation where the novel target data is few. Perez-Rua et al. (2020) introduced the incremental few-shot object detection (iFSD) paradigm, where new classes are made available gradually through different sessions. Perez proposed the ONCE model to solve the iFSD problem, which is based on the CenterNet (Zhou, Wang, & Krähenbühl, 2019) one-stage detection method. First, the model uses abundant base dataset to train a class-generic feature extractor. Then, a meta-learning algorithm is used to train a class-specific code generator for each novel category to register new classes. Incrementally appearing new class samples only need to be registered in the meta-training phase through forward propagation without revisiting base classes or iteratively updating, making it suitable for deployment on embedded devices. Most subsequent methods employ class-agnostic feature extractors trained on abundant base data, following the BPNF strategy (see Definition 2), and continuously register new embeddings when novel classes emerge. Cheng et al. (2022) also utilize CenterNet as the fundamental framework, similar to Perez-Rua et al. (2020), but introduce a novel meta-learning method for fine-tuning the model, thus retaining the knowledge related to base classes. During meta-learning optimization, they draw inspiration from the model-agnostic meta-learning algorithm (Finn et al., 2017), a few-shot meta-learning algorithm that uses gradient descent to identify an appropriate initialization that can quickly adapt to the few samples of unseen classes. However, due to overfitting of the feature extractor on base class samples, the model's generalization of output features is inadequate, limiting the proposed model's performance on new classes. Yin et al. (2022) proposed a hypernetwork framework for iFSD called Sylph. It uses a base detector and hypernetwork architecture similar to ONCE. Unlike ONCE, they trained a base detector with class-agnostic localization capability on abundant base dataset, thus decoupling localization from classification. This simplifies the task, but when the size of the base dataset is small or the dataset quality is poor, the class-agnostic detector's localization ability is poor. They further improved the detection accuracy by modifying the network structure and adding normalization to the predicted parameters. Dong et al. (2022) first introduced the DETR object detector (Zhu et al., 2020) into few-shot object detection and proposed Incremental-DETR. They still followed the BPNF guideline. The entire model is divided into two stages. First, the entire network is pre-trained using a large amount of data from base classes, and a self-supervised algorithm is used to fine-tune the class-specific projection layer and classification head. Then, the backbone is frozen, and the class-specific projection layer and classification head are fine-tuned for novel classes. In contrast, Feng et al. (2022) proposed a multi-class head model that mimics the mechanism of maintaining new knowledge by building new connections in human cells. The classification header is continuously added as new data appears. The classification head performs classification detection by using features learned from the data, simulating the way humans learn and maintain new knowledge. Furthermore, by adding a new backbone to the multi-class head model, a bi-path multi-class head model is formed to achieve the transfer from old knowledge to new knowledge.

In practical applications, Tambwekar et al. (2021) proposed a few-shot batch incremental road object detection method specifically designed for road objects. The DualFusion architecture they proposed consists of a Faster R-CNN used for base classes detection, a novel class detection network, and a fusion network. When detecting each new class, only 10 annotated instances are used. The limitation of this method is that although access to the base dataset is only required once, all novel few-shot data must be retained to permanently access novel class data. In the field of hot-rolled steel strip surface defects, Sun et al. (2022) proposes a new knowledge distillation network called dual knowledge align network. Following the BPNF guidelines, a knowledge distillation framework is designed for fine-tuning. They convert NEU-DET (Song & Yan, 2013) into an incremental few-shot dataset, and the experiment shows that they achieve great performance compared to other methods. Furthermore, the few-shot incremental object learning problem for robotic vision is highly valuable. Previous studies have explored the use of a small set of visual examples to incrementally train robots and enhance their recognition capabilities (Ayub & Wagner, 2020b). However, the few-shot incremental object learning problem for robotic vision remains unresolved (Ayub & Wagner, 2021).

6.1.3. Applications in image segmentation

Unlike image classification and object detection, image segmentation requires classification of each pixel, making it more challenging than the other two tasks. Instance segmentation, a subtask of image segmentation, is even more difficult than semantic segmentation as it requires distinguishing boundaries between different instances, while semantic segmentation only requires distinguishing objects and background. In the following, we will discuss some applications of FSCIL in semantic and instance segmentation. Cermelli et al. (2021) proposed the first attempt to solve incremental few-shot semantic segmentation. They proposed PIFS, which combines prototype learning with knowledge distillation. In the base stage, PIFS trains the network on base data to develop the capability of feature extraction. In the FSL stage, PIFS exploits prototypes to initialize classifiers of new classes and fine-tunes the network to refine its feature representation. The subsequently added prototype-based distillation loss enables the model to avoid overfitting and forgetting. Shi et al. (2022) proposed the Embedding adaptive-update and Hyper-class representation Network (EHNet) for incremental few-shot learning. The category embedding describes exclusive semantic properties, and the hyper-class knowledge expresses class-shared semantic properties. The category embedding is stored in

the memory pool and can be updated adaptively. Subsequently, in the segmentation stage, EHNet guides the query image to segment the corresponding category.

For more challenging incremental few-shot instance segmentation, Ganea et al. (2021) introduced Model agnostic methods and proposed the first approach to solving this problem: iMTFA. It repurposes the Mask R-CNN network (He, Gkioxari, Dollár, & Girshick, 2017) to train feature extractors to generate discriminative embeddings for different instances. The average of those class embeddings is used as the representation for each class in the cosine similarity classifier. Thanks to the ability to predict localization and segmentation in a class-agnostic manner, adding new classes simply uses the representation of each class. When a new class appears, Nguyen and Todorovic (2022) fine-tunes the Mask-RCNN that was pre-trained on base classes. Specifically, they use Bayesian learning to estimate the class-weight distribution to modify the classification head and compute the uncertainty of prediction to modify the bounding-box head. This results in better performance than iMTFA on the COCO dataset. However, they do not successfully explain why their estimation of the uncertainty of bounding-box localization surpasses a Gaussian-based uncertainty estimation (He, Zhu, Wang, Savvides, & Zhang, 2019).

6.2. Few-shot incremental learning in natural language processing

FSIL is first proposed in the computer vision field, but with its increasing influence, many studies have applied its ideas to natural language processing (NLP). For instance, in few-shot intent recognition used for text data, Zhang et al. (2022) proposes constructing an undirected fully connected geometry structure based on the spatial distribution of selected samples in the embedding space. Subsequently, they apply a multisource contrastive-based loss to prevent the forgetting of the base classes and avoid overfitting of the novel classes.

Qin and Joty (2022) define relation learning in few-shot and incremental scenarios as continual few-shot relation learning and propose a method based on embedding space regularization and data augmentation to solve this problem. Wang, Yu, et al. (2022) use the generation-replay method to solve FSCIL for named entity recognition, which generates synthetic data of old entity classes for distillation. Qin and Joty (2021) propose a unified framework for lifelong few-shot language learning, LFPT5, based on prompt tuning of T5. LFPT5 performs well on three different tasks: sequence labeling, text classification, and text generation, and is suitable for real-world applications.

In addition, FSIL has also been applied to the fusion field of images and NLP. For example, in the label-to-image translation field, which uses deep learning algorithms to learn the mapping relationship from semantic space to image space. Chen et al. (2022) propose a FSIL method for label-to-image translation, which solves this task with semantically-adaptive filters and normalization.

6.3. Few-shot incremental learning in graph

Recent studies have applied FSCIL to graphs (Lu et al., 2022; Tan et al., 2022). To maintain consistency with existing literature, we refer to this as graph Few-shot class incremental learning (GFSCIL). One of the pioneering studies in this field is the HAG-Meta method proposed by Tan et al. (2022), which incorporates the previously mentioned Prospective Learning concept. HAG-Meta is based on the graph pseudo incremental learning paradigm and enables the model to learn new classes incrementally by cyclically adopting them from the base classes. Furthermore, it addresses class imbalance problems using hierarchical-attention-based modules. Lu et al. (2022) proposed Geometer to tackle GFSCIL problems. Geometer predicts the label of a node by identifying the nearest class prototype in the metric space and adjusts the attention-based prototypes by observing the geometric proximity, uniformity, and separability of novel classes. To mitigate catastrophic forgetting

and unbalanced labeling issues, teacher–student knowledge distillation and biased sampling are also introduced. However, both of these methods are unable to handle dynamic graph structures.

7. Future works

In this section, we discuss three key directions for the further development of FSCIL, namely, (i) theories, (ii) FSCIL settings and (iii) applications.

7.1. Theories

In order to further advance the field of FSCIL, there are several key areas that require attention in future research. Firstly, researchers should aim to enhance the efficiency of the algorithm by considering both performance and complexity. While many studies have solely focused on improving performance, it is important to also take into account the resource requirements of these methods. Secondly, it is crucial to improve testing standards to more accurately evaluate performance across multiple tasks and on the base dataset. Although the average accuracy metric is widely used, it fails to account for the issue of imbalanced base classes and novel classes data. Additionally, the performance dropping rate solely focuses on accuracy of the base and final tasks, without considering the accuracy of intermediate processes. In comparison, relative performance dropping rate (Tan et al., 2022) and harmonic accuracy (Peng et al., 2022) offer more comprehensive means of measuring model performance. Thirdly, as the ViT (Dosovitskiy et al., 2021) continues to gain importance, it may be worthwhile to explore its potential for use in FSCIL, as exemplified in Zhou, Ye, et al. (2022). By addressing these key areas, future research can build upon the current state-of-the-art and continue to advance this important area of machine learning.

7.2. FSCIL settings

The current experimental guidelines for FSCIL largely follow the setting proposed in Tao et al. (2020), which assumes a fixed number of new classes and samples per class in each incremental phase. However, this setting is difficult to meet in real-world applications. To better address this issue, Ahmad, Dhamija, Jafarzadeh, et al. (2022) extended FSCIL to be variable, where in each incremental session, a learning agent can expect up to N ways and up to K shots. Additionally, Kalla and Biswas (2022) proposed a more general setting, where novel classes have different numbers of samples, known as FSCIL-imbalanced, and the number of base classes is not abundant, known as FSCIL-less base. Exploring approaches closer to real-world applications, such as how to handle variable numbers of new classes and shots in different sessions, has practical significance. It is also worth investigating the fusion of FSL with Task-IL and Domain-IL, which are promising research directions.

7.3. Applications

The application of FSCIL in various interdisciplinary fields is a promising avenue for exploration in the future. For instance, recent research has introduced FSIL into the field of audio (Wang, Bryan, Cartwright, Bello, & Salamon, 2021), dynamic few-shot learning for multi-label audio classification (Gidaris & Komodakis, 2018), automatic radar modulation recognition (Luo, Si, & Deng, 2022), intrusion detection (Wang, Lv, Hu, & Sun, 2021), and medical time-series classification (Sun, Zhang, Wang, & Tiwari, 2023). However, these methods are limited to single-scene settings, thus lacking scalability. Therefore, establishing a unified theoretical framework that is applicable to a wide range of scenarios is one of the future directions to address complex and multimodal tasks.

8. Conclusion

Few-shot class-incremental learning is a challenging yet crucial task. It reflects how humans learn in real-world scenarios where high-quality data is often limited and learning data is continually presented. In this paper, we have provided a comprehensive survey of existing FSCIL approaches and attempted to categorize them into five families, including traditional machine learning methods, meta learning-based methods, feature and feature space-based methods, replay-based methods, and dynamic network structure-based methods. Integrating these methodologies to balance performance, scalability, efficiency, and complexity may provide a direction for future research. We have also discussed the performance of classic FSCIL methods and the applications of FSCIL in various fields of deep learning. However, FSCIL remains an underexplored area, and further research is required to explore its potential applications and theories. Due to limitations of space, some theoretical derivations of the content were not extensively introduced. With the increasing demand for real-world AI applications, FSCIL research will continue to attract more attention and drive new innovations in the field of deep learning.

CRediT authorship contribution statement

Songsong Tian: Conceptualization, Methodology, Writing – review & editing. **Lusi Li:** Formal analysis, Visualization, Software. **Weijun Li:** Supervision, Methodology, Writing – review & editing. **Hang Ran:** Methodology, Writing – review & editing. **Xin Ning:** Methodology, Writing – review & editing. **Prayag Tiwari:** Supervision, Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interests.

Data availability

Its survey paper and we will make it open access.

Acknowledgments

References

Achituve, I., Navon, A., Yemini, Y., Chechik, G., & Fetaya, E. (2021). GP-tree: A Gaussian process classifier for few-shot incremental learning. In *Proceedings of the 38th international conference on machine learning* (pp. 54–65). PMLR.

Agarwal, A., Banerjee, B., Cuzzolin, F., & Chaudhuri, S. (2022). Semantics-driven generative replay for few-shot class incremental learning. In *Proceedings of the 30th ACM international conference on multimedia* (pp. 5246–5254).

Ahmad, T., Dhamija, A. R., Cruz, S., Rabinowitz, R., Li, C., Jafarzadeh, M., et al. (2022). Few-shot class incremental learning leveraging self-supervised features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3900–3910).

Ahmad, T., Dhamija, A. R., Jafarzadeh, M., Cruz, S., Rabinowitz, R., Li, C., et al. (2022). Variable few shot class incremental and open world learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops* (pp. 3688–3699).

Akyürek, A. F., Akyürek, E., Wijaya, D. T., & Andreas, J. (2022). Subspace regularizers for few-shot class incremental learning.

Aljundi, R., Chakravarty, P., & Tuytelaars, T. (2017). Expert gate: Lifelong learning with a network of experts. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3366–3375).

Ayub, A., & Wagner, A. R. (2020a). Cognitively-inspired model for incremental learning using a few examples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 222–223).

Ayub, A., & Wagner, A. R. (2020b). Tell me what this is: few-shot incremental object learning by a robot. In *2020 IEEE/RSJ international conference on intelligent robots and systems* (pp. 8344–8350). IEEE.

Ayub, A., & Wagner, A. R. (2021). F-SIOL-310: A robotic dataset and benchmark for few-shot incremental object learning. In *2021 IEEE international conference on robotics and automation* (pp. 13496–13502).

Azadi, S., Fisher, M., Kim, V. G., Wang, Z., Shechtman, E., & Darrell, T. (2018). Multi-content gan for few-shot font style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7564–7573).

Bai, J., Yuan, A., Xiao, Z., Zhou, H., Wang, D., Jiang, H., et al. (2020). Class incremental learning with few-shots based on linear programming for hyperspectral image classification. *IEEE Transactions on Cybernetics*.

Belouadah, E., Popescu, A., & Kanellos, I. (2021). A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Networks, 135*, 38–54.

Bojchevski, A., & Günnemann, S. (2018). Deep Gaussian embedding of graphs: Unsupervised inductive learning via ranking. In *International conference on learning representations*.

Caesar, H., Uijlings, J., & Ferrari, V. (2018). Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1209–1218).

Castro, F. M., Marín-Jiménez, M. J., Guil, N., Schmid, C., & Alahari, K. (2018). End-to-end incremental learning. In *Proceedings of the European conference on computer vision* (pp. 233–248).

Cavalin, P., Ribeiro, V. H. A., Appel, A., & Pinhanez, C. (2020). Improving out-of-scope detection in intent classification by using embeddings of the word graph space of the classes. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 3952–3961).

Cermelli, F., Mancini, M., Xian, Y., Akata, Z., & Caputo, B. (2021). Prototype-based incremental few-shot semantic segmentation. In *British machine vision conference*.

Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., & Liu, Z. (2020). Dynamic convolution: Attention over convolution kernels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11030–11039).

Chen, K., & Lee, C.-G. (2021). Incremental few-shot learning via vector quantization in deep embedded space. In *International conference on learning representations*.

Chen, L., Yu, C., & Chen, L. (2019). A new knowledge distillation for incremental object detection. In *2019 International joint conference on neural networks* (pp. 1–7). IEEE.

Chen, P., Zhang, Y., Li, Z., & Sun, L. (2022). Few-shot incremental learning for label-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3697–3707).

Cheng, M., Wang, H., & Long, Y. (2022). Meta-learning-based incremental few-shot object detection. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(4), 2158–2169.

Cheraghian, A., Rahman, S., Fang, P., Roy, S. K., Petersson, L., & Harandi, M. (2021). Semantic-aware knowledge distillation for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2534–2543).

Cheraghian, A., Rahman, S., Ramasinghe, S., Fang, P., Simon, C., Petersson, L., et al. (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* (pp. 8661–8670).

Chi, Z., Gu, L., Liu, H., Wang, Y., Yu, Y., & 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* (pp. 14166–14175).

Cui, Y., Xiong, W., Tavakolian, M., & Liu, L. (2021). Semi-supervised few-shot class-incremental learning. In *2021 IEEE international conference on image processing* (pp. 1239–1243). IEEE.

De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., et al. (2021). A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(7), 3366–3385.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.

Deng, Y., Luo, P., Loy, C. C., & Tang, X. (2014). Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM international conference on multimedia* (pp. 789–792).

Dong, S., Hong, X., Tao, X., Chang, X., Wei, X., & Gong, Y. (2021). Few-shot class-incremental learning via relation knowledge distillation. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(2), 1255–1263.

Dong, N., Zhang, Y., Ding, M., & Lee, G. H. (2022). Incremental-DETR: incremental few-shot object detection via self-supervised learning.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International conference on learning representations*.

Elsken, T., Staffler, B., Metzen, J. H., & Hutter, F. (2020). Meta-learning of neural architectures for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12365–12375).

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision, 88*(2), 303–339.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2012). The PASCAL visual object classes challenge 2012 (VOC2012) results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

Fe-Fei, L., et al. (2003). A Bayesian approach to unsupervised one-shot learning of object categories. In *Proceedings ninth IEEE international conference on computer vision* (pp. 1134–1141). IEEE.

Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(4), 594–611.

Feng, H., Zhang, L., Yang, X., & Liu, Z. (2022). Incremental few-shot object detection via knowledge transfer. *Pattern Recognition Letters*, *156*, 67–73.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126–1135). PMLR.

Ganea, D. A., Boom, B., & Poppe, R. (2021). Incremental few-shot instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1185–1194).

Gidaris, S., & Komodakis, N. (2018). Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4367–4375).

Gupta, A., Dollar, P., & Girshick, R. (2019). Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5356–5364).

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, *30*.

Han, X., Zhu, H., Yu, P., Wang, Z., Yao, Y., Liu, Z., et al. (2018). FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4803–4809). Brussels, Belgium: Association for Computational Linguistics.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

He, Y., Zhu, C., Wang, J., Savvides, M., & Zhang, X. (2019). Bounding box regression with uncertainty for accurate object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2888–2897).

Hemphill, C. T., Godfrey, J. J., & Doddington, G. R. (1990). The ATIS spoken language systems pilot corpus. In *Speech and natural language: Proceedings of a workshop*.

Hersche, M., Karunaratne, G., Cherubini, G., Benini, L., Sebastian, A., & Rahimi, A. (2022). Constrained few-shot class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9057–9067).

Hochreiter, S., & Schmidhuber, J. (1997). Flat minima. *Neural Computation*, *9*(1), 1–42.

Hospedales, T., Antoniou, A., Micaelli, P., & Storkey, A. (2022). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(9), 5149–5169.

Hou, S., Pan, X., Loy, C. C., Wang, Z., & Lin, D. (2018). Lifelong learning via progressive distillation and retrospection. In *Proceedings of the European conference on computer vision* (pp. 437–452).

Hou, S., Pan, X., Loy, C. C., Wang, Z., & Lin, D. (2019). Learning a unified classifier incrementally via rebalancing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Jadon, S. (2020). An overview of deep learning architectures in few-shot learning domain. arXiv preprint arXiv:2008.06365.

Jamal, M. A., & Qi, G.-J. (2019). Task agnostic meta-learning for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11719–11727).

Kalla, J., & Biswas, S. (2022). S3C: Self-supervised stochastic classifiers for few-shot class-incremental learning. In *Computer vision–ECCV 2022: 17th European conference, Tel Aviv, Israel, October 23–27, 2022, proceedings, Part XXV* (pp. 432–448). Springer.

Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., & Darrell, T. (2019). Few-shot object detection via feature reweighting. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8420–8429).

Kim, D.-Y., Han, D.-J., Seo, J., & Moon, J. (2023). Warping the space: Weight space rotation for class-incremental few-shot learning. In *International conference on learning representations*.

Koch, G., Zemel, R., Salakhutdinov, R., et al. (2015). Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop, vol. 2*. Lille.

Kong, C., Kim, J., Han, D., & Kwak, N. (2022). Few-shot image generation with mixup-based distance learning. In *European conference on computer vision* (pp. 563–580). Springer.

Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3D object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision (ICCV) workshops*.

Krizhevsky, A., Hinton, G., et al. (2009). Learning multiple layers of features from tiny images. *Handbook of Systemic Autoimmune Diseases*, *1*(4).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Kukleva, A., Kuehne, H., & Schiele, B. (2021). Generalized and incremental few-shot learning by explicit learning and calibration without forgetting. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9020–9029).

Li, D.-W., & Huang, H. (2022). Few-shot class-incremental learning via compact and separable features for fine-grained vehicle recognition. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 21418–21429.

Li, D., Zhang, Z., Chen, X., Ling, H., & Huang, K. (2016). A richly annotated dataset for pedestrian attribute recognition. arXiv preprint arXiv:1603.07054.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6-12, 2014, proceedings, Part V 13* (pp. 740–755). Springer.

Liu, S., Davison, A., & Johns, E. (2019). Self-supervised generalisation with meta auxiliary learning. *Advances in Neural Information Processing Systems*, *32*.

Liu, H., Gu, L., Chi, Z., Wang, Y., Yu, Y., Chen, J., et al. (2022). Few-shot class-incremental learning via entropy-regularized data-free replay.

Liu, L., Kuang, Z., Chen, Y., Xue, J.-H., Yang, W., & Zhang, W. (2020). Incdet: In defense of elastic weight consolidation for incremental object detection. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(6), 2306–2319.

Liu, J., Song, L., & Qin, Y. (2020). Prototype rectification for few-shot learning. In *Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16* (pp. 741–756). Springer.

Liu, B., Wang, X., Dixit, M., Kwitt, R., & Vasconcelos, N. (2018). Feature space transfer for data augmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 9090–9098).

Liu, B., Yang, B., Xie, L., Wang, R., Tian, Q., & Ye, Q. (2022). Learnable distribution calibration for few-shot class-incremental learning.

Lu, B., Gan, X., Yang, L., Zhang, W., Fu, L., & Wang, X. (2022). Geometer: Graph few-shot class-incremental learning via prototype representation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1152–1161).

Luo, J., Si, W., & Deng, Z. (2022). New classes inference, few-shot learning and continual learning for radar signal recognition. *IET Radar, Sonar & Navigation*, *16*(10), 1641–1655.

Luo, Z., Zou, Y., Hoffman, J., & Fei-Fei, L. F. (2017). Label efficient learning of transferable representations acrosss domains and tasks. *Advances in Neural Information Processing Systems*, *30*.

Mazumder, P., Singh, P., & Rai, P. (2021). Few-shot lifelong learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(3), 2337–2345.

McAuley, J., Pandey, R., & Leskovec, J. (2015). Inferring networks of substitutable and complementary products. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794).

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: the sequential learning problem. In G. H. Bower (Ed.), *Psychology of learning and motivation, vol. 24* (pp. 109–165). Academic Press.

Munkhdalai, T., & Yu, H. (2017). Meta networks. In *International conference on machine learning* (pp. 2554–2563). PMLR.

Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B., et al. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023.

Nguyen, K., & Todorovic, S. (2022). iFS-RCNN: An incremental few-shot instance segmenter. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7010–7019).

Nichol, A., & Schulman, J. (2018). Reptile: a scalable metalearning algorithm. (p. 4). arXiv preprint arXiv:1803.02999, 2(3).

Palatucci, M., Pomerleau, D., Hinton, G. E., & Mitchell, T. M. (2009). Zero-shot learning with semantic output codes. *Advances in Neural Information Processing Systems*, *22*.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.

Pan, Z., Yu, X., Zhang, M., & Gao, Y. (2023). SSFE-Net: Self-supervised feature enhancement for ultra-fine-grained few-shot class incremental learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 6275–6284).

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, *113*, 54–71.

Park, W., Kim, D., Lu, Y., & Cho, M. (2019). Relational knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3967–3976).

Peng, C., Zhao, K., Wang, T., Li, M., & Lovell, B. C. (2022). Few-shot class-incremental learning from an open-set perspective. In *Computer vision–ECCV 2022: 17th European conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXV* (pp. 382–397). Springer.

Perez-Rua, J.-M., Zhu, X., Hospedales, T. M., & Xiang, T. (2020). Incremental few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13846–13855).

Qi, H., Brown, M., & Lowe, D. G. (2018). Low-shot learning with imprinted weights. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5822–5830).

Qin, C., & Joty, S. (2021). LFPT5: A unified framework for lifelong few-shot language learning based on prompt tuning of t5. arXiv preprint arXiv:2110.07298.

Qin, C., & Joty, S. (2022). Continual few-shot relation learning via embedding space regularization and data augmentation. In *Proceedings of the 60th annual meeting of the association for computational linguistics (Volume 1: Long papers)* (pp. 2776–2789).

Rebuffi, S.-A., Kolesnikov, A., Sperl, G., & Lampert, C. H. (2017). icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2001–2010).

Ren, M., Liao, R., Fetaya, E., & Zemel, R. (2019). Incremental few-shot learning with attention attractor networks. In *Advances in neural information processing systems, vol. 32*. Curran Associates, Inc.

Ren, M., Triantafillou, E., Ravi, S., Snell, J., Swersky, K., Tenenbaum, J. B., et al. (2018). Meta-learning for semi-supervised few-shot classification. arXiv preprint arXiv:1803.00676.

Riemer, M., Cases, I., Ajemian, R., Liu, M., Rish, I., Tu, Y., et al. (2019). Learning to learn without forgetting by maximizing transfer and minimizing interference. In *International conference on learning representations*.

Rosenfeld, A., & Tsotsos, J. K. (2018). Incremental learning through deep adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(3), 651–663.

Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., et al. (2016). Progressive neural networks. arXiv preprint arXiv:1606.04671.

Rusu, A. A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., et al. (2019). Meta-learning with latent embedding optimization. In *International conference on learning representations*.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, *30*.

Sang, E. F., & De Meulder, F. (2003). Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. arXiv preprint cs/0306050.

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., & Lillicrap, T. (2016). Meta-learning with memory-augmented neural networks. In *International conference on machine learning* (pp. 1842–1850). PMLR.

Shaban, A., Bansal, S., Liu, Z., Essa, I., & Boots, B. (2017). One-shot learning for semantic segmentation. In *British machine vision conference 2017*.

Shi, G., Chen, J., Zhang, W., Zhan, L.-M., & Wu, X.-M. (2021). Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima. In *Advances in neural information processing systems, vol. 34* (pp. 6747–6761). Curran Associates, Inc.

Shi, G., Wu, Y., Liu, J., Wan, S., Wang, W., & Lu, T. (2022). Incremental few-shot semantic segmentation via embedding adaptive-update and hyper-class representation. In *Proceedings of the 30th ACM international conference on multimedia* (pp. 5547–5556).

Shmelkov, K., Schmid, C., & Alahari, K. (2017). Incremental learning of object detectors without catastrophic forgetting. In *Proceedings of the IEEE international conference on computer vision* (pp. 3400–3409).

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, *30*.

Song, Y., Wang, T., Cai, P., Mondal, S. K., & Sahoo, J. P. (2023). A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*.

Song, K., & Yan, Y. (2013). A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, *285*, 858–864.

Song, Z., Zhao, Y., Shi, Y., Peng, P., Yuan, L., & 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* (pp. 24183–24192).

Sun, C., Gao, L., Li, X., & Gao, Y. (2022). A new knowledge distillation network for incremental few-shot surface defect detection. arXiv preprint arXiv:2209.00519.

Sun, L., Zhang, M., Wang, B., & Tiwari, P. (2023). Few-shot class-incremental learning for medical time series classification. *IEEE Journal of Biomedical and Health Informatics*.

Tambwekar, A., Agrawal, K., Majee, A., & Subramanian, A. (2021). Few-shot batch incremental road object detection via detector fusion. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3070–3077).

Tan, Z., Ding, K., Guo, R., & Liu, H. (2022). Graph few-shot class-incremental learning. In *Proceedings of the fifteenth ACM international conference on web search and data mining* (pp. 987–996).

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 990–998).

Tao, X., Hong, X., Chang, X., Dong, S., Wei, X., & Gong, Y. (2020). Few-shot class-incremental learning. In *2020 IEEE/CVF conference on computer vision and pattern recognition* (pp. 12180–12189). Seattle, WA, USA: IEEE.

Van de Ven, G. M., & Tolias, A. S. (2019). Three scenarios for continual learning. arXiv preprint arXiv:1904.07734.

Varma, G., Subramanian, A., Namboodiri, A., Chandraker, M., & Jawahar, C. (2019). IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments. In *2019 IEEE winter conference on applications of computer vision* (pp. 1743–1751). IEEE.

Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al. (2016). Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, *29*.

Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2011). *The caltech-ucsd birds-200–2011 dataset*. California Institute of Technology.

Wang, Y., Bryan, N. J., Cartwright, M., Bello, J. P., & Salamon, J. (2021). Few-shot continual learning for audio classification. In *ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing* (pp. 321–325). IEEE.

Wang, Q., Liu, X., Liu, W., Liu, A.-A., Liu, W., & Mei, T. (2020). Metasearch: Incremental product search via deep meta-learning. *IEEE Transactions on Image Processing*, *29*, 7549–7564.

Wang, T., Lv, Q., Hu, B., & Sun, D. (2021). A few-shot class-incremental learning approach for intrusion detection. In *2021 International conference on computer communications and networks* (pp. 1–8). IEEE.

Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, *53*(3), 1–34.

Wang, R., Yu, T., Zhao, H., Kim, S., Mitra, S., Zhang, R., et al. (2022). Few-shot class-incremental learning for named entity recognition. In *Proceedings of the 60th annual meeting of the association for computational linguistics (Volume 1: Long papers)* (pp. 571–582). Dublin, Ireland: Association for Computational Linguistics.

Wang, Z., Zhang, Z., Ebrahimi, S., Sun, R., Zhang, H., Lee, C.-Y., et al. (2022). Dualprompt: Complementary prompting for rehearsal-free continual learning. In *European conference on computer vision* (pp. 631–648). Springer.

Wang, F.-Y., Zhou, D.-W., Ye, H.-J., & Zhan, D.-C. (2022). Foster: Feature boosting and compression for class-incremental learning. In *European conference on computer vision* (pp. 398–414). Springer.

Wei, X.-S., Cui, Q., Yang, L., Wang, P., & Liu, L. (2019). RPC: A large-scale retail product checkout dataset. arXiv preprint arXiv:1901.07249.

Wei, K., Deng, C., Yang, X., & Tao, D. (2021). Incremental zero-shot learning. *IEEE Transactions on Cybernetics*, 1–12.

Wei, K., Deng, C., Yang, X., et al. (2020). Lifelong zero-shot learning. In *IJCAI* (pp. 551–557).

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., et al. (2013). *Ontonotes release 5.0 ldc2013t19, vol. 23*. Philadelphia, PA: Linguistic Data Consortium.

Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y., et al. (2019). Large scale incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 374–382).

Xiang, L., Jin, X., Ding, G., Han, J., & Li, L. (2019). Incremental few-shot learning for pedestrian attribute recognition. In *International joint conference on artificial intelligence* (pp. 3912–3918).

Xie, S., Li, Y., Lin, D., Lay Nwe, T., & Dong, S. (2019). Meta module generation for fast few-shot incremental learning. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*.

Yang, B., Lin, M., Liu, B., Fu, M., Liu, C., Ji, R., et al. (2021). Learnable expansion-and-compression network for few-shot class-incremental learning.

Yang, B., Lin, M., Zhang, Y., Liu, B., Liang, X., Ji, R., et al. (2022). Dynamic support network for few-shot class incremental learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yang, L., Luo, P., Change Loy, C., & Tang, X. (2015). A large-scale car dataset for fine-grained categorization and verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Yang, Y., Yuan, H., Li, X., Lin, Z., Torr, P., & Tao, D. (2023). Neural collapse inspired feature-classifier alignment for few-shot class-incremental learning. In *International conference on learning representations*.

Yao, G., Zhu, J., Zhou, W., & Li, J. (2022). Few-shot class-incremental learning based on representation enhancement. *Journal of Electronic Imaging*, *31*(4), Article 043027.

Ye, H.-J., Hu, H., & Zhan, D.-C. (2021). Learning adaptive classifiers synthesis for generalized few-shot learning. *International Journal of Computer Vision*, *129*(6), 1930–1953.

Yin, L., Perez-Rua, J. M., & Liang, K. J. (2022). Sylph: a hypernetwork framework for incremental few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9035–9045).

Yoon, S. W., Kim, D.-Y., Seo, J., & Moon, J. (2020). XtarNet: Learning to extract task-adaptive representation for incremental few-shot learning. In *Proceedings of the 37th international conference on machine learning* (pp. 10852–10860). PMLR.

Yoon, J., Madjid, S., Hwang, S. J., Yoo, C.-D., et al. (2023). On the soft-subnetwork for few-shot class incremental learning. In *International conference on learning representations*.

Yu, X., Zhao, Y., Gao, Y., Yuan, X., & Xiong, S. (2021). Benchmark platform for ultra-fine-grained visual categorization beyond human performance. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10285–10295).

Zeng, H., Zhou, H., Srivastava, A., Kannan, R., & Prasanna, V. (2020). GraphSAINT: Graph sampling based inductive learning method. In *International conference on learning representations*.

Zhang, X., Jiang, M., Chen, H., Zheng, J., & Pan, Z. (2022). Incorporating geometry knowledge into an incremental learning structure for few-shot intent recognition. *Knowledge-Based Systems*, *251*, Article 109296.

Zhang, C., Song, N., Lin, G., Zheng, Y., Pan, P., & Xu, Y. (2021). Few-shot incremental learning with continually evolved classifiers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12455–12464).

Zhang, X., Sung, F., Qiang, Y., Yang, Y., & Hospedales, T. M. (2018). Deep comparison: Relation columns for few-shot learning. arXiv preprint arXiv:1811.07100.

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in Neural Information Processing Systems*, *28*.

Zhang, Y., Zhong, V., Chen, D., Angeli, G., & Manning, C. D. (2017). Position-aware attention and supervised data improve slot filling. In *Conference on empirical methods in natural language processing*.

Zhao, H., Fu, Y., Kang, M., Tian, Q., Wu, F., & Li, X. (2021). MgSvF: Multi-grained slow vs. fast framework for few-shot class-incremental learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1.

Zheng, G., & Zhang, A. (2021). Few-shot class-incremental learning with meta-learned class structures. In *2021 International conference on data mining workshops* (pp. 421–430).

Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. arXiv preprint arXiv: 1904.07850.

Zhou, D.-W., Wang, Q.-W., Qi, Z.-H., Ye, H.-J., Zhan, D.-C., & Liu, Z. (2023). Deep class-incremental learning: A survey. arXiv preprint arXiv:2302.03648.

Zhou, D.-W., Wang, F.-Y., Ye, H.-J., Ma, L., Pu, S., & Zhan, D.-C. (2022). Forward compatible few-shot class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9046–9056).

Zhou, D.-W., Ye, H.-J., Ma, L., Xie, D., Pu, S., & Zhan, D.-C. (2022). Few-shot class-incremental learning by sampling multi-phase tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–16.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 633–641).

Zhu, K., Cao, Y., Zhai, W., Cheng, J., & Zha, Z.-J. (2021). Self-promoted prototype refinement for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6801–6810).

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2020). Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), 43–76.

Zou, Y., Zhang, S., Li, Y., & Li, R. (2022). Margin-based few-shot class-incremental learning with class-level overfitting mitigation. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in neural information processing systems*.