A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities

YISHENG SONG and TING WANG, East China Normal University, China
PUYU CAI, Michigan State University, United States
SUBROTA K. MONDAL, Macau University of Science and Technology, China
JYOTI PRAKASH SAHOO, Siksha “O” Anusandhan University, India

Few-shot learning (FSL) has emerged as an effective learning method and shows great potential. Despite the recent creative works in tackling FSL tasks, learning valid information rapidly from just a few or even zero samples remains a serious challenge. In this context, we extensively investigated 200+ FSL papers published in top journals and conferences in the past three years, aiming to present a timely and comprehensive overview of the most recent advances in FSL with a fresh perspective and to provide an impartial comparison of the strengths and weaknesses of existing work. To avoid conceptual confusion, we first elaborate and contrast a set of relevant concepts including few-shot learning, transfer learning, and meta-learning. Then, we inductively extract prior knowledge related to few-shot learning in the form of a pyramid, which summarizes and classifies previous work in detail from the perspective of challenges. Furthermore, to enrich this survey, we present in-depth analysis and insightful discussions of recent advances in each subsection. What is more, taking computer vision as an example, we highlight the important application of FSL, covering various research hotspots. Finally, we conclude the survey with unique insights into technology trends and potential future research opportunities to guide FSL follow-up research.

CCS Concepts: • Computing methodologies → Artificial intelligence; Machine learning; Learning paradigms;

Additional Key Words and Phrases: Few-shot learning, one-shot learning, zero-shot learning, low-shot learning, meta-learning, prior knowledge

ACM Reference format:
Yisheng Song, Ting Wang, Puyu Cai, Subrota K. Mondal, and Jyoti Prakash Sahoo. 2023. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. ACM Comput. Surv. 55, 13s, Article 271 (July 2023), 40 pages.
https://doi.org/10.1145/3582688

This work was supported by the National Key R&D Program of China (No. 2021ZD0114600).

Authors’ addresses: Y. Song and T. Wang (corresponding author), MoE Engineering Research Center of Software/Hardware Co-design Technology and Application, Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China, 200062; emails: 71205902054@stu.ecnu.edu.cn, twang@sei.ecnu.edu.cn; P. Cai, Department of Computer Science and Engineering, Michigan State University, Michigan; email: caipuyu@msu.edu; S. K. Mondal, Faculty of Information Technology, Macau University of Science and Technology, Macao, China; email: skmondal@must.edu.mo; J. P. Sahoo, Department of Computer Science & Information Technology, Institute of Technical Education and Research, Siksha “O” Anusandhan University, Odisha, India; email: jpsahoo@ieee.org.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
0360-0300/2023/07-ART271 $15.00
https://doi.org/10.1145/3582688

ACM Computing Surveys, Vol. 55, No. 13s, Article 271. Publication date: July 2023.
1 INTRODUCTION

Recent advances in hardware and information technology have accelerated the interconnection of billions of devices in various IoT-enabled application domains. Smart and adaptive devices are increasingly deployed in critical infrastructures such as health, transportation, industrial production, environmental detection, and home automation. The massive number of terminal devices have been generating a huge amount of data every hour and moment. In the cloud-based service architecture, these data need to be sent back to the data center for central processing and storage. Although the total quantity of generated data at the edge is very large, the volume of every dataset generated by a single device or a single scene is extremely scarce. Traditional data-driven and domain-specific algorithms do not perform well in FSL settings. To this end, numerous research has been conducted in exploring effectual learning paradigms based on few samples or even cross-domain scenarios. Few-shot learning (FSL), as well as meta-learning, have inevitably emerged as a promising method to address such issues. However, how to effectively introduce prior knowledge that can be quickly generalized on new tasks with only a small amount of data remains the greatest challenge.

Empirically, the data distribution in real-world scenarios often has long-tail effects. It is difficult to generalize the same model across different data distributions when there is rarely any data. Taking the industrial inspection for smart manufacturing as an example, poor generalization has become one of the critical challenges that largely affect the performance of their intelligent models, particularly with dynamic changes in lighting conditions and scarce defect samples. In fact, similar problems involve some high cost and privacy scenarios. In the field of cosmic ray discovery, researchers hope to find high-energy cosmic rays that carry a particular message from the thousands of cosmic rays. The entire process is time-consuming, can take more than a decade, and is costly. Nowadays, few-shot learning is active in all walks of personal life, involving credit fraud [1], ticket identification [2], intent recognition [3], cold start recommendation [4], and gesture recognition [5]. Table 1 provides a summary of application scenarios and main challenges of FSL.

To address these issues more effectively, numerous creative works have been proposed. Wang et al. [6] conduct an extensive investigation on how to minimize empirical risk, including data, models, and algorithms. In our survey, from the level of abstraction of prior knowledge, we represent the work of FSL as data level, feature level, task level, and multimodal level. In contrast to the previous discussion on how to minimize empirical risk, our survey focuses on prior knowledge itself and attempts to refine the abstraction levels of prior knowledge from the perspective of challenges. The data level is the lowest level, which mainly aims to increase the data diversity as much as possible by transforming metric functions or directly generating new data. The feature level is the second level, which can be used as the basic statistical knowledge of a domain between the data and the corresponding labels. The task level is the third level, which is independent of specific data and domains and accomplishes the mapping from seen to unseen tasks by customizing the learning paradigm. Finally, the multimodal level is the highest level of the prior knowledge in FSL. It can address a certain type of challenges using multiple data sources including images, text, and audio in a generalized way. The multimodal level is expected to be the ultimate path to end the FSL issues.

Correspondingly, different levels of prior knowledge face multiple challenges. At the data level, the data volume is too scarce, so models can not accurately evaluate the true data distribution by relying on only one or few shot samples. At the feature level, models are trained with large-scale base datasets. If the base dataset and support dataset exist clear domain gap, then the feature level may learn an inconsistent representation space that misguides the parameter updates. At the
Table 1. A Summary of Application Scenarios as Well as Main Challenges of FSL

| Scenarios                        | Industry                | Issues                                      |
|----------------------------------|-------------------------|---------------------------------------------|
| Quality Inspection Line          | Smart Manufacturing     | Scarcity of sample space                    |
| High Energy Cosmic Ray Detection| Air & Space             | Differences in domains distribution         |
| Drug Toxicity Discovery          | Chemical                | Scarcity of sample space                    |
| Credit Fraud                     | Finance                 | Scarcity of sample space                    |
| Ticket Identification            | Optical Character Recognition | Sample distribution in local space          |
| Intent Recognition               | Natural Language Processing | Scarcity of sample space                  |
| Cold Start Recommendation        | Recommendation          | Scarcity of sample space                    |
| Gesture Recognition              | Action Recognition      | Differences in domains distribution         |

Table 2. A List of Key Acronyms

| NOMENCLATURE                     |
|----------------------------------|
| **Full Form**       | **Abbreviation** | **Full Form**       | **Abbreviation** |
| Artificial Intelligence        | AI               | Few-Shot Learning   | FSL             |
| Deep Learning                  | DL               | Machine Learning    | ML              |
| Zero-Shot Learning             | ZSL              | One-Shot Learning   | OSL             |
| Neural Architecture Search     | NAS              | Convolutional Neural Network | CNN         |
| K-NearestNeighbor              | KNN              | Support Vector Machine | SVM          |
| Nearest Centroid Classifier    | NCC              | Graph Neural Network | GNN             |
| Variational Auto Encoders      | VAE              | Few-Shot Object Detection | FSOD        |
| Long Short-Term Memory         | LSTM             | Data Augmentation   | DA              |
| Few-Shot Cross-Domain          | FSCD             | Contrast Learning   | CL              |
| Language Model                 | LM               | Prompt Tuning       | PT              |

So far, several existing surveys have investigated few-shot learning from different perspectives. Shu et al. [7] divide few-shot learning into experience learning and concept learning. Lu et al. [8] classify few-shot learning into generative models and discriminant models. Recently, Wang et al. [6] investigate the minimization of empirical risk and define the few-shot learning in terms of experience $E$, task $T$, and performance $P$. The whole taxonomy is based on optimizing the parameter space. To our best knowledge, no paper has yet provided a taxonomy from the perspective of prior knowledge itself. In line with the levels of abstraction of prior knowledge, our survey tries to summarize each level separately in detail from the perspective of challenges. Logically, data, model, and algorithm are three parallel aspects that play different roles in parameter updating. Instead, the level of prior knowledge itself is hierarchically progressive. By summarizing the challenges of different levels, readers can better grasp the motivation and principle behind the FSL. A list of key acronyms used in this article is summarized in Table 2.
1.1 Organization of the Survey

The remainder of this survey is organized as follows: Section 2 provides an overview of FSL, comparatively analyzing machine learning, meta-learning, and transfer learning along with a summary of current variants of FSL and challenges. In addition, to address the obstacles systematically, in this section, we demonstrate an innovative taxonomy to classify the existing FSL-related works. Section 3 to Section 6 present a systematic investigation of current mainstream research from the perspective of challenges in FSL and provide a comparative analysis from various aspects. With this taxonomy, a discussion and summary are provided at the end of each section. Section 7 takes computer vision as an example, counting the latest progress of FSL in representative tasks. Section 8 explores the current challenges faced by FSL and how to seek breakthroughs in each branch. The overall outline of this article is shown in Figure 1.

The main contributions of this survey can be summarized as follows:

- We start with the edge computing scenario, in which the FSL challenges arise, explaining and clarifying several analogous concepts that are easily confused. This will be beneficial to help readers establish the relationship between few-shot learning, transfer learning, and meta-learning.
- We comprehensively investigate the FSL-related work from the perspective of prior knowledge itself through knowledge graphs. With this taxonomy, we divide the FSL work into four levels according to the abstraction of prior knowledge, where the highest level is multimodal, and the first, second, third are data level, feature level, and task level, respectively.
- We investigate an adequate number of papers in recent three years and summarize the main achievements of FSL involving a wide range of benchmark datasets and tasks. Notably, we also provide a cutting-edge insightful discussion at the end of each section.
- Finally, in response to these challenges in line with the taxonomy, we discuss the current hot directions of FSL development and make recommendations in each potential area. We aim to inspire readers on how to find breakthroughs based on existing work and jointly promote FSL in a more practical direction.
2 CONCEPTS AND PRELIMINARIES

As a branch of machine learning, FSL aims to address new tasks with scarce data settings. What is FSL, and how does it relate to traditional machine learning, transfer learning, and meta-learning? What are the issues of the existing FSL variants? What is the benchmark dataset of FSL in computer vision and natural language processing? In this section, we will resolve the obstacles to FSL for readers by answering these questions.

2.1 What Is Few Shot Learning?

The concept of FSL is inspired by the robust reasoning and analytical capabilities of humans. In 2020, Wang et al. [6] give a standard definition through experience (E), task (T), and performance (P) of machine learning: A computer program is said to learn from experience E with respect to some classes of task T and performance measure P if its performance can improve with E on T measured by P. It is worth mentioning that E in FSL is very scarce. Relevant neural scientific evidence [9] has proved that innate human abilities are related to various memory systems, including parametric slow learning neocortical systems and non-parametric fast hippocampal learning systems, which correspond to FSL’s data-based slow learning and feature-based fast learning, respectively.

Currently, theoretical research on FSL mainly focuses on domain adaption [10] and probably approximately correct (PAC) [11]. In common machine learning tasks, it is assumed that the real data X obeys a distribution D. Normally, we can get sufficient samples x_1, x_2, and x_3 so the model can map the parameter space $H^{all}$ to be as consistent as possible with the target concept $h^*$ by minimizing the approximation error. As for the parameter space $H^{few}$ of FSL, Yang et al. [12] explain the mean and variance of FSL data distribution are inaccurate due to few samples being a very small subset of the total dataset. But similar classes usually have similar means and variances. Figure 2 illustrates the estimation and approximation error for FSL in PAC. In this figure, $h^{meta}$ is the optimal initialization point searched by meta-learning, $h^{few}$ is the optimization goal based on the $H^{few}$ space, $h^{few} - h^*$ is the approximation error, and $h^{few} - h$ is the estimation error. Theoretical evidence [13] proves that the difference between estimation and approximation error is less than a value with some probability. In conclusion, constrained parameters mapping space, uncertainty optimization path, and poor initialization state all lead to poor model performance. FSL tasks are typically represented with the N-way-K-shot. The number of samples per class has a significant impact on the model performance [14]. As the sample size increases, the improvement of model performance will decrease, the sensitivity of estimation error to intra-class variance bias will decrease, and the approximation error to inter-class variance bias will increase.
In line with the taxonomy, FSL still exists many challenges, which are generated from various levels of prior knowledge. In this section, we illustrate in detail the challenges of FSL that are still not well addressed:

- **Inaccurate data distribution assessment**: In FSL, it is difficult to label a large amount of semi-supervised dataset due to costs, ethical, legal, or other reasons. Relying on only a few samples may produce biases in estimating the true data distribution, which is harmful to some tasks. At the data level, maximizing the exploration of data distributions with scarce information is the most significant challenge.

- **Feature reuse sensitivity**: Through accumulating large-scale data-label domain knowledge, transfer learning easily obtains feature-level prior knowledge with the pre-trained model. However, as far as cross-domain FSL is concerned, the pre-trained feature extractor does not have sufficient generalization, which leads to misguidance for unseen tasks.

- **Generality of future tasks**: As represented by meta-learning and episodic training, task-level methods enable dual sampling of data and tasks and map from seen tasks to unseen tasks as quickly as possible. Nevertheless, meta-learning has recently been shown to be effective only when training and testing tasks are similar enough. Moreover, meta-learning is highly dependent on network structure under various domains in the real world.

- **Effectiveness of multimodal information fusion**: Multimodal learning enables interactions with the environment such as image, text, and audio. It has shown strong generality integrating image embedding and text vector. A parametric multimodal pre-trained model can easily address various downstream tasks. However, how to embed texts, images, and other messages into the same space with minimal loss is still the core challenge of multimodal FSL.

### 2.2 How Does Few-shot Learning Relate to Traditional Machine Learning?

Traditional machine learning is a data-hungry method. It uses large-scale datasets as input, and its inference is based on statistical results extracted from historical prior knowledge. Nowadays, the burgeoning of 5G provides massive connectivity for millions of end devices, enabling the interconnection of everything. The total size of data generated by terminal devices is huge, but the single dataset is extremely scarce. Hence, traditional ML does not perform well in FSL data settings. To this end, FSL emerges and provides a promising way to handle data scarcity scenarios.

In recent years, the research on FSL has been extensively conducted and has achieved significant progress. Figure 3 exhibits the statistics of paper publications and citations related to FSL. From 2014 to 2017, the number of publications per year did not exceed 50. During this period, almost all work still focused on pre-training and fine-tuning. Since 2018, with the advent of metric learning and episodic training of meta-learning, extensive work has been completed based on various baselines of metric learning and meta-learning and achieved competitive performance in FSL. Up to now, the 5-way-1-shot task has achieved 95.3% accuracy, and the 5-way-5-shot task has achieved 98.4% in mini-ImageNet few-shot classification tasks. Figure 5 provides a knowledge map covering the hot research topics and cutting-edge developments in the field of FSL, including but not limited to zero-shot learning, one-shot learning, transfer learning, meta-learning, and multitask learning. Note that computer vision with predominant green color is the most active research field.

FSL is just a specialized setting of machine learning, which is difficult to obtain enough valid data in fragmented scenarios. In FSL episodic training, the input of the model is generally given in the tasks. Through continuously collecting related tasks, FSL is enabled to measure the similarity between support-query pairs. When the model faces an unseen task, a good initialization can be quickly accomplished with just a few iterations. Compared to traditional machine learning,
it requires optimization through a specific loss function based on large-scale base datasets. In conclusion, no matter FSL or traditional machine learning, both of them are just tools to produce algorithms that eventually serve to fit their respective application scenarios.

2.3 How Does Few-shot Learning Relate to Transfer Learning?

Transfer learning imitates the thought processing of the human brain; whereby, after addressing a problem, a better, faster solution is generated when a new, related problem arrives. Contrasted to transfer learning, the limited amount of training data, domain variations, and task modifications make FSL more challenging. There are many variants of FSL, including one-shot learning (OSL), zero-shot learning (ZSL), and cross-domain few-shot learning. These variants are considered special cases of transfer learning in terms of sample size and domain learning.
The three-step training involved in meta-learning includes:

1. **Stage 1**: Find the learning algorithm. It can produce specific function.
2. **Stage 2**: Define loss function using tasks.
3. **Stage 3**: Find parameter that can minimize loss function.

### One-Shot Learning
One-shot learning has one available sample per class in the support dataset. The model only needs to answer yes or no based on the support set. In fact, one-shot learning does not classify the data specifically, but simply makes a cluster to learn the similarity metric function.

### Zero-Shot Learning
Zero-shot learning considers a more extreme case extending FSL. In the absence of any support samples, zero-shot learning completely relies on semantic features as a bridge to infer the unseen samples. Zero-shot learning represents the raw data using high-dimensional semantic features instead of low-dimensional pixels feature. One-shot learning and FSL can be regarded as special zero-shot learning.

### Cross-Domain Few-shot Learning
Cross-domain few-shot learning combines the challenges of transfer learning and FSL. Due to the existence of domain gaps, the intersection of classes in the source and target domains is empty. In addition, the number of samples in the target domain is extremely scarce.

### How Does Few-shot Learning Relate to Meta Learning?

Meta-learning is a general paradigm that provides episodic training. Figure 6 illustrates the three-step training involved in meta-learning [17]. It focuses on improving generalization for unseen tasks with the assistance of prior knowledge. If the prior knowledge is used to assist the model in learning on a specific task, then meta-learning can be regarded as a variant of FSL. Meta-learning is not equivalent to FSL. FSL is more of an ultimate goal that aims to achieve robust representations without relying on large-scale datasets. Through dual sampling of data and task space, meta-learning is enabled to construct a large number of auxiliary tasks related to unseen tasks. Even if some work does not involve meta-learning, it is likely to improve performance if episodic training can be considered, such as meta reinforcement learning [18, 19], meta video detection [20], and so on.

Nevertheless, meta-learning has its own limitations: when the training and testing tasks exist obvious domain gap, meta-learning rarely initializes parameter weights. In addition, meta-learning is highly dependent on the structure of the network and needs to be redesigned for widely varying tasks. In spite of this, meta-learning is still one of the most effective methods to address FSL issues.

### Datasets

Before the FSL benchmark dataset was presented, most of experiments use a wide variety of datasets in N-way-K-shot tasks to evaluate the performance of FSL models. However, several simple datasets cannot reflect the complexity of the real world. Through 20 years of development, FSL
benchmark datasets have completed the transition from single-domain, single-dataset to cross-domain, and multi-dataset. Table 3 summarizes these datasets in a variety of dimensions. Taking it a step further, in Figure 4, we choose computer vision tasks to present the state-of-the-art during 2019–2022. If the benchmark dataset involves multiple datasets, then we calculate the average accuracy of report results. Except for the Meta-Dataset [21] and BSCD-FSL [22], all other benchmark datasets were evaluated using 5-way-1-shot. It is clear that mini-ImageNet [23] and CIFAR-FS [24] few-shot image classifications are relatively well addressed. But several cross-domain image classification tasks, especially non-natural scenarios, are still in the exploration stage. The best result achieved by the RDC-FT [25] is only 53.13%, which is clearly below the human level.

In natural language processing, FewGLUE_64_labeled [26] is a collection of natural language tasks in English, including text inference, textual entailment, sentiment analysis, and semantic similarity, whose corpus consists of individual sentences. Recently, FewCLUE [27] has become the official benchmark in the Chinese language, including IFLYTEK (long paper classification), CLUEWSC (pronoun disambiguation), CSL (paper keyword recognition), CSLDCP, and other tasks. The human performance on FewCLUE [27] is an average of 82.49%. However, the model’s best performance [28] is only 54.34%. The results indicate that FSL still has great potential for improvement in natural language processing.

### 2.6 Taxonomy

In line with the modality of prior knowledge, FSL work can be clearly divided into a unimodal learning stage and a multimodal learning stage. In the unimodal learning stage, our survey further abstracts the prior knowledge in data, features, and task levels. Figure 7 vividly demonstrates the FSL’s taxonomy under the abstract of prior knowledge in the form of the pyramid.

- **Data Level**: Data level is mainly used to evaluate the true data distribution by increasing the number of features or samples. The most direct method is to generate additional data
Fig. 7. The entire taxonomy is presented in the form of a pyramid. The bottom level represents the “cloud-edge-terminal” edge computing scenario, which is characterized by few-shot real-time computation under high traffic. Based on this, the challenges of FSL are classified into four levels according to the degree of integration of the prior knowledge.

Based on semantic space [43] or label the similar auxiliary labeled datasets [44]. As for semi-supervised datasets, contrast learning [45, 46] and latent augmentation [45] are effective representations frameworks.

- **Feature Level**: Feature level is mainly used to build a data-to-label mapping from support set to query set. A good feature embedding is essential for extracting discriminative representations. Most pre-training and fine-tuning work need to apply with effective regularisation. In addition, if maximizing the retaining parameters of the pre-trained model, then prompt-tuning largely breaks the constraint of the data by giving some manual prompts to the model.

- **Task Level**: Task-level is mainly used to refine parameters in the task space, including model parameters and meta-learned parameters. Distinguishing multi-task learning, meta-learning tries to learn prior knowledge from related tasks without expanding the learned parameters or sacrificing efficiency at inference. Meta-learning, metric learning, and graph neural networks are the dominant methods during this processing.

- **Multimodal Level**: Multimodal level enables text, visuals, and other messages to be embedded in the same space with minimal loss. With the assistance of language models, images also could be embedded in the form of patches. Multimodal learning contains a wealth of knowledge that leads FSL to enter the area of multimodal large models + small samples.

3 DATA LEVEL: EVALUATE THE TRUE DATA DISTRIBUTION WITH MAXIMUM PROBABILITY

**Few-shot learning (FSL)** aims to learn new categories with a few or even zero samples per class. In real-world FSL tasks, the number of samples used for training is always limited due to privacy, collection, and labeling costs. To address the data scarcity, a direct way is to increase the number of samples available. In our survey, we continue to follow the abstraction of prior knowledge level and subdivide data augmentation into data expansion and feature expansion.

3.1 Data Expansion

Data expansion is mainly adding new available data to original training dataset, either as generated pseudo-labeled data based on unlabelled data [45, 47, 48] or transform on original dataset data.
Fig. 8. Data expansion includes applying transformation rules directly on the training set, learning transformation rules from other semi-supervised datasets, or selecting suitable samples from similar semi-supervised datasets.

Fig. 9. Feature augmentation is mainly performed with the assistance of unlabelled query sets and semi-supervised reference datasets, with discriminative feature representation on the feature space.

[43, 49, 50]. Figure 8 summarizes methods to expand the dataset on FSL data augmentation. When a sufficient dataset is available, FSL is expected to regress to the machine learning paradigm so all machine learning techniques can be used to solve FSL problems.

Among them, the most widely used method is to transform various rules on the image involving flipping, Gaussian blurring, cropping, scaling, and so on [51]. Furthermore, References [52] and [53] combine these simple sub-strategies named auto data augmentation. However, pixel information has limited improvement for FSL. Several works attempt to superimpose semantic information [54], complement areas of erasure regions [55], and transform inter-class or intra-class variation [56]. An early work [57] is using the Hausdorff distance metric in similar classes. Later on, References [58–60] use auto-encoders to capture linear or non-linear changes in class attributes, such as spatial and appearance. As far as natural language processing, data expansion mainly includes simple back-translation, word replacement, paraphrasing, and belief state annotations [50].

No matter image or text, the corresponding labels assign the original labels. Recently, inversion label [61] could also be found the performance will be improved on special tasks.

Compared to generating more diverse samples on the original dataset, a better direction is to introduce a large amount of unlabelled datasets. In gesture recognition, T. Pfister et al. [62] first co-train a similar gesture in a large library of unlabelled gestures. However, in many other cases, the model cannot complete the selection of unlabelled data at once. The whole process requires constant training, inference, and addition. Specifically, Wu et al. [63] selected labeled data as the initial model, Wang et al. [64] trained with both labeled and unlabelled data together, and Reference [65] defines unlabelled data and its nearest labeled neighbors in the same metric space by loss distribution. However, if there exists a domain transfer between unlabelled and labeled data, then the generated pseudo-labels may bring the noise from the target class. An earlier work [66] modeled this relationship as two generators, one for mapping large samples to small samples and another for mapping small samples to large samples. Much work has been used to generate better pseudo labels, including fusing data distributions [67], contrast learning [45], and embedding augmentation [45].

In addition to the unlabelled datasets, the query sets can also be trained together in FSL, which is named transductive few-shot learning [68]. By adding test data in the training stage, more discriminative global representation features can be obtained. STARTUP [46] proposes to train the model with contrast loss functions on unlabelled query datasets, and the clustering of the model results as a label for these unlabelled query datasets. If inter-domain differences are large, then it may be suboptimal to use a fixed pre-trained model. Islam et al. [69] discard the contrast loss function and
use dynamic distillation to learn a better representation of the target domain. Currently, FSL based on transductive learning is tightly integrated with the meta-learning task, and more work will be introduced in later sections.

3.2 Feature Augmentation

Feature augmentation is a higher level of data augmentation and focuses on a series of feature transformations. Figure 9 summarizes the main methods in the feature space. In feature augmentation, samples are passed through the feature extractor and the valid feature is further compressed into texture, position, or attributes. Laso [70] is an early and representative work. Due to each sample having multiple labels, more related samples can be generated in the hidden space by defining the intersection and difference sets of the feature space. PFEMed [71] concatenates pre-trained general features with specific features to further enhance the semantic information of FSL features. Chen et al. [72] propose to add a set of reference images, which consists of many pairs of images of the same category. The features of the reference pairs are subtracted or added in the embedding space. It not only enriches diversity but also introduces reference features. In addition, AFHN [73] explores the use of conditional generative adversarial networks to produce a greater variety of recognition features. Recently, an assumption [12] is made that, when the base class and new class are semantically similar, the means and variances could be shared to a large extent. The mean and variance of the base class could correct the new class data distribution. Furthermore, Xu et al. [44] propose a framework for decomposing the variance in the dataset, where one represents the intra-class variance and others represent the embedding of the discriminative information. By repeated sampling, the intra-class variance can be added to the discriminative features. In this way, the features are learned to the maximum extent while maintaining a large intra-class variance.

3.3 Discussion and Summary

To maximize the evaluation of the real data distribution of FSL settings, several hypotheses have emerged. When the training and test domains are similar, a reasonable range of data is generated in the test domain using the mean and variance of the training domain. When the training and test domains are not similar enough, one approach is to introduce unlabelled dataset from the test domain, and another approach learns the intra-class or inter-class variance of the training domain to generate or transform new samples based on original data. Table 4 summarizes the methods in different dimensions, including whether the query set was used and whether additional labeled or unlabelled data were used. Table 5 provides a fair performance comparison of representative methods at the data level, where the methods are classified into two groups based on the backbone and ranked based on correctness rates from high to low in 5-way-1-shot task. It can be revealed from Tables 4 and 5 that the use of auxiliary datasets is more effective than the absence of auxiliary datasets, and feature augmentation is commonly more effective than data expansion. In the future, FSL data augmentation will continue to move towards more generalization and effectiveness.

4 FEATURE LEVEL: BUILDS DATA-TO-LABEL MAPPINGS FOR SPECIFIC PROBLEMS

The ability of deep learning models to represent features is far beyond humans with large-scale datasets and powerful computing power. FSL leverages feature knowledge to share pre-training parameters to varying degrees and redefine downstream tasks. In this section, following the direction of prior knowledge, we summarize the related work as transfer learning and multitask learning. Figure 10 shows the paradigm for transfer learning in FSL, including fine-tuning and prompt learning.
A Comprehensive Survey of FSL: Evolution, Applications, Challenges, and Opportunities

Table 4. Summary of the FSL Data Augmentation Work in Various Dimensions

| Method                               | Venue       | Type              | Experimental Dataset | Query set | External labeled set | External unlabeled set |
|--------------------------------------|-------------|-------------------|----------------------|-----------|----------------------|------------------------|
| Congealing [57]                      | CVPR'00     | Data Expansion    | MNIST                | ✓         |                      |                        |
| DAD [62]                             | ECCV'14     | Data Expansion    | Sign Language*       | ✓         |                      |                        |
| FTT [60]                             | CVPR'16     | Data Expansion    | TADB                 | ✓         |                      |                        |
| Low-shot-shrink-halfnurcaine [58]    | ICCV'17     | Data Expansion    | ImageNet             | ✓         |                      |                        |
| FSL [74]                             | ICCV'17     | Data Expansion    | CUB*                 | ✓         |                      |                        |
| Explore the Unknown Gradually [65]   | CVPR'18     | Data Expansion    | MARS                 | ✓         |                      |                        |
| Large-scale Diffusion [75]           | CVPR'18     | Data Expansion    | ImageNet             | ✓         |                      |                        |
| Δ -encoder [76]                      | NIPS'18     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| CF-AAN [66]                          | NIPS'18     | Data Expansion    | ImageNet             | ✓         |                      |                        |
| LST [65]                             | NIPS'19     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| Global Class Representations [77]    | ICCV'19     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| AutoAugment [52]                     | CVPR'19     | Data Expansion    | CIFAR10              |           |                      |                        |
| EDA [53]                             | EMNLP IJCNLP'19 | Data Expansion  | SST-2                |           |                      |                        |
| ΔSAO [70]                            | CVPR'19     | Feature Augmentation | Mlearned          | ✓         |                      |                        |
| ΔDeMe-Net [54]                       | CVPR'19     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| Learned Transformations [19]         | CVPR'19     | Data Expansion    | Biomedical Image     | ✓         |                      |                        |
| AΦHN [73]                            | CVPR'20     | Feature Augmentation | mini-ImageNet       | ✓         |                      |                        |
| DTN [72]                             | AAAJ'20     | Feature Augmentation | mini-ImageNet*       | ✓         |                      |                        |
| Neural-Snowball [78]                 | AAAJ'20     | Feature Augmentation | FewRel            | ✓         |                      |                        |
| STARTUP [46]                         | ICLR'21     | Data Expansion    | BSCD-FSL             | ✓         |                      |                        |
| Few_Shot_Distribution_Calibration [12]| ICLR'21   | Feature Augmentation | mini-ImageNet*       | ✓         |                      |                        |
| Relations+MT [79]                     | ICML'21     | Feature Augmentation | mini-ImageNet*       | ✓         |                      |                        |
| AdarGCN [67]                         | WACV'21     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| CSEI [55]                            | AFAJ'21     | Feature Augmentation | mini-ImageNet*       | ✓         |                      |                        |
| VFD [44]                             | ICCV'21     | Feature Augmentation | CUB                | ✓         |                      |                        |
| PAS [50]                             | ICCV'21     | Data Expansion    | iNat2019 CL*         | ✓         |                      |                        |
| PLCM [81]                            | ICCV'21     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| ΔlPC [47]                            | ICCV'21     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| Jr0s. [69]                           | NIPS'21     | Data Expansion    | BSCD-FSL             | ✓         |                      |                        |
| WSI [45]                             | ICLR'22     | Data Expansion    | NCT-CBCH-10K         | ✓         |                      |                        |
| FlipDA [41]                          | ACL'22      | Data Expansion    | ALBERT               | ✓         |                      |                        |
| PromDA [49]                          | ACL'22      | Data Expansion    | CoNLL03*             | ✓         |                      |                        |
| TODST [50]                           | ACL'22      | Data Expansion    | MultiWOZ             | ✓         |                      |                        |
| RSVAE [43]                           | CVPR'18     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |
| CLUSTER-FSL [48]                      | CVPR'22     | Data Expansion    | mini-ImageNet*       | ✓         |                      |                        |

The mark * means that there is more than one dataset in experiments.

Table 5. Comparing the Performance of Data-level Representative Methods Using the Mini-ImageNet as Benchmark Dataset

| Method                          | Backbone | Setting       | 5-way-1-shot | 5-way-5-shot |
|---------------------------------|----------|---------------|--------------|--------------|
| AFHN [73]                       | ResNet-18| Inductive     | 62.38% ± 0.72% | 78.16% ± 0.56% |
| Δ -encoder [76]                 | ResNet-18| Inductive     | 59.8%        | 67.9%        |
| ΔDeMe-Net [54]                  | ResNet-18| Inductive     | 59.14% ± 0.86% | 74.63% ± 0.74% |
| Relation+MT [79]                | ResNet-18| Transductive  | 55.7%        | 70.29%       |
| CLUSTER-FSL [48]                | ResNet-12| Semi-supervised | 77.81% ± 0.81% | 85.55% ± 0.41% |
| RSVAE [43]                      | ResNet-12| Inductive     | 73.35% ± 0.37% | 80.95% ± 0.31% |
| PLCM [81]                       | ResNet-12| Semi-supervised | 72.06% ± 1.08% | 83.71% ± 0.63% |
| LST [65]                        | ResNet-12| Semi-supervised | 70.1% ± 1.9% | 78.7% ± 0.8% |
| iLPC [47]                       | ResNet-12| Transductive  | 69.79% ± 0.99% | 79.82% ± 0.55% |
| CSEI [55]                       | ResNet-12| Inductive     | 68.94% ± 0.28% | 85.07% ± 0.5% |
| DTN [72]                        | ResNet-12| Inductive     | 63.45% ± 0.86% | 77.91% ± 0.62% |

4.1 Transfer Learning

From the perspective of transfer learning, FSL can be regarded as a cross-domain learning task. When relevant tasks have been extensively obtained, FSL attempts to generalize certain prior knowledge that is useful for the target task. Standard transfer learning in FSL can also be divided into two phases: learn on source tasks and transfer on the target task. Among them, the second stage can be implemented in different ways.
Fig. 10. Transfer learning can be divided into pre-training and fine-tuning stages, where the baseline model can be combined with other techniques to improve the model performance.

Fig. 11. Prompt learning reconstructs different tasks to adapt pre-trained language model. Specifically, prompts can be manual, abstract, and automatically generated.

4.1.1 Pre-training. During the pre-training stage, a high-volume network, including ResNet [82] families, VIT [83] families, ELMo [84], BERT [85], and GPT [86] families, is trained on a large base class dataset that covers similar downstream tasks. Pre-training processing can be conducted in supervised learning or unsupervised learning. Supervised learning is a common approach that requires the design of special loss functions. Note that our survey focuses on self-supervised and unsupervised learning works, which can achieve more competitive results compared with supervised learning in FSL.

The most representative work is Transformer [83], which is an unsupervised training method in natural language processing. Based on this, BERT [85] uses Transformer in the encoder and focuses on contextual information during training. Alternatively, GPT [86] uses Transformer in the decoder. Compared to RNN, it is more easily parallelizable. There exist many different versions of GPT, each with an increasing number of parameters and lengths of word sequences to process. GPT-2 [87] introduces multi-task learning, where the whole training process requires more parameters and a larger dataset. GPT-3 [88] is inspired by meta-learning, and its inner loop and outer loop are responsible for updating different parameters, respectively. Recently, iGPT [89] extends GPT into computer vision, which predicts the bottom half of an image on a pixel-by-pixel basis from the top half of the retained image using auto-regressive. Subsequently, MAE [90] tries to partly mask the image with the assistance of self-supervised learning. By reconstructing the original image, models learn how to represent semantics features.

FSL leverages pre-trained models’ feature representation capacity to largely reduce intra-class variation [91], which leads to the model focusing on more discriminative regions. Experiments prove that the supervised training approach would distort instances of different classes and ignore semantic information. On the contrary, unsupervised training focuses on aligning the features for downstream tasks.

4.1.2 Fine-tuning. When high-level semantic features are closely related to the target task, fine-tuning paradigm assumes that ground-level feature knowledge can be reused in downstream tasks. Thus, it is vital to identify the specific layers and learning rates during fine-tuning.

Baseline [92] uses a standard feature extractor followed by a fully connected layer. Baseline++ [92] replaces the fully connected layer with a cosine metric function. When the computing power is sufficient, evolution strategies [93] are good choices to determine the number of retrain layers and learning rates. In addition, experiences convince that just re-randomizing the top-level parameters [94] and self-distillation of the same architecture knowledge [95] are also beneficial to FSL. Recently, many works [96, 97] explore useful tricks, including self-supervised learning [96],
Table 6. Summary of the Pre-training and Fine-tuning Methods in Various Dimensions Including Serval Tricks

| Method          | Venue       | Backbone Parameters | Tricks                          | Refining Parameters | Knowledge distillation | Attention |
|-----------------|-------------|---------------------|---------------------------------|--------------------|------------------------|-----------|
| Squad [101]     | EMNLP'16    | Logistic Regression |                   |                    |                        |           |
| Attention in All You Need [102] | NIPS 17    | Transformer 110M    |                   |                    |                        |           |
| Baseline [92]   | ICLR'19     | Conv64F 0.22M       |                   |                    |                        |           |
| Baseline++ [92] | ICLR'19     | Conv64F 0.22M       |                   |                    |                        |           |
| CAN [98]        | NIPS 19     | ResNet12 12.47M     |                   |                    |                        |           |
| GLUE [103]      | ICLR'19     | ELMo 94M            |                   |                    | ✓                      |           |
| RPS-distill [95]| ECCV'20     | ResNet-12 12.47M    |                   |                    | ✓                      |           |
| LEE et al [104] | WPT 20      | GPT-2 1508M         |                   |                    | ✓                      |           |
| CTX [97]        | NIPS 20     | Transformer 110M    |                   |                    | ✓                      |           |
| SKD [96]        | CVPR'20     | ResNet12 12.47M     |                   |                    | ✓                      |           |
| P-Transfer [93] | AAAI'21     | ResNet-12 12.47M    |                   |                    | ✓                      |           |
| BERT Fine-tuning [105] | CL 21 | BERT 110M          |                   |                    | ✓                      |           |
| COSOC [106]     | NIPS 21     | ResNet-12 12.47M    |                   |                    | ✓                      |           |
| P-M-F [100]     | CVPR'22     | VIT 307M            |                   |                    | ✓                      |           |
| ReFine [94]     | CVPR'22     | ResNet-10 9.86M     |                   |                    | ✓                      |           |

Table 7. A Detailed Summary of FSL Prompt Learning in Natural Language Processing

| Method          | Venue     | Backbone | Parameters | Type      | Shape          | Highlight                  | Highlight                              |
|-----------------|-----------|----------|------------|-----------|----------------|----------------------------|----------------------------------------|
| PT [10]         | ACL'21    | GPT-3    | Manual     | Prefix Prompt |                | Combine input vector and prefix prompts |                                        |
| Brown [88]      | NIPS 20   | GPT-3    | Manual     | Prefix Prompt |                | Design task-related prompt |                                        |
| Petroni et al. [87] | EMNLP'19 | BERT     | Manual Prompt | Cloze-style |                | Use close-style prompt to learn prior knowledge |                                        |
| Huang et al. [114] | ACL-NLP'21 | BERT     | Manual     | Prefix Prompt |                | Use close to constraint tasks in entity classification |                                        |
| PET [112]       | LAL '23   | PLM      | Manual     | Prompt      | Cloze-style     | Introduces fine-tuning and knowledge distillation |                                        |
| LPAQA [113]     | MT '20    | LMS      | Auto-Prompt | Cloze-style  |                | Collect manual templates as template library |                                        |
| AutoPrompt [114] | ACL '20   | MLMs     | Auto-Prompt | Cloze-style  |                | Automatically search template by collect prior prompt |                                        |
| LM-Prompt [113] | ACL '21   | GPT-3    | Auto-Prompt | Cloze-style  |                | Restore the span of the mask sentences |                                        |
| PT [108]        | ACL '21   | GPT-3    | Latent     | Prompt      |                | Add prefix ID in front of input text |                                        |
| CP-Tuning [112] | ACL '22   | PLM      | Latent     | Prompt      |                | Continuous prompt embedding instead of manually designed prompt |                                        |
| OptIPrompt [113] | NAACL '21 | BERT     | Latent     | Prompt      |                | Initialize with prompt templates and fine-tuning |                                        |
| OPT-understands [112] | Arxiv '21 | GPT-3    | Latent     | Prompt      |                | Model output is regarded as hidden vector to optimize |                                        |

attention module [98, 99], and mask generation module [97]. Clearly, P > M > F [100] summarizes the training process across pre-training, meta-learning, and fine-tuning. It achieves the state-of-the-art level of FSL on multiple benchmark datasets. Among them, the feature extraction backbone is the dominant factor in FSL performance.

Similarly, in natural language processing, pre-training and fine-tuning have greatly accelerated FSL development. Among them, BERT [85] is mainly for language understanding and the GPT [86] is for text generation. The language model trains different tasks on a large corpus so it has the ability to understand [103] or even generate target text [101]. Zhang et al. [105] explore the main tricks of fine-tuning BERT [85] including error correction, weight initialization, L2 regularization, and freezing. However, Lee et al. [104] fine-tune GPT-2 and achieve the state-of-the-art on the automatic patent assertion task. Aside from performance, the biggest advantage of fine-tuning is that it does not require task-level model structure design. Table 6 summarizes the main work of pre-training + fine-tuning under FSL in terms of backbone and fine-tuning tricks.

4.1.3 Prompt Learning. Prompt learning is a novel paradigm that achieves competitive performance in zero or few sample scenarios. In contrast to fine-tuning, prompt learning is not necessary to design a loss function but requires a slight manual prompting to the pre-trained model. It retains the maximum capability of the pre-trained model so it can better understand the downstream task. According to the location of the prompt, prompt learning can be divided into prefix prompt [107, 108] and cloze-style prompt [87, 107, 109]. Specifically, prefix prompt indicates that subsequent text is combined with prefix words corresponding to different tasks [88, 110]. Cloze-style prompt means to fill suitable words in the blanks of sentences. Different tasks [107] could have different blanks positions to test whether the model has learned the corresponding semantic knowledge [107, 111]. Table 7 and Figure 11 summarize the widely used prompting strategies for FSL.
The significant problem of manual prompting templates is that the quality of prompting directly affects the model’s performance. Jiang et al. [113] collect some transforms as templates in a huge amount of text anticipation. New tasks could search for templates in the huge amount of experiences. Another method is prompt paraphrasing. Based on the seed prompt, back translation and keyword substitution expand it to more prompts. The most representative work is Auto-Prompt [114], which masks the words in the template that auto-replaces them with other suitable words to maximize the probability of labeling. Gao et al. [115] use the T5 model for training and comparing to generate each position of a blank. The whole process is relatively computationally intensive.

Going further, a more abstract type of prompt is prompting hidden spaces. The inputs can be embedding vectors, and outputs are no longer concrete words. The most representative method is PT [108], which adds several prefix IDs in front of the normal input. Except for initialized parameters, all other parameters are frozen and updated only when the prefix is adjusted. Inspired by this idea, CP-Tuning [116] uses continuous prompt embedding instead of the end-to-end manual design of prompt templates. Afterward, several works start to explore how to initialize prompts in hidden spaces. PPT [119] points out that better initialization can be obtained by adding soft prompts to the pre-train stage. Another way [117, 118] is to use the prompt as an initial state and find a more appropriate latent prompt in the process of fine-tuning.

### 4.2 Multitask Learning

Compared to transfer learning, multi-task learning requires multiple loss functions to optimize the model. Extended tasks tend to update in different directions in the embedding space, which offsets some noises to a certain extent. In FSL multi-tasking works, parameter updates can be classified as hard parameter sharing and soft parameter sharing.

To the best of our knowledge, FSL multitask learning [120] was initially applied in video event detection tasks. In feature representation, by controlling the position of parameter updates [121–123], multi-task learning enables maximum confusion of shared layer parameters. This operation of separating features [124] maximizes the balance between the learning of generic and discriminative features. This idea was applied to automatic charge prediction [125] by adding several discriminative attributes, eliminating attribute irrelevant features, and using attribute-relevant features for accusation prediction. Multi-task learning can also be used for data augmentation. FGVC [126] proposes to add a sample selection task to the pre-trained task. Each training classifier shares a base network to achieve better results.

Following this process, FSL multi-task learning was associated with self-supervised learning for a long time. Figure 12 summarizes the self-supervised training of the main pretext tasks. Specifically, BF3S [127] combines the FSL classification task with predicting the rotation angle of images. Similarly, The pre-tasks were also used to color the image [128] and predict the relative position of each patch [129, 130], local Fisher discriminant [131]. Extensive experiments show that predicting the rotation angle and relative position of the image patch are more effective methods for FSL. PSST [132] further combines FSL classification, image position prediction, and angle prediction to achieve the state-of-the-art. The intuition was that by solving these tasks, the backbone network could extract more semantic features not only related labels. As meta-learning is combined with well-designed multi-task learning, it is possible to achieve competitive results and reduce the training time by 10 times.

### 4.3 Discussion and Summary

FSL based on the feature level is a wide research direction. Indeed, the selection of a feature extractor is particularly important. A good feature extractor can extract discriminative information
Contrast learning is involved in FSL multitask work, including coloring, puzzles, auto-view, and foreground extraction.

Table 8. Performance Comparison of Feature-level Representative Methods Using the Mini-ImageNet as Benchmark Dataset

| Methods               | Backbone      | Setting          | 5-way-1-shot   | 5-way-5-shot   |
|-----------------------|---------------|------------------|----------------|----------------|
| Baseline++ [92]        | Conv-4-64     | Fine-tuning      | 63.85% ± 0.48% | 66.43% ± 0.63% |
| BF3S(Rotation) [127]   | Conv-4-64     | Multitask Learning | 54.83% ± 0.43% | 71.86% ± 0.33% |
| BF3S(patch location) [127] | Conv-4-64   | Multitask Learning | 54.30% ± 0.42% | 71.58% ± 0.33% |
| Baseline [92]          | Conv-4-64     | Fine-tuning      | 42.11% ± 0.71% | 62.53% ± 0.69% |
| COSOC [106]            | ResNet-12     | Fine-tuning      | 69.28% ± 0.49% | 85.16% ± 0.42% |
| SKD [96]               | ResNet-12     | Fine-tuning      | 65.93% ± 0.81% | 83.15% ± 0.54% |
| RFS-distill [95]       | ResNet-12     | Pre-trained      | 64.82% ± 0.6%  | 82.14% ± 0.43% |
| P-Transfer [93]        | ResNet-12     | Fine-tuning      | 64.21% ± 0.77% | 80.38% ± 0.59% |
| PSST [132]             | ResNet-12     | Multitask Learning | 64.05% ± 0.49% | 80.24% ± 0.45% |
| CAN [98]               | ResNet-12     | Fine-tuning      | 63.85% ± 0.48% | 79.44% ± 0.34% |

from a few samples. On this basis, multi-task learning can better guide the updating of model parameters if there are multiple similar tasks in reality. Of course, if the relevant task is difficult to collect, then fine-tuning is also a good option. In conclusion, there are two main types of fine-tuning paradigms: pre-training and prompt learning. Among them, prompt learning retains the better capability of the pre-trained model for each task than fine-tuning.

During this processing, contrast learning throws off the labels and learns a more rich knowledge from the data itself. Note that contrast learning requires sufficient computing resources. Table 8 compares feature-level representative methods in a fair way. As the prompt learning experiments do not involve the mini-ImageNet, the comparison is not included here. For tiny-capacity networks, multi-task learning has certain advantages over fine-tuning, and for large-capacity networks, fine-tuning has a definite advantage. Much work has shown that most FSL tasks can be well solved even only using fine-tuning. In the future, contrast learning combined with multi-task learning and fine-tuning will be a hot direction to explore.

5 TASK LEVEL: DERIVE META-KNOWLEDGE TO TARGET TASK MAPPINGS INDEPENDENT OF SPECIFIC PROBLEMS

Task level is different from data level and feature level, where it performs dual sampling for data and tasks to extract meta-knowledge. Meta-knowledge is independent of the specific problem and searches for an optimal parameter in task space. Broadly speaking, task level learns to
optimize parameters, generates metric functions, and summarizes knowledge transfer. Among them, learning optimization parameters also includes optimizing meta-learned parameters and optimizing existing model parameters; generating metric algorithms includes feature embedding, external memory-based, metric learning, graph neural networks, and other similarity-based algorithms.

5.1 Learning Optimization Meta-learned Parameters
The key idea of learning meta-learned parameters is to search for a global initialization state that is general to unseen tasks. Traditional initialization methods, such as uniform distribution and normal distribution, easily fall into local optimum. Considering the specificities of FSL, **MAML** (model-agnostic meta-learning) [133] proposes a new method based on episodic training. MAML is a typical two-stage parameter model. The first stage is updating each task, namely, a local update. The second update is an average of the task query losses in each batch, namely, a global update. The volume of calculations is the biggest problem. To this end, extensive work aims to improve MAML, including simplifying [134], momentum update [135], or even neglecting second-order derivatives [136, 137], using evolutionary algorithms [138] to avoid back-propagation to run well on the CPU [139] and considering gradient update directions and learning rates [140].

In addition, relevant initialization parameters can also be generated by the other models. **AWGIM** [141] is based on low-dimensional [142] introducing mutual information and attention, which generates weights containing more query information. Furthermore, MAML is combined with probability distributions to derive **LLAMA** [143], **PLATIPUS** [144], **Bayesian MAML** [145], **DKT** [146], and **ABML** [147]. As far as enlarging parameter space, **MT-NET** [148] adds transformation matrix and a binary mask matrix; **TAML** [149] introduces regularization conditions; **OOD-MAML** [150] draws on dummy samples; **ARM** [151] constructs meta-knowledge graphs; **WGD-MAML** [152] places non-linear activation function between learned layers; **Reference** [153] introduces teacher-student network widened longer internal horizon. **UNICORN-MAML** [154] summarizes the above work and gives a few suggestions on how to train MAML to compete with the state-of-the-art methods. Recently, **METADOCK** [155] compresses the meta-models by dynamically selecting the kernels. It can be easily deployed on edge devices. Figure 13 provides an overview and summary of these related works.
In addition to optimizing initialization parameters, another important direction is model-based mechanisms. It requires an external memory to store prior knowledge in the form of key-value pairs. The key stores the output of the model embedding and the value stores various labels. To our best knowledge, MANN [156] is the earliest work that uses external memory to explicitly store feature embedding in each episode. Kaiser et al. [157] expand the key-value memory module to use a triplet for preservation. Additional vectors are used to select one that has not been updated for a long time. APL [158] further summarizes the model-based knowledge as a probability distribution. Different from existing work that applies static memory matrix, DMIN [159] uses dynamic memory routing to learn dynamic memory modules. Similarly, Reference [160] models the relationship of individual historical tasks in the memory module, while Reference [161] maintains a stable prototype representation during training processing. In several extremely cross-domain scenarios, the keys in the memory module cannot be simply generalized to semantics but should be multi-layered [162].

Inspired by MAML and Reptile, network architecture search could also achieve fast convergence in a few rounds of updates. The shared [163] or randomly selected supernet weights [164] were early works named one-shot NAS. But the performance of one-shot NAS still falls relatively behind the traditional NAS. Subsequently, Zhao et al. [165] propose the few-shot NAS. The core idea is to divide the super-networks into multiple sub-supernets to search different regions of the search space. Due to a slight increase in the number of supernets, the accuracy of few-shot NAS has been greatly improved. MetaNAS [166] is the first method that completely integrates meta-learning and traditional NAS. It replaces the weighted summation of DARTS [167] to reduce different operations. Experiences show that MetaNAS is more adaptable to downstream learning tasks.

5.2 Learning Metric Algorithm

Learning metric algorithm aims to learn a mapping that computes the abstract distance between the support-query pairs. In References [168, 169], metric learning is presented as a separate chapter in meta-learning. Generally speaking, metric learning could also be seen as learning of a similarity measure that integrates the idea of learning to learn. Table 9 summarizes the extensive work on FSL metric learning.

Siamese neural network [170] is the relatively early model. The input consists of a set of positive or negative pairs, and the output is simply a binary classification problem. Based on siamese neural network, more works consider adding an anchor sample [171], or even more samples [172], and selecting a hard sample [173]. Siamese neural network is a general framework. When the backbone model does not share weights, siamese neural network will become the pseudo siamese network.

Compared to the siamese neural network, the prototype network cited in Reference [174] based on feature averaging realizes the true meaning of classification. But simply averaging feature is easily disturbed by noise. Several works [175–177] propose different types of adaptive boundary distances between similar categories and dissimilar categories. In addition to this, selecting specific parts of the image [178] to enforce the interaction of higher-order semantic features [179] can guide the model to learn prototype representations in the new metric space [180]. APLCNE [181] is the most representative work, which replaces the CNN with a capsule network to encode spatial location information. Prototypes are calculated using weighted summation. If the sample is more sparse, then it is fine to use regularization to constrain the length of prototypes in different modalities [182] as well as introduce attention to reconstruct the prototype representation [183–185].

Matching Networks [23] combine the parametric and non-parametric algorithms to model the distance distribution. The most commonly used metric includes Euclidean distance, Manhattan distance, and cosine similarity. Recently, DEEPEMD [186] first proposes a cross-referencing
Table 9. A Summary of Metric Learning Methods Based on Baseline Approaches (Highlighted in Bold)

| Method                        | Venue         | Sample Selection | Metric Function | Purpose                                                                 |
|-------------------------------|---------------|------------------|-----------------|-------------------------------------------------------------------------|
| Siamese Neural Network [170]  | ICML'15       | Pair of Positive and Negative | European Distance | The distance between two similar inputs as small as possible, and the distance between two different categories of inputs as large as possible. |
| Triple loss [171]             | SBMAD'15      | Anchor, Positive, Negative | European Distance | The distance between Anchor and Positive is smaller than that with Negative. |
| E-navigate sample [173]       | CVPR'17       | Hard Sample      | Fully Connected Layers | Not only considers the relative distance between positive and negative samples, but also adds the absolute distance between positive samples. |
| K-tuple [172]                 | Neurocomputing'20 | K-tuple Input    | Fully Connected Layers | Revised the classical triplet network and extended it to a K-tuple network. |
| MSN [199]                     | IEEE 21       | Pair of Positive and Negative | Multiple Distance Metrics | Introduce distance between Anchor and Prototype to assist the optimization. |
| Muller et al. [200]           | ACL 22        | Random Input and Label Texts | Dot Product | The adjusted tag embedding scores each tag pair, and assigns weights an action network and object network. |
| PSN [201]                     | SIGIR'21      | Pairs of Action Sentence and Object Sentence | Fully Connected Layers | PSN consists of two identical subnetworks with the same structure but different weights an action network and object network. |
| Prototype Network [174]       | NIPS 17       | Episodic Training | European Distance | Compare the distance between the prototype and the query sample. |
| BFIS [175]                    | CVPR'20       | Episodic Training | Cosine Similarity | Adaptive boundary distances between similar categories larger than between dissimilar categories. |
| Negative Margin [176]         | CVPR'20       | Episodic Training | Cosine Similarity | Proposing negative margin for alleviating the problem of learning multiplex with few samples. |
| Cao et al. [178]              | ICLR 21       | Episodic Training | Euclidean Distance | Using concepts shared across many images as prototype. |
| Embedding propagation [179]   | ECCV'20       | Episodic Training | GNN | Improve higher-order interaction information between representations. |
| SGAP-Net [180]               | AAAI 20       | Episodic Training | PS-Module | Learns a new metric space through a semantically guided model. |
| APLCNE [181]                 | ECCV'20       | Episodic Training | Triplet Loss | Further considers the spatial correlation of image. |
| SEN [182]                     | ECCV'20       | Episodic Training | Euclidean Distance | MDP feature similarity between the length of the prototype. |
| RestoreNet [183]              | AAAI 20       | Episodic Training | Fully Connected Layers | Prototype as the sum of the original vector and the vector reconstructed via MLP layer. |
| PANet [202]                   | ICCV'19       | Episodic Training | Cosine Similarity | Matching the similarity of each pixel to the prototype. |
| HATT-Proto [184]              | AAAI 19       | Episodic Training | Euclidean Distance | Design hybrid networks of attention for instance and prototype features. |
| CAD [185]                     | CVPR'22       | Episodic Training | Fully Connected Layers | Calculates the attention scores between features and then produces a representation via a mapping header. |
| MetaNODE [205]               | AAAI 22       | Episodic Training | Cosine Similarity | Prototype bias can continue to be addressed by meta-optimization. |
| Matching Network [23]         | NIPS 16       | Episodic Training | Cosine Similarity | Using external memory to improve its learning ability. |
| DIEEPDM [186]                | CVPR'20       | Episodic Training | Earth's Mover's Distance | Using Earth’s Mover’s Distance end-to-end training and proposed a cross-referencing mechanism. |
| Adap-EMD [187]                | IEEE 22       | Episodic Training | Earth’s Mover’s Distance | Dynamically computes the fine-grained relationships between each set of vectors. |
| Deep Brownian Distance Covariance [188] | CVPR'22 | Episodic Training | Brownian Distance | Learn image representations by measuring the difference between the joint distribution of embedded features and the product of marginal distributions. |
| Relation Network [189]        | CVPR'18       | Episodic Training | Fully Connected layers | Calculating similarity by using neural networks. |
| RSNET [190]                  | IEEE 20       | Episodic Training | Bi-Fully Connected layers | Dual similarity module by combining the cosine module with other modules. |
| SAML [191]                   | ICCV'19       | Episodic Training | Fully Connected layers | Association matrix is passed through the MLP layer to obtain a similarity score. |
| GNN-FSL [193]                | ICLR'18       | Episodic Training | GNN | Feeding the vector of samples after the feature extractor directly into a graph neural network for training to predict labels. |
| EGNN [194]                   | CVPR'19       | Episodic Training | GNN | Designing the edge-labeled graph neural network, which not only predicts edge node labels but also predicts different classes. |
| Meta-CCN [191]               | ICLR'20       | Episodic Training | GNN | Updating of weights of graphs could also be optimized according to the gradient descent steps. |
| TIPN [196]                   | ICCV'20       | Episodic Training | GNN | Considers relationships between pairs of support queries for the first time and models them explicitly as graph nodes. |
| HGN [197]                    | AAAI 19       | Episodic Training | GNN | Considers GNN as a label propagation tool that can be co-trained with feature embedding networks. |
| DPNG [198]                   | CVPR 20       | Episodic Training | GNN | Introduces the concept of point and distribution maps. |
Fig. 14. Metric learning can be divided into three parts: sampling strategies, backbone networks, and metric methods, where red line and orange line are based on siamese neural network models, blue line is based on prototype networks, and brown line is based on matching networks, relational networks, and graph neural networks. Among them, $P$ refers to a positive sample, $N$ refers to a negative sample, $A$ refers to anchor sample, $S$ refers to the support set, and $Q$ refers to the query set.

mechanism with Earth’s Mover’s distance. It uses the dot product of the node feature and the average node feature in another structure to generate a relevance score as a weight value. DEEPEMD could apply in remote sensing scenes [187]. Deep Brownian Distance Covariance [188] is another lower computing method that learns image representations by measuring the difference between the joint distribution of embedded features and the product of marginal distributions. Experiments have shown that Brownian Distance Covariance has great potential and wide applications in FSL metric learning, see Figure 14.

Relation network [189] differs from the above-mentioned models in that its similarity is calculated by using a neural network combined with the cosine similarity [190], attention module [191], and graph neural network [192]. Among them, graph neural networks are used to solve a large class of FSL problems. One of the earliest works is GNN-FSL [193], which feeds the vector representation of samples into a graph neural network. On this basis, by designing different embedding nodes [194–196] and label propagation mechanisms [197, 198], graph neural networks can model the relationships between supported query pairs quite competently.

5.3 Discussion and Summary

Task level prior knowledge covers all learning to learn methods, including learning to optimize parameters and learning metric functions. Specifically, the work involves metric learning, transfer learning, meta-learning, model-based memory, embedding learning, graph neural networks, and many more. Learning meta-learned parameters can largely obtain a good initialization for unseen tasks. Certainly, several algorithms achieve better results by fine-grained domain transfer rules. However, this is a considerable challenge in terms of computing resources. Table 10 compares representative meta-learning methods, where the MAML-based approaches [133, 146] further optimize the inner and outer loop to achieve new results on few-shot classification tasks, and the approaches based on new metrics [186, 188] have been greatly improved compared to the approaches based on cosine similarity [204] and Euclidean distance [174].

In the future, more cross-modal tasks are expected to be integrated into the meta-learning framework, where tasks between different modalities can be mutually reinforcing through the interaction of the intelligence with the environment, making it possible to train models without having to train parameters from scratch, especially when it encounters fine-grained FSL tasks.
Table 10. Performance Comparison of Task-level Representative Methods Using the Mini-ImageNet Dataset as an Example

| Method                  | Backbone     | Setting                      | 5-way-1-shot        | 5-way-5-shot        |
|-------------------------|--------------|------------------------------|---------------------|---------------------|
| DKT [146]               | Conv-4-64    | Meta Learning                | 49.73% ± 0.07%      | -                   |
| MAML [133]              | Conv-4-64    | Meta Learning                | 48.70% ± 1.75%      | 63.11% ± 0.92%      |
| DeepBDC [188]           | ResNet-12    | Brownian Distance            | 67.83% ± 0.43%      | 85.45% ± 0.29%      |
| DEEPEMD [186]           | ResNet-12    | Earth Mover’s Distance       | 65.91% ± 0.82%      | 82.41% ± 0.56%      |
| Meta-baseline [204]     | ResNet-12    | Cosine Similarity            | 63.17% ± 0.23%      | 79.26% ± 0.17%      |
| Prototype Network [174] | ResNet-12    | Euclidean Distance           | 60.37% ± 0.83%      | 78.02% ± 0.57%      |

Fig. 15. Multimodal learning models various modal information by fusion, alignment, and translation.

6 MULTIMODAL: LOSSLESS REPRESENTATION OF MULTIMODAL INFORMATION

Although several works [185–187] have provided excellent results for specific tasks with FSL, FSL is still difficult to make large breakthroughs with more generalized prior knowledge, such as BSCD-FSL [22]. Nowadays, due to the effectiveness of prefix tuning and prompt tuning, FSL performs surprisingly well in zero-shot learning. Figure 15 shows several key techniques for multimodal FSL across text, image, and audio.

CLIP [205] gives a basic idea for multimodal training: joint training of text backbone and visual backbone. Based on CLIP, CADA-VAE [206] uses VAE as the backbone to map image features and label text into the same latent space. Wang et al. [207] decompose the image into the original image, foreground image, and background image to obtain fused visual features through CNN backbone. Li et al. [208] apply unsupervised clustering to obtain layer-level semantic information, and Schwartz et al. [209] use multiple MLP layers to extract the semantic prototypes with labels vector through the textual backbone. Peng et al. [210] use a GCN to generate the semantic information of the categories into corresponding classification weights, and the FSL classifier is completed using visual classification weights in combination with semantic generation weights. Xing et al. [211] add text vector to a prototype network together with visual features to train an adaptive classification. Similar to Reference [211], Pahde et al. [212] use semantic generation to aid visual features. CCAM [213] replaces the real labels by encoding contextual prototypes and classifies them by comparing the distance between the individual prototypes. Using backbone network embedding of the respective domains is simple to train, but good language models and visual models co-trained can largely damage the performance of the language model.

Multimodal prompt learning allows visual embeddings to consistently adapt to the language model by prompting or adding prefixes, which preserves to a large extent the powerful feature representation capabilities of the language model. The prompts of MCNLG [214] are very intuitive, which places multimodal sequences as prefix prompts in front of input sequences that decoder in turn the shared multimodal information. The visual backbone is ResNet-152 and the text backbone is the embedding network. ActionCLIP [215] is the application of CLIP [205] in action.
Table 11. Performance Comparison of Task-level Representative Methods Using the Mini-ImageNet as Benchmark Dataset

| Method         | Backbone | Setting  | 5-way-1-shot | 5-way-5-shot     |
|----------------|----------|----------|--------------|------------------|
| Schwartz et al. [209] | ResNet-10 | Discriminant | 67.3%        | 82.1%            |
| Peng et al. [210]   | ResNet-10 | Generation | 54.34% ± 0.77% | 69.02% ± 0.65%  |
| Xing et al. [211]  | ResNet-12 | Discriminant | 65.30% ± 0.49% | 78.10% ± 0.36%  |
| Li et al. [208]    | WRN-28-10 | Discriminant | 64.40% ± 0.43% | 83.05% ± 0.28%  |

In the future, with the emergence of more and more multimodal pre-trained models, traditional FSL will enter the era of large models + few sample fine-tuning. Multimodal pre-training models have powerful generality, which can make up for the lack of language models in multimodal scenario applications end-to-end. Multimodal FSL brings the capabilities of FSL one step closer to humans.

6.1 Discussion and Summary

Compared to unimodal learning, multimodal FSL is still in the development stage. Table 11 compares the representative methods of multimodal FSL. The experimental results show that, under the same condition, the discriminative methods achieve slightly better performance than the generative methods. The discriminative methods mainly focus on the design of semantic information prototypes and the selection of metric algorithms.

Currently, multimodal FSL also exists with many additional challenges: how to integrate data from heterogeneous domains; how to deal with the different levels of noise that arises during the combination of different modalities; and how to align common learning feature representations in the same space. The various experiments show that prompt learning is significantly better than co-training and is more suited to the practical situation where few samples are learned to update a small number of parameters. In multimodal FSL, a good feature representation should be able to assist in the missing modal information based on the observed modal information. Next, more multimodal pre-training models will emerge across text, image, and audio, together driving FSL towards environmental interaction.

7 FSL APPLICATIONS IN COMPUTER VISION

At present, FSL is deeply integrated into a wide range of industries. Researchers expect FSL to handle reasoning tasks in a variety of extreme conditions as easily as humans. The performance of computer vision tasks largely represents the highest level of prior knowledge. In this section, we
provide a detailed graphical summary of four main tasks in computer vision: classification, object detection, instance segmentation, and semantic segmentation. This section will provide readers with a comprehensive overview of the latest achievements in the field of computer vision.

7.1 Few-shot Image Classification

As the number of parameters of the pre-trained backbone grows towards tens and hundreds of billions, the regular few-shot image classification tasks have been well addressed. In some extreme cross-domain conditions, FSL is still not well resolved, such as EuroSAT [39], ISIC2018 [40], Plant Disease, and ChestX-Ray8 [41]. With different levels of prior knowledge, few-shot image classification could be summarized in various dimensions. The best result [100] is achieved based on the pre-training + fine-tuning combined with the meta-learning. It largely surpasses transductive few-shot learning. In this section, we investigate the few-shot image classification work from 2016 to 2022 and count the best performance on the mini-ImageNet benchmark dataset according to 5-way-1-shot and 5-way-5-shot tasks. Table 12 and Figure 16 illustrate our investigation results.

7.2 Few-shot Object Detection

Traditional object detection focuses on extracting features using a powerful pre-trained backbone network. However, apart from the backbone network, other detector components, such as detector heads and feature pyramid networks, are still randomly initialized. In FSL object detection, most of the work is based on improvements to the original models, such as Faster R-CNN [245] and YOLO [246]. In a recent paper [247], using a visual transformer as the backbone network and embedding a decoder into the detector head while removing the pyramidal feature network from the feature extractor, it does not need any additional conditions. Table 13 and Figure 17 show recent advances in object detection in FSL.

7.3 Few-shot Semantic Segmentation

To the best of our knowledge, the few-shot semantic segmentation was first proposed [255] in 2017. And it has been widely used in medical images and driverless cars. In contrast to traditional semantic segmentation, few-shot semantic segmentation has less pixel annotation on the support dataset. Few-shot semantic segmentation can be broadly classified into supervised semantic segmentation, unsupervised semantic segmentation, and video semantic segmentation. In the machine learning area, the classical method is to use probabilistic mappings as prior knowledge for derivation. Recently, Reference [256] has made significant improvements to few-shot semantic segmentation by
proposing a concise paradigm where only the classifier is meta-learned and the feature encoding decoder remains trained using a conventional segmentation model. Table 13. The Overview Table of the State-of-the-art on MS COCO Object Detection with Some Special Tricks

| Method   | Venue   | Type                     | Backbone | Tricks | mini-ImageNet | Available Code |
|----------|---------|--------------------------|----------|--------|---------------|----------------|
| P-M+F [140] | CVPR’22 | Fine-tuning               | ViT      | ✓✓✓✓   | ✓              | ✓              |
| SOT [208] | CVPR’22 | Fine-tuning               | ResNet-12 | ✓✓✓✓   | ✓              | ✓              |
| ESFR [221] | PMLR’21 | Embedding                 | ResNet-18 | ✓✓✓✓   | ✓              | ✓              |
| Multi-task Learning [222] | PMLR’21 | Fine-tuning               | ResNet-12 | ✓✓✓✓   | ✓              | ✓              |
| IS-FSL [223] | BMVC’21 | Multimodality Transformer | ✓✓✓✓     | ✓✓✓✓   | ✓              | ✓              |
| Invariance-equivariance [224] | CVPR’21 | Embedding                 | ResNet-12 | ✓✓✓✓   | ✓              | ✓              |
| DSL [225] | Arxiv’21 | Embedding                 | ResNet-12 | ✓✓✓✓   | ✓              | ✓              |
| MATAN [226] | Arxiv’21 | Metric                   | Conv-64F | ✓✓✓✓   | ✓              | ✓              |
| SSL-FE [227] | ICASSP’21 | Embedding                 | AadimNet | ✓✓✓✓   | ✓              | ✓              |
| MORN [228] | R-PR’21 | Meta                     | ResNet-12 | ✓✓✓✓   | ✓              | ✓              |
| M1Unet [229] | CVPR’21 | Metric                   | ResNet-18 | ✓✓✓✓   | ✓              | ✓              |
| Self-organizing Map [230] | NIPS’20 | Transfer Learning         | WRN      | ✓✓✓✓   | ✓              | ✓              |
| RCN [231] | Arxiv’20 | Metric                   | ResNet-12 | ✓✓✓✓   | 57.4%          | 75.1%          |
| MFS-distil [99] | ECCV’20 | Embedding                 | ResNet-12 | ✓✓✓✓   | 83.27%         | 86.46%         |
| inp-clipinset [232] | P-ResNet-18 | Metric                 | ResNet-18 | ✓✓✓✓   | 75.57%         | 84.73%         |
| Transductive CNAsPs [233] | WACV’20 | Meta                     | ResNet-18 | ✓✓✓✓   | ✓              | 79.9%          | 91.5%          |
| SJK [98] | Arxiv’20 | Embedding                 | ResNet-12 | ✓✓✓✓   | 67.04%         | 85.34%         |
| PT-SAP [234] | IJCNN’20 | Metric                   | WRN      | ✓✓✓✓   | 82.92%         | 88.82%         |
| FKAML [175] | CVPR’20 | Metric                   | ResNet-12 | ✓✓✓✓   | 63.70%         | 79.54%         |
| SIB [235] | KLR’20 | Meta                     | WRN      | ✓✓✓✓   | 70.61%         | 79.2%          |
| IC [64] | CVPR’20 | Embedding                 | ResNet-12 | ✓✓✓✓   | 69.66%         | 81.11%         |
| EPN [179] | ECCV’20 | Embedding                 | WRN      | ✓✓✓✓   | ✓              | -              | 88.05%         |
| SSL [236] | AIL’20 | Meta                     | 4-Block CNN | ✓✓✓✓   | 34.11%         | 83.87%         |
| MetaFun [237] | ICRL’20 | Meta                     | ResNet-18 | ✓✓✓✓   | 84.13%         | 80.82%         |
| taskLevelAug [238] | Arxiv’20 | Meta                     | ResNet-12 | ✓✓✓✓   | 65.38%         | 82.13%         |
| DK [146] | NIPS’20 | Meta                     | Bayesian model | ✓✓✓✓   | 62.96%         | 64%            |
| S2MAR [239] | WACV’20 | Embedding                 | ResNet-18 | ✓✓✓✓   | 64.93%         | 83.18%         |
| LTV [150] | NIPS’19 | Meta                     | ResNet-12 | ✓✓✓✓   | 70.1%          | 78.7%          |
| TapNet [240] | PMLR’19 | Meta                     | ResNet-12 | ✓✓✓✓   | 61.65%         | 76.36%         |
| ECNN [234] | CVPR’19 | Metric                   | CNN      | ✓✓✓✓   | -              | 76.37%         |
| ACC [241] | Arxiv’19 | Transfer Learning         | ResNet50 | ✓✓✓✓   | 62.21%         | 80.75%         |
| Tap [152] | CVPR’20 | Embedding                 | ResNet-18 | ✓✓✓✓   | 63.72%         | 78.38%         |
| DNN [242] | CVPR’19 | Meta                     | Conv-64  | ✓✓✓✓   | 51.24%         | 71.02%         |
| MIC2 [244] | NIPS’19 | Meta                     | WRN      | ✓✓✓✓   | 55.73%         | 70.33%         |
| PT Indicates Pre-train, FT Indicates Fine-tuning, DA means using Data Augmentation during the Train Stage or Test Stage. KD means knowledge distillation, SS means self-supervision and MTL means multitasking Learning. |

Table 13. The Overview of the State-of-the-arts on MS COCO Object Detection with Some Special Tricks

| Method | Venue | Backbone | Type | Tricks | mini-ImageNet | Available Code |
|--------|-------|----------|------|--------|---------------|----------------|
| mTRef- ViT-FB [247] | Arxiv’22 | ViT | Transfer learning | ✓✓✓✓ | ✓ | ✓ | 22.5AP | 30.2AP |
| Meta-DETR [19] | Arxiv’22 | ResNet-101 | Meta | ✓✓✓✓ | ✓ | ✓ | 17.8AP | 29.9AP |
| ESF [248] | CVPR’21 | ResNet-101 | Transfer learning | ✓✓✓✓ | ✓ | ✓ | 11.1AP | 13.5AP |
| SSF-FSL [249] | CVPR’21 | ResNet-101 | Multimodality | ✓✓✓✓ | ✓ | ✓ | 11.3AP | 14.7AP |
| FaDetNet View [250] | ECCV’20 | R-CNN | Meta | ✓✓✓✓ | ✓ | ✓ | 12.5AP | 14.7AP |
| MPSR [251] | ECCV’20 | ResNet-101 | Embedding | ✓✓✓✓ | ✓ | ✓ | 9.8AP | 14.1AP |
| TFA [252] | ICML’20 | ResNet-101 | Transfer learning | ✓✓✓✓ | ✓ | ✓ | 10AP | 13.7AP |
| Meta R-CNN [253] | ICCV’19 | R-CNN | Meta | ✓✓✓✓ | ✓ | ✓ | - | 12.4AP |
| MetaDet [254] | ICCV’19 | VGG16 | Meta | ✓✓✓✓ | ✓ | ✓ | 7.1AP | 11.3AP |

PT indicates pre-training, FT indicates fine-tuning, DA means using data augmentation during the train stage or test stage, AT means attention technology, SS means self-supervision, and MTL means multitasking learning.

7.4 Few-shot Instance Segmentation

Compared to semantic segmentation, FSL instance segmentation requires not only identifying each pixel in the image but also labeling the corresponding pixel. Recently, few works are dealing with the problem of segmenting a few samples of instances. In addition, some work presents how to
Fig. 18. Performance improvement progress of FSL semantic segmentation tasks during 2018–2021.

Fig. 19. Performance improvement progress of FSL instance segmentation tasks during 2017–2021.

Table 14. The Overview of the State-of-the-arts on COCO-20i Semantic Segmentation with Some Special Tricks

| Method       | Venue | Backbone | Type | Tricks | COCO-20i(MPA50) | Available Code |
|--------------|-------|----------|------|--------|-----------------|----------------|
| MSANet       | Arxiv'22 | ResNet-101 | Metric | ✓✓✓ | 69.13 73.99 | ✓ |
| CyCTR        | NIPS'21 | ResNet-50 | Metric | ✓✓✓ | 64.3 66.6 | ✓ |
| HSNNet       | ICCV'21 | ResNet-50 | Metric | ✓✓✓ | 66.2 70.4 | ✓ |
| RPMM         | ECCV'20 | ResNet-50 | Metric | ✓✓ | 56.3 - | ✓ |
| SVF          | Arxiv'22 | ResNet-50 | Transfer learning | ✓✓✓ | 60.8 - | ✓ |
| PPNet        | ECCV'20 | ResNet-50 | Metric | ✓✓✓ | 51.5 62.0 | ✓ |
| FWB          | ICCV'19 | ResNet-101 | Metric | ✓✓✓ | 56.2 59.9 | ✓ |
| PANet        | ICCV'19 | VGG16 | Metric | ✓✓✓ | 48.1 55.7 | ✓ |
| CANet        | CVPR'19 | ResNet | Metric | ✓✓✓✓ | 55.4 57.1 | ✓ |

PT indicates pre-training, FT indicates fine-tuning, DA means using data augmentation during the training stage or test stage, and AT means attention technology.

Table 15. The Overview of the State-of-the-arts on COCO-20i Instance Segmentation with Some Special Tricks

| Method       | Venue | Backbone | Type | Tricks | COCO-20i | Available Code |
|--------------|-------|----------|------|--------|----------|----------------|
| iMTFA        | CVPR'21 | R-CNN | Metric | ✓✓✓ | 20.13 18.22 | ✓ |
| FAPS         | CVPR'21 | ResNet-50 | Metric | ✓✓✓ | 16.3 18.2 | ✓ |
| FGN          | CVPR'20 | ResNet-101 | Meta | ✓✓✓ | 16.2 - | ✓ |
| SG-One       | IEEE'20 | VGG-16 | Metric | ✓✓✓ | 14.8 - | ✓ |
| Siamese Mask R-CNN | Arxiv'18 | ResNet-50 | Metric | ✓✓✓ | 14.5 - | ✓ |

PT indicates pre-training, FT indicates fine-tuning, DA means using data augmentation during the training stage or test stage, AT means attention technology, SS means self-supervision, and MTL means multitask learning.

improve RCNNs using some effective modules. The most representative work is Reference [264], which proposes an incremental few-shot instance segmentation algorithm that greatly improves the performance of the benchmark datasets. Table 15 and Figure 19 show the research progress of few-shot instance segmentation from 2019 to 2021.

8 FUTURE DIRECTION AND OPPORTUNITIES OF FSL

With a significant amount of work making promising progress on various task settings for FSL, the more challenging scenarios appear. For example, training and validation datasets are both
scarce, where there are no available external data and no similar domain training task or validation datasets. In meta-learning, it may not involve enough tasks to initialize the model parameters, and multimodal learning suffers from a distinct modal ordering problem. In line with the taxonomy, we present several possible future research directions corresponding to each level.

8.1 Better Evaluation of Data Distribution

Machine learning is difficult to train a model with excellent generalization from a very small number of samples. If the algorithms are trained on data with biases, then it will destroy the generalization of the model. Yang et al. [12] make the first meaningful attempt in this direction. It assumes that both the base class data and the new class data share a normal distribution. By calculating the mean and variance of the base class, the values can be transferred to the new class. If this assumption is sufficiently accurate, then it can bridge the gap between FSL and traditional machine learning to some extent. However, it is still a relatively strict assumption of the data distribution. If there is a large gap between the base class and the novel class, then sufficiently complex correction rules need to be explored using some assisted modules like a relation network. In the future, this is an exciting direction that considers relaxing the assumptions and exploring more generalization methods.

Another issue is that mainstream FSL benchmark datasets exist more or less some problems. For example, the mini-ImageNet dataset obtains some unsuitable samples that are too difficult for model evaluation, such as entity occlusion, and multiple objects. However, other simple datasets, such as Omniglot, cifar10, cifar100, and so on, have been largely addressed. FSL needs more challenging datasets like BSCD-FSL [22]. Likewise, natural language processing and multimodal field expect more benchmark datasets such as FEWCLUE and FEWGlue, which are closely integrated with hotspots of research and application. But until now, there is no benchmark dataset to evaluate a model’s ability to generalize at a fine-grained and full scene level.

8.2 Improving the Robustness of Data-to-label Mapping

The advent of the BSCD-FSL [22] benchmark brings a new challenge to FSL. It explores and reveals the limitations of current FSL for cross-domain learning. Recent research has produced skillfully designed models, more sophisticated hyperparameter tuning, additional auxiliary datasets, and extraction of domain-irrelevant features that are effective for FSL. Currently, fine-tuning is already performing very robustly at the intersection of transfer learning and meta-learning. Pre-training can be seen as learning many categories of tasks; it is single-task learning. Meta-learning, however, is a multi-task learning approach. It is worth exploring whether there is a better model that can integrate meta-learning and fine-tuning to maximize the performance of the model while reducing the computational complexity during meta-learning worth exploring.

Nowadays, the feature level still has a lot of potential for parameter space optimization. P > M > F [100] explores this simple pipeline including pre-training, meta-learning, and fine-tuning. It achieves the state-of-the-art in FSL classification task on mini-ImageNet. Subsequent work continues to experiment with replacing the pre-trained backbone, adding various tricks to the fine-tuning, or introducing multi-task to fully exploit the performance of fine-tuning and meta-learning.

8.3 Learn Meta-knowledge More Effectively from Historical Tasks

Meta-learning is a very general concept. Using meta-learning in combination with other methods can further improve the performance of FSL tasks. However, meta-learning is limited to specific task spaces under a defined network structure. In the case of classification tasks, only related classification tasks are currently considered. Does there exist a framework that takes into account classification, detection, prediction, and generation at the same time? This would enable meta-learning
separated from the conception of tasks to some extent. Recent work is attempting to optimize each batch as a whole. In this case, how to optimize the inner loop will be an important direction. In the future, the combination of meta-learning and fine-tuning will become the mainstream algorithms to address FSL.

Note that meta-learning is still exploring the correlation between tasks. No relevant theory has yet to emerge to explain the causal relationship behind meta-learning. In the future, meta-learning may tend to become a more general framework as the causal theory framework evolves.

8.4 Full Convergence of Multimodal Information

Multimodal learning is currently an emerging approach to address FSL problem, which automatically learns heterogeneous data from edge scenes without supervised labels and quickly transfers to different downstream tasks. Multimodal learning is widely regarded as an exploration from weak AI in limited domains to general AI. The advent of multimodal FSL has brought FSL into the era of pretrained multimodal models + small sample fine-tuning. Technically, the pre-trained multimodal model with a sufficient number of parameters can address any downstream tasks in terms of promotion. At this stage, researchers are encouraged to try more modal fusion learning, such as speech and video. Another direction is how to quantify the importance of each type of information given the fusion of multiple information so different weights can be obtained in the training stage.

9 CONCLUSION

As an important branch of machine learning, few-shot learning does not require a large amount of data but rather chooses a flexible approach to solve problems. By leveraging prior knowledge at different levels of abstraction, a wide range of work has emerged that attempts to address FSL issues. In this context, we provide a comprehensive survey of FSL in the form of questions and answers that easily distinguish confused conceptions and summarize the rich benchmark datasets under FSL. Besides, we provide unique insights into the challenges in the evolution of FSL following a new taxonomy. The relevant research methods are analyzed in depth according to the degree of integration of prior knowledge at each level. Furthermore, for the sake of completeness of the exposition, we also compare and analyze the recent advances of FSL in the field of computer vision. Finally, we present a list of possible future research directions and opportunities in light of the extensive recent literature. On the whole, this survey provides an overall comprehensive summary of the frontier advances in FSL over the past three years and is expected to contribute to the synergistic development of FSL and its related fields.

REFERENCES

[1] Wenbo Zheng, Lan Yan, Chao Gou, and Fei-Yue Wang. 2021. Federated meta-learning for fraudulent credit card detection. In Proceedings of the 29th International Conference on International Joint Conference on Artificial Intelligence. 4654–4660.

[2] Adrian El Baz, Ihsan Ullah, et al. 2022. Lessons learned from the NeurIPS 2021 MetaDL challenge: Backbone fine-tuning without episodic meta-learning dominates for few-shot learning image classification. In NeurIPS 2021 Competitions and Demonstrations Track. PMLR, 80–96.

[3] Yutai Hou, Xinghao Wang, Cheng Chen, Bohan Li, Wanxiang Che, and Zhigang Chen. 2022. FewJoint: Few-shot learning for joint dialogue understanding. Int. J. Mach. Learn. Cyber. (2022), 1–15.

[4] Huimin Sun, Jiajia Xu, Kai Zheng, Pengpeng Zhao, Pingfu Chao, and Xiaofang Zhou. 2021. MFNP: A meta-optimized model for few-shot next POI recommendation. In Proceedings of the International Joint Conference on Artificial Intelligence. 3017–3023.

[5] Elahe Rahimian, Soheil Zabihi, Amir Asif, S. Farokh Atashzar, and Arash Mohammadi. 2021. Few-shot learning for decoding surface electromyography for hand gesture recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 1300–1304.
A Comprehensive Survey of FSL: Evolution, Applications, Challenges, and Opportunities

[6] Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. ACM Comput. Surv. 53, 3 (2020), 1–34.

[7] Jun Shu, Zongben Xu, and Deyu Meng. 2018. Small sample learning in big data era. arXiv preprint arXiv:1808.04572 (2018).

[8] Jiang Lu, Pinghua Gong, Jieping Ye, and Changshui Zhang. 2020. Learning from very few samples: A survey. arXiv preprint arXiv:2009.02653 (2020).

[9] Endel Tulving. 1985. How many memory systems are there? Amer. Psychol. 40, 4 (1985), 385.

[10] Biqi Wang, Yang Xu, Zebin Wu, Tianming Zhan, and Zhilu Wei. 2022. Spatial-spectral local domain adaption for cross domain few shot hyperspectral images classification. IEEE Trans. Geosci. Remote Sens. 60 (2022), 1–15. DOI: http://dx.doi.org/10.1109/TGRS.2022.3208897

[11] Luiz F. O. Chamon and Alejandro Ribeiro. 2020. Probably approximately correct constrained learning. In Proceedings of the Annual Conference on Neural Information Processing Systems. Retrieved from https://proceedings.neurips.cc/paper/2020/hash/c291b01517f3e6797c774c306591cc32-Abstract.html.

[12] Ehtesham Iqbal, Sirojbek Safarov, and Seongdeok Bang. 2022. MSANet: Multi-similarity and attention guidance for few-shot learning. In Proceedings of the 9th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=JWOIYxMG92s.

[13] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one-shot learning. Adv. Neural Inf. Process. Syst. 29 (2016), 3630–3638.

[14] Technical Report.

[15] Biqi Wang, Yang Xu, Zebin Wu, Tianming Zhan, and Zhilu Wei. 2022. Spatial-spectral local domain adaption for cross domain few shot hyperspectral images classification. IEEE Trans. Geosci. Remote Sens. 60 (2022), 1–15. DOI: http://dx.doi.org/10.1109/TGRS.2022.3208897

[16] Luiz F. O. Chamon and Alejandro Ribeiro. 2020. Probably approximately correct constrained learning. In Proceedings of the Annual Conference on Neural Information Processing Systems. Retrieved from https://proceedings.neurips.cc/paper/2020/hash/c291b01517f3e6797c774c306591cc32-Abstract.html.

[17] Luo, Y., & Xiang, Y. (2022). FewCLUE: A Chinese few-shot learning evaluation benchmark. ArXiv Preprint ArXiv:2206.09667. (2022).

[18] Yang, J., Zheng, Y., & Liu, Y. (2021). Few-shot learning in big data era. arXiv Preprint ArXiv:2107.07498. (2021).

[19] Patrizio Bellan, Han van der Aa, Mauro Dragoni, Chiara Ghidini, and Simone Paolo Ponzetto. 2022. PET: A new dataset for process extraction from natural language text. arXiv e-prints (2022), arXiv–2203.

[20] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).
A Comprehensive Survey of FSL: Evolution, Applications, Challenges, and Opportunities 271:31

[54] Zitian Chen, Yanwei Fu, Yu-Xiong Wang, et al. 2019. Image deformation meta-networks for one-shot learning. In Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference. 8680–8689.

[55] Junjie Li, Zilei Wang, and Xiaoming Hu. 2021. Learning intact features by erasing-inpainting for few-shot classification. In Proceedings of the AAAI Conference on Artificial Intelligence. 8401–8409.

[56] Yuhang Huang, Fangle Chang, Yu Tao, Yangfan Zhao, Longhua Ma, and Hongye Su. 2022. Few-shot learning based on AttN-CutMix and task-adaptive transformer for the recognition of cotton growth stage. Comput. Electron. Agric. 202 (2022), 107406. DOI: http://dx.doi.org/10.1016/j.compag.2022.107406

[57] Erik G. Miller, Nicholas E. Matsakis, and Paul A. Viola. 2000. Learning from one example through shared densities on transforms. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 464–471.

[58] Bharath Hariharan and Ross Girshick. 2017. Low-shot visual recognition by shrinking and hallucinating features. In Proceedings of the IEEE International Conference on Computer Vision. 3018–3027.

[59] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V. Guttag, and Adrian V. Dalca. 2019. Data augmentation using learned transformations for one-shot medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8543–8553.

[60] Roland Kwitt, Sebastian Hegenbart, and Marc Niethammer. 2016. One-shot learning of scene locations via feature trajectory transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 78–86.

[61] Jing Zhou, Yanan Zheng, Jie Tang, Li Jian, and Zhilin Yang. 2022. FlipDA: Effective and robust data augmentation for few-shot learning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 8646–8665. DOI: http://dx.doi.org/10.18653/v1/2022.acl-long.592

[62] Tomas Pfister, James Charles, and Andrew Zisserman. 2014. Domain-adaptive discriminative one-shot learning of gestures. In Proceedings of the European Conference on Computer Vision. Springer, 814–829.

[63] Yu Wu, Yutian Lin, Xuanli Dong, Yan Yan, Wanli Ouyang, and Yi Yang. 2018. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5177–5186.

[64] Yikai Wang, Chengming Xu, Chen Liu, Li Zhang, and Yanwei Fu. 2020. Instance credibility inference for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 12836–12845.

[65] Xinze Li, Qianru Sun, Yaoao Liu, Qin Zhou, Shibao Zheng, Tat-Seng Chua, and Bernt Schiele. 2019. Learning to self-train for semi-supervised few-shot classification. Adv. Neural Inf. Process. Syst. 32 (2019), 10276–10286.

[66] Hang Gao, Zheng Shou, Alirez A Zareian, Hanwang Zhang, and Shih-Fu Chang. 2018. Low-shot learning via covariance-preserving adversarial augmentation networks. In Proceedings of the Annual Conference on Neural Information Processing Systems. 983–993. Retrieved from https://proceedings.neurips.cc/paper/2018/hash/81448138f5f163ccdba4acc69819f280-Abstract.html

[67] Jianhong Zhang, Manli Zhang, Zhiwu Lu, and Tao Xiang. 2021. AdarGCN: Adaptive aggregation GCN for few-shot learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 3482–3491.

[68] Haofeng Zhang, Li Liu, Yang Long, Zheng Zhang, and Ling Shao. 2020. Deep transductive network for generalized zero shot learning. Pattern Recog. 105 (2020), 107370. DOI: http://dx.doi.org/10.1016/j.patcog.2020.107370

[69] Ashraful Islam, Chun-Fu Richard Chen, Rameswar Panda, Leonid Karlinsky, Rogerio Feris, and Richard J. Radke. 2021. Dynamic distillation network for cross-domain few-shot recognition with unlabeled data. Adv. Neural Inf. Process. Syst. 34 (2021), 3584–3595.

[70] Amit Alfassy, Leonid Karlinsky, Amit Aides, Joseph Shtok, Sivan Harary, Rogerio Feris, Raja Giryes, and Alex M. Bronstein. 2019. LaSO: Label-set operations networks for multi-label few-shot learning. In Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference. 6548–6557.

[71] Zhiyong Huang, Fangle Chang, Yu Tao, Yangfan Zhao, Longhua Ma, and Hongye Su. 2022. Few-shot learning based on Attn-CutMix and task-adaptive transformer for the recognition of cotton growth state. Comput. Electron. Agric. 202 (2022), 107406. DOI: http://dx.doi.org/10.1016/j.compag.2022.107406

[72] Mengting Chen, Yuxin Fang, Xinggang Wang, Heng Luo, Yifeng Geng, Xinyu Zhang, Chang Huang, Wenyu Liu, and Zhiyong Dai, Jianjun Yi, Lei Yan, Qingwen Xu, Liang Hu, Qi Zhang, Jiahui Li, and Guoqiang Wang. 2023. PFEMed: Few-shot medical image classification using prior guided feature enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8580–8589.

[73] Roland Kwitt, Sebastian Hegenbart, and Marc Niethammer. 2016. One-shot learning of scene locations via feature trajectory transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 78–86.

[74] Jing Zhou, Yanan Zheng, Jie Tang, Li Jian, and Zhilin Yang. 2022. FlipDA: Effective and robust data augmentation for few-shot learning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 8646–8665. DOI: http://dx.doi.org/10.18653/v1/2022.acl-long.592

[75] Tomas Pfister, James Charles, and Andrew Zisserman. 2014. Domain-adaptive discriminative one-shot learning of gestures. In Proceedings of the European Conference on Computer Vision. Springer, 814–829.

[76] Yu Wu, Yutian Lin, Xuanli Dong, Yan Yan, Wanli Ouyang, and Yi Yang. 2018. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5177–5186.

[77] Roland Kwitt, Sebastian Hegenbart, and Marc Niethammer. 2016. One-shot learning of scene locations via feature trajectory transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 78–86.

[78] Jing Zhou, Yanan Zheng, Jie Tang, Li Jian, and Zhilin Yang. 2022. FlipDA: Effective and robust data augmentation for few-shot learning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 8646–8665. DOI: http://dx.doi.org/10.18653/v1/2022.acl-long.592

[79] Ehretd Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Abhishek Kumar, Rogério Schmidt Feris, Raja Giryes, and Alexander M. Bronstein. 2018. Delta-encoder: An effective sample synthesis method for few-shot object recognition. In Proceedings of the Annual Conference on Neural Information Processing Systems. 2850–2860. Retrieved from https://proceedings.neurips.cc/paper/2018/hash/1714726c817af50457d810ae94d27a2e-Abstract.html.

ACM Computing Surveys, Vol. 55, No. 13s, Article 271. Publication date: July 2023.
A Comprehensive Survey of FSL: Evolution, Applications, Challenges, and Opportunities

[99] Xilang Huang and Seon Han Choi. 2023. SAPENet: Self-attention based prototype enhancement network for few-shot learning. *Pattern Recog.* 135 (2023), 109170. DOI: https://doi.org/10.1016/j.patcog.2022.109170

[100] Shell Xu Hu, Da Li, Jan Stihmier, et al. 2022. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9068–9077.

[101] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100, 000+ questions for machine comprehension of text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 2383–2392. DOI: https://doi.org/10.18653/v1/d16-1264

[102] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30 (2017).

[103] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 7th International Conference on Learning Representations*. OpenReview.net. Retrieved from https://openreview.net/forum?id=rJ4km2R97.

[104] Jieh-Sheng Lee and Jieh Hsiang. 2020. Patent claim generation by fine-tuning OpenAI GPT-2. *World Patent Inf.* 62 (2020), 101983.

[105] Tianyi Zhang, Felix Wu, Arzoo Katiyar, Kilian Q. Weinberger, and Yoav Artzi. 2021. Revisiting few-sample BERT fine-tuning. In *Proceedings of the 9th International Conference on Learning Representations*. OpenReview.net. Retrieved from https://openreview.net/forum?id=cOlH43yUF.

[106] Xu Luo, Longhui Wei, Liangjian Wen, Jinrong Yang, Lingxi Xie, Zenglin Xu, and Qi Tian. 2021. Rectifying the shortcut learning of background for few-shot learning. *Adv. Neural Inf. Process. Syst.* 34 (2021), 13073–13085.

[107] Leyang Cui, Yu Wu, Jian Liu, Sen Yang, and Yue Zhang. 2021. Template-based named entity recognition using BART. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP (Findings of ACL)*. Association for Computational Linguistics, 1835–1845. DOI: https://doi.org/10.18653/v1/2021.findings-acl.161

[108] Xiang Li Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*. ACL, 4582–4597. DOI: https://doi.org/10.18653/v1/2021.acl-long.353

[109] Ning Ding, Shengding Hu, Weilin Zhao, Yulin Chen, Zhiyuan Liu, Haitao Zheng, and Maosong Sun. 2022. OpenPrompt: An open-source framework for prompt-learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 105–113. DOI: https://doi.org/10.18653/v1/2022.acl-demo.10

[110] Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 3045–3059. DOI: https://doi.org/10.18653/v1/2021.emnlp-main.243

[111] Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066* (2019).

[112] Timo Schick and Hinrich Schütze. 2021. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. ACL, 255–269. DOI: https://doi.org/10.18653/v1/2021.eacl-main.20

[113] Zhenghao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. 2020. How can we know what language models know? *Trans. Assoc. Computat. Ling.* 8 (2020), 423–438.

[114] T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace, and S. Singh. 2020. AutoPrompt: Eliciting knowledge from language models with automatically generated prompts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 4222–4235. DOI: https://doi.org/10.18653/v1/2020.emnlp-main.346

[115] Tianyu Gao, Adam Fisch, and Danqi Chen. 2021. Making pre-trained language models better few-shot learners. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*. 3816–3830. DOI: https://doi.org/10.18653/v1/2021.acl-long.295

[116] Ziyun Xu, Chengyu Wang, Minghui Qiu, Fuli Luo, Runxin Xu, Songfang Huang, and Jun Huang. 2022. Making pre-trained language models end-to-end few-shot learners with contrastive prompt tuning. *arXiv preprint arXiv:2204.00166* (2022).

[117] Zexuan Zhong, Dan Friedman, and Danqi Chen. 2021. Factual probing is [MASK]: Learning vs. learning to recall. *arXiv preprint arXiv:2104.05240* (2021).

[118] Xiao Liu, Yanan Zheng, Zhengxiao Du, et al. 2021. GPT understands, too. *arXiv preprint arXiv:2103.10385* (2021).

[119] Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. 2022. PPT: Pre-trained prompt tuning for few-shot learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 8410–8423. DOI: https://doi.org/10.18653/v1/2022.acl-long.576

[120] Wang Yan, Jordan Yap, and Greg Mori. 2015. Multi-task transfer methods to improve one-shot learning for multimedia event detection. In *Proceedings of the British Machine Vision Conference*. 37–1.
[121] Saeid Motiian, Quinn Jones, Seyed Iranmanesh, and Gianfranco Doretto. 2017. Few-shot adversarial domain adaptation. Adv. Neural Inf. Process. Syst. 30 (2017).

[122] Sagie Benaim and Lior Wolf. 2018. One-shot unsupervised cross domain translation. Adv. Neural Inf. Process. Syst. 31 (2018).

[123] Samaneh Azadi, Matthew Fisher, Vladimir G. Kim, Zhaowen Wang, Eli Shechtman, and Trevor Darrell. 2018. Multi-content GAN for few-shot font style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7564–7573.

[124] Bo Liu, Xudong Wang, Mandar Dixit, Roland Kwitt, and Nuno Vasconcelos. 2018. Feature space transfer for data augmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 9090–9098.

[125] Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and MiaoSong Sun. 2018. Few-shot charge prediction with discriminative legal attributes. In Proceedings of the 27th International Conference on Computational Linguistics. 487–498.

[126] Yabin Zhang, Hui Tang, and Kui Jia. 2018. Fine-grained visual categorization using meta-learning optimization with sample selection of auxiliary data. In Proceedings of the European Conference on Computer Vision (ECCV). 233–248.

[127] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Perez, and Matthieu Cord. 2019. Boosting few-shot visual learning with self-supervision. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 8059–8068.

[128] R. Zhang, P. Isola, and A. A. Efros. 2016. Colorful image colorization. In Proceedings of the European Conference on Computer Vision. Springer, 649–666.

[129] Jong-Chyi Su, Subhransu Maji, and Bharath Hariharan. 2020. When does self-supervision improve few-shot learning? In Proceedings of the European Conference on Computer Vision. Springer, 645–666.

[130] Zijun Li, Zhengping Hu, Weimei Luo, and Xiao Hu. 2023. SaberNet: Self-attention based effective relation network for few-shot learning. Pattern Recog. 133 (2023), 109024. DOI: http://dx.doi.org/10.1016/j.patcog.2022.109024

[131] Rui Feng, Hongbing Ji, Zhigang Zhu, and Lei Wang. 2022. SelNet: A semi-supervised local Fisher discriminant network for few-shot learning. Neurocomputing 512 (2022), 352–362. DOI: http://dx.doi.org/10.1016/j.neucom.2022.09.012

[132] Zhengyu Chen, Jixie Ge, Heshen Zhan, Siteng Huang, and Donglin Wang. 2021. Pareto self-supervised training for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 13663–13672.

[133] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning. PMLR, 1126–1135.

[134] Aravind Rajeswaran, Chelsea Finn, Sham M. Kakade, and Sergey Levine. 2019. Meta-learning with implicit gradients. Adv. Neural Inf. Process. Syst. 32 (2019).

[135] Xiaomeng Zhu and Shuxiao Li. 2022. MGML: Momentum group meta-learning for few-shot image classification. Neurocomputing 514 (2022), 351–361. DOI: http://dx.doi.org/10.1016/j.neucom.2022.10.012

[136] Alex Nichol, Joshua Achiam, and John Schulman. 2018. On first-order meta-learning algorithms. arXiv preprint arXiv:1803.02999 (2018).

[137] Alex Nichol and John Schulman. 2018. Reptile: A scalable metalearning algorithm. arXiv preprint arXiv:1803.02999 2, 3 (2018), 4.

[138] Ondrej Bohdal, Yongxin Yang, and Timothy Hospedales. 2021. EvoGrad: Efficient gradient-based meta-learning and hyperparameter optimization. Adv. Neural Inf. Process. Syst. 34 (2021), 22234–22246.

[139] Xingyou Song, Wenbo Gao, Yuxiang Yang, Krzysztof Choromanski, Aldo Pacchiano, and Yunhao Tang. 2019. ES-MAML: Simple Hessian-free meta learning. arXiv preprint arXiv:1910.01215 (2019).

[140] Z. Li, F. Zhou, F. Chen, and H. Li. 2017. Meta-SGD: Learning to learn quickly for few-shot learning. arXiv preprint arXiv:1707.09835 (2017).

[141] Yiluan Guo and Ngai-Man Cheung. 2020. Attentive weights generation for few shot learning via information maximization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 13499–13508.

[142] Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. 2019. Meta-learning with latent embedding optimization. In Proceedings of the International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=BJgklhAcK7.

[143] Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas L. Griffiths. 2018. Recasting gradient-based meta-learning as hierarchical Bayes. In Proceedings of the International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=BJ_UL-kb.

[144] C. Finn, K. Xu, and S. Levine. 2018. Probabilistic model-agnostic meta-learning. Adv. Neural Inf. Process. Syst. 31 (2018).

[145] Jaesik Yoon, Taesup Kim, Ousmane Dia, Sungwoong Kim, Yoshua Bengio, and Sungjin Ahn. 2018. Bayesian model-agnostic meta-learning. Adv. Neural Inf. Process. Syst. 31 (2018).

[146] Massimiliano Patacchiola, Jack Turner, Elliot J. Crowley, Michael O’Boyle, and Amos Storkey. 2020. Bayesian meta-learning for the few-shot setting via deep kernels. (2020).
A Comprehensive Survey of FSL: Evolution, Applications, Challenges, and Opportunities  271:35

[147] Sachin Ravi and Alex Beatson. 2018. Amortized Bayesian meta-learning. In Proceedings of the International Conference on Learning Representations.

[148] Yoonho Lee and Seungjiin Choi. 2018. Gradient-based meta-learning with learned layerwise metric and subspace. In Proceedings of the International Conference on Machine Learning. PMLR, 2927–2936.

[149] Muhammad Abdullah Jamal and Guo-Jun Qi. 2019. Task agnostic meta-learning for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 11719–11727.

[150] Taewon Jeong and Heeyoung Kim. 2020. OOD-MAML: Meta-learning for few-shot out-of-distribution detection and classification. Adv. Neural Inf. Process. Syst. 33 (2020), 3907–3916.

[151] Huaxiu Yao, Xian Wu, Zhiqiang Tao, Yaliang Li, Bolin Ding, Ruirui Li, and Zhenhui Li. [n. d.]. Automated relational meta-learning. ([n. d.]).

[152] Sebastian Fltenerhag, Andrei A. Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, and Raia Hadsell. 2019. Meta-learning with warped gradient descent. arXiv preprint arXiv:1909.00025 (2019).

[153] Muhammad Abdullah Jamal, Liqiang Wang, and Boqing Gong. 2021. A lazy approach to long-horizon gradient-based meta-learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 6577–6586.

[154] Han-Jia Ye and Wei-Lun Chao. 2022. How to train your MAML to excel in few-shot classification. In Proceedings of the 10th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=49h_lkpJtAE.

[155] Arnav Chavan, Rishabh Tiwari, Udbhav Bamba, and Deepak K. Gupta. 2022. Dynamic kernel selection for improved generalization and memory efficiency in meta-learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9851–9860.

[156] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. 2016. Meta-learning with memory-augmented neural networks. In Proceedings of the International Conference on Machine Learning. PMLR, 1842–1850.

[157] Lukasz Kaiser, Ofir Nachum, Aurko Roy, and Samy Bengio. 2017. Learning to remember rare events. In Proceedings of the 5th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=SJTQLdqlg.

[158] Tiago Ramalho and Marta Garnelo. 2019. Adaptive posterior learning: Few-shot learning with a surprise-based memory module. In Proceedings of the 7th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=Bye5dsC9Km.

[159] Ruiying Geng, Binhua Li, Yongbin Li, Jian Sun, and Xiaodan Zhu. 2020. Dynamic memory induction networks for few-shot text classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 1087–1094. DOI: http://dx.doi.org/10.18653/v1/2020.acl-main.102

[160] Xuncheng Liu, Xudong Tian, Shaohui Lin, Yanyun Qu, Lizhuang Ma, Wang Yuan, Zhizhong Zhang, and Yuan Xie. 2021. Learn from concepts: Towards the purified memory for few-shot learning. In Proceedings of the International Joint Conference on Artificial Intelligence. 888–894.

[161] Tianqin Li, Zijie Li, Andrew Luo, Harold Rockwell, Amir Barati Farimani, and Tai Sing Lee. 2021. Prototype memory and attention mechanisms for few shot image generation. In Proceedings of the International Conference on Learning Representations.

[162] Ying-Jun Du, Xiantong Zhen, Ling Shao, and Cees G. M. Snoek. 2020. Hierarchical variational memory for few-shot learning across domains. In Proceedings of the 10th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=i3RI65sR7N.

[163] Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, and Jeff Dean. 2018. Efficient neural architecture search via parameters sharing. In Proceedings of the International Conference on Machine Learning. PMLR, 4095–4104.

[164] Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, and Quoc Le. 2018. Understanding and simplifying one-shot architecture search. In Proceedings of the International Conference on Machine Learning. PMLR, 550–559.

[165] Yiyang Zhao, Linnan Wang, Yuandong Tian, Rodrigo Fonseca, and Tian Guo. 2021. Few-shot neural architecture search. In Proceedings of the International Conference on Machine Learning. PMLR, 12707–12718.

[166] Thomas Elsken, Benedikt Staffler, Jan Hendrik Metzen, and Frank Hutter. 2020. Meta-learning of neural architectures for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 12365–12375.

[167] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2018. DARTS: Differentiable architecture search. arXiv preprint arXiv:1806.09055 (2018).

[168] Jorge Gonzalez-Zapata, Ivan Reyes-Amezcua, Daniel Flores-Araiza, Mauricio Mendez-Ruiz, Gilberto Ochoa-Ruiz, and Andres Mendez-Vazquez. 2022. Guided deep metric learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1481–1489.

[169] You Zhou, Changlin Chen, and Shukun Ma. 2021. Few-shot ship classification based on metric learning. Multim. Syst. (2021), 1–10.
[170] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. 2015. Siamese neural networks for one-shot image recognition. In ICML Deep Learning Workshop, Vol. 2. Lille.

[171] Elad Hoffer and Nir Ailon. 2015. Deep metric learning using triplet network. In Proceedings of the International Workshop on Similarity-based Pattern Recognition. Springer, 84–92.

[172] Xiaomeng Li, Lequan Yu, Chi-Wing Fu, Meng Fang, and Pheng-Ann Heng. 2020. Revisiting metric learning for few-shot image classification. Neurocomputing 406 (2020), 49–58.

[173] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. 2017. Beyond triplet loss: A deep quadruplet network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 403–412.

[174] Jake Snell, Kevin Swersky, and Richard S. Zemel. 2017. Prototypical networks for few-shot learning. In Proceedings of the International Conference on Neural Information Processing Systems. 4077–4087. Retrieved from https://proceedings.neurips.cc/paper/2017/hash/cb8da6767461f2812ae4290eac7cbe42-Abstract.html.

[175] Aoxue Li, Weiran Huang, Xu Lan, Jiashi Feng, Zhenguo Li, and Liwei Wang. 2020. Boosting few-shot learning with adaptive margin loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 12576–12584.

[176] Bin Liu, Yue Cao, Yutong Lin, Qi Li, Zheng Zhang, Mingsheng Long, and Han Hu. 2020. Negative margin matters: Understanding margin in few-shot classification. In Proceedings of the European Conference on Computer Vision. Springer, 438–455.

[177] Xiaoxu Li, Yalan Li, Yixiao Zheng, Rui Zhu, Zhanyu Ma, Jing-Hao Xue, and Jie Cao. 2023. ReNAP: Relation network with adaptive prototypical learning for few-shot classification. Neurocomputing 520 (2023), 356–364. DOI: http://dx.doi.org/10.1016/j.neucom.2022.11.082

[178] Kaidi Cao, Maria Brbic, and Jure Leskovec. 2021. Concept learners for few-shot learning. In Proceedings of the 9th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=eJIJF3-LoZO.

[179] Pau Rodriguez, Issam Laradji, Alexandre Drouin, and Alexandre Lacoste. 2020. Embedding propagation: Smoother manifold for few-shot classification. In Proceedings of the European Conference on Computer Vision. Springer, 121–138.

[180] Zhong Ji, Xiyao Liu, Yanwei Pang, and Xuelong Li. 2020. SGAP-Net: Semantic-guided attentive prototypes network for few-shot human-object interaction recognition. In Proceedings of the AAAI Conference on Artificial Intelligence. 11085–11092.

[181] Fangyu Wu, Jeremy S. Smith, Wenjin Lu, Chaoyi Pang, and Bailing Zhang. 2020. Attentive prototype few-shot learning with capsule network-based embedding. In Proceedings of the European Conference on Computer Vision. Springer, 237–253.

[182] Van N. Nguyen, Sigurd Lakse, Kristoffer Wickstrom, Michael Kampffmeyer, Davide Roverso, and Robert Jenssen. 2020. SEN: A novel feature normalization dissimilarity measure for prototypical few-shot learning networks. In Proceedings of the European Conference on Computer Vision. Springer, 118–134.

[183] Wanqi Xue and Wei Wang. 2020. One-shot image classification by learning to restore prototypes. In Proceedings of the AAAI Conference on Artificial Intelligence. 6558–6565.

[184] Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. 2019. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In Proceedings of the AAAI Conference on Artificial Intelligence. 6407–6414.

[185] Philip Chikontwe, Soopil Kim, and Sang Hyun Park. 2022. CAD: Co-adapting discriminative features for improved few-shot classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 14554–14563.

[186] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. 2020. DeepEMD: Few-shot image classification with differentiable earth mover’s distance and structured classifiers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 12203–12213.

[187] Yidan Nie, Chunjiang Bian, and Ligang Li. 2022. Adap-EMD: Adaptive EMD for aircraft fine-grained classification in remote sensing. IEEE Geosci. Remote Sens. Lett. 19 (2022), 1–5.

[188] Jiangtao Xie, Fei Long, Jiaming Lv, Qilong Wang, and Peihua Li. 2022. Joint distribution matters: Deep Brownian distance covariance for few-shot classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7972–7981.

[189] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H. S. Torr, and Timothy M. Hospedales. 2018. Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1199–1208.

[190] Xiaoxu Li, Jijie Wu, Zhuo Sun, Zhanyu Ma, Jie Cao, and Jing-Hao Xue. 2020. BSNet: Bi-similarity network for few-shot fine-grained image classification. IEEE Trans. Image Process. 30 (2020), 1318–1331.

[191] Fusheng Hao, Fengxiang He, Jun Cheng, Lei Wang, Jianzhong Cao, and Dacheng Tao. 2019. Collect and select: Semantic alignment metric learning for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 8460–8469.
A Comprehensive Survey of FSL: Evolution, Applications, Challenges, and Opportunities

[192] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S. Yu Philip. 2020. A comprehensive survey on graph neural networks. IEEE Trans. Neural Netw. Learn. Syst. 32, 1 (2020), 4–24.

[193] Victor Garcia Satorras and Joan Bruna Estrach. 2018. Few-shot learning with graph neural networks. In Proceedings of the 6th International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=BJj6qGrRW.

[194] Jongmin Kim, Taesup Kim, Sungwoong Kim, and Chang D. Yoo. 2019. Edge-labeling graph neural network for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 11–20.

[195] Avishek Joey Bose, Ankit Jain, Piero Molino, and William L. Hamilton. 2019. Meta-graph: Few shot link prediction via meta learning. arXiv preprint arXiv:1912.09867 (2019).

[196] Yuqing Ma, Shihao Bai, Shan An, Wei Liu, Aishan Liu, Xiantong Zhen, and Xianglong Liu. 2020. Transductive relation-propagation network for few-shot learning. In Proceedings of the International Joint Conference on Artificial Intelligence. 804–810.

[197] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. 2019. Hypergraph neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence. 3558–3565.

[198] Ling Yang, Liangliang Li, Zilun Zhang, Xinyu Zhou, Erjin Zhou, and Yu Liu. 2020. DPGN: Distribution propagation graph network for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 13390–13399.

[199] Huihai Shao, Dexing Zhong, Xuefeng Du, Shaoyi Du, and Raymond N. J. Veldhuis. 2021. Few-shot learning for palmprint recognition via meta-Siamese network. IEEE Trans. Instrum. Measur. 70 (2021), 1–12.

[200] Thomas Müller, Guillermo Pérez-Torró, and Marc Franco-Salvador. 2022. Few-shot learning with siamese networks and label tuning. arXiv preprint arXiv:2203.14655 (2022).

[201] Congying Xia, Caiming Xiong, and Philip Yu. 2021. Pseudo siamese network for few-shot intent generation. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2005–2009.

[202] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. 2019. PANet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 9197–9206.

[203] Baoquan Zhang, Xutao Li, Shanshan Feng, Yunning Ye, and Rui Ye. 2022. MetaNODE: Prototype optimization as a neural ODE for few-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence. 9014–9021.

[204] Yinbo Chen, Zhuang Liu, Hujuan Xu, Trevor Darrell, and Xiaolong Wang. 2021. Meta-baseline: Exploring simple meta-learning for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 9042–9051. DOI: http://dx.doi.org/10.1109/ICCV48922.2021.00893.

[205] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning. PMLR, 8748–8763.

[206] Edgar Schonfeld, Sayna Ebrahimi, Samarath Sinha, Trevor Darrell, and Zeynep Akata. 2019. Generalized zero-and few-shot learning via aligned variational autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8247–8255.

[207] Shuo Wang, Jun Yue, Jianzhuan Liu, Qi Tian, and Meng Wang. 2020. Large-scale few-shot learning via multi-modal knowledge discovery. In Proceedings of the European Conference on Computer Vision. Springer, 718–734.

[208] Aoxue Li, Tiange Luo, Zhiwu Lu, Tao Xiang, and Liwei Wang. 2019. Large-scale few-shot learning: Knowledge transfer with class hierarchy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7212–7220.

[209] Eli Schwartz, Leonid Karlinsky, Rogerio Feris, Raja Giryes, and Alex Bronstein. 2022. Baby steps towards few-shot learning with multiple semantics. Pattern Recog. Lett. 160 (2022), 142–147.

[210] Zhiqiao Peng, Zechao Li, Junge Zhang, Yan Li, Guo-Jun Qi, and Jinhu Tang. 2019. Few-shot image recognition with knowledge transfer. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 441–449.

[211] Chen Xing, Negar Rostamzadeh, Boris Oreshkin, and Pedro O. O. Pinheiro. 2019. Adaptive cross-modal few-shot learning. Adv. Neural Inf. Process. Syst. 32 (2019), 4847–4857.

[212] Frederik Pahde, Mihai Puscas, Tassilo Klein, and Moin Nabi. 2021. Multimodal prototypical networks for few-shot learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2644–2653.

[213] Mathieu Pagé Fortin and Braham Chaib-draa. 2021. Towards contextual learning in few-shot object classification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 3279–3288.

[214] Michael Sollami and Aashish Jain. 2021. Multimodal conditionality for natural language generation. arXiv preprint arXiv:2109.01229 (2021).

[215] Mengmeng Wang, Jiacheng Xing, and Yong Liu. 2021. ActionCLIP: A new paradigm for video action recognition. arXiv preprint arXiv:2109.08472 (2021).
[216] Dongxu Li, Junnan Li, Hongdong Liu, Juan Carlos Niebles, and Steven C. H. Hoi. 2022. Align and prompt: Video-and-language pre-training with entity prompts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 4953–4963.

[217] Yuan Yao, Ao Zhang, Zhengyan Zhang, Zhiyuan Liu, Tat-Seng Chua, and Maosong Sun. 2021. CPT: Colorful prompt tuning for pre-trained vision-language models. arXiv preprint arXiv:2109.11797 (2021).

[218] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. 2022. Learning to prompt for vision-language models. Int. J. Comput. Vis. (2022), 1–12.

[219] Maria Tsimpoukelli, Jacob L. Menick, Serkan Cabi, S. M. Eslami, Oriol Vinyals, and Felix Hill. 2021. Multimodal few-shot learning with frozen language models. Adv. Neural Inf. Process. Syst. 34 (2021), 200–212.

[220] Daniel Shalam and Simon Korman. 2022. The self-optimal-transport feature transform. arXiv e-prints (2022), arXiv–2204.

[221] Dong Hoon Lee and Saé-Young Chung. 2021. Unsupervised embedding adaptation via early-stage feature reconstruction for few-shot classification. arXiv preprint arXiv:2106.11486 (2021).

[222] Haoxiang Wang, Han Zhao, and Bo Li. 2021. Bridging multi-task learning and meta-learning: Towards efficient training and effective adaptation. In Proceedings of the 38th International Conference on Machine Learning. PMLR, 10991–11002. Retrieved from http://proceedings.mlr.press/v139/wang21ad.html.

[223] Mohamed Afham, Salman Khan, Muhammad Haris Khan, Muzammal Naseer, and Fahad Shahbaz Khan. 2022. Enhancing few-shot image classification with unlabelled examples. In Proceedings of the International Conference on Pattern Recognition. IEEE, 4765–4771. DOI: http://dx.doi.org/10.1109/ICPR56361.2022.9955637

[224] Mamshad Nayeem Rizve, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. 2021. Exploring complementary strengths of invariant and equivariant representations for few-shot learning. In Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference. 10836–10846.

[225] Reza Esfandiarpoor, Amy Pu, Mohsen Hajabdollahi, and Stephen H. Bach. 2020. Extended few-shot learning: Exploiting existing resources for novel tasks. arXiv preprint arXiv:2012.07176 (2020).

[226] Haoxing Chen, Huaxiong Li, Yaohui Li, and Chunlin Chen. 2022. Multi-scale adaptive task attention network for few-shot learning. In Proceedings of the 26th International Conference on Pattern Recognition. IEEE, 4765–4771. DOI: http://dx.doi.org/10.1109/ICPR56361.2022.9955637

[227] Da Chen, Yuefeng Chen, Yuhong Li, Feng Mao, Yuan He, and Hui Xue. 2021. Self-supervised learning for few-shot image classification. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 1745–1749.

[228] Xian Zhong, Cheng Gu, Wenxin Huang, Lin Li, Shuqin Chen, and Chia-Wen Lin. 2021. Complementing representation deficiency in few-shot image classification: A meta-learning approach. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR). IEEE, 2677–2684.

[229] Bowen Wang, Liangzhi Li, Manisha Verma, Yuta Nakashima, Ryo Kawasaki, and Hajime Nagahara. 2020. Match them up: Visually explainable few-shot image classification. arXiv preprint arXiv:2011.12527 (2020).

[230] Lyes Khacef, Vincent Gripon, and Benoit Miramond. 2020. GPU-based self-organizing maps for post-labeled few-shot unsupervised learning. In Proceedings of the International Conference on Neural Information Processing. Springer, 404–416.

[231] Zhiyu Xue, Lixin Duan, Wen Li, Lin Chen, and Jiebo Luo. 2020. Region comparison network for interpretable few-shot image classification. arXiv preprint arXiv:2009.03558 (2020).

[232] Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed. 2020. Laplacian regularized few-shot learning. In Proceedings of the International Conference on Machine Learning. PMLR, 11660–11670.

[233] Peyman Bateni, Jarred Barber, Jan-Willem van de Meent, and Frank Wood. 2022. Enhancing few-shot image classification with unlabelled examples. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 1597–1606. DOI: http://dx.doi.org/10.1109/WACV51458.2022.00166

[234] Yuqing Hu, Vincent Gripon, and Stéphane Pateux. 2021. Leveraging the feature distribution in transfer-based few-shot learning. In Artificial Neural Networks and Machine Learning - 30th International Conference on Artificial Neural Networks, Vol. 12892. Springer, 487–499. DOI: http://dx.doi.org/10.1007/978-3-030-86340-1_39

[235] Shell Xu Hu, Pablo Garcia Moreno, Yang Xiao, Xi Shen, Guillaume Obozinski, Neil D. Lawrence, and Andreas C. Damianou. 2020. Empirical Bayes transductive meta-learning with synthetic gradients. In Proceedings of the International Conference on Learning Representations. OpenReview.net. Retrieved from https://openreview.net/forum?id=Hkg-xgr1VH.

[236] Cuong Nguyen, Thanh-Toan Do, and Gustavo Carneiro. 2020. PAC-Bayesian meta-learning with implicit prior and posterior. arXiv preprint arXiv:2003.02453 (2020).

[237] Jin Xu, Jean-Francois Ton, Hyunjik Kim, Adam Kosiorek, and Yee Whye Teh. 2020. MetaFun: Meta-learning with iterative functional updates. In Proceedings of the International Conference on Machine Learning. PMLR, 10617–10627.
[238] Jialin Liu, Fei Chao, and Chih-Min Lin. 2020. Task adaptation by rotating for meta-learning. arXiv preprint arXiv:2003.00804 (2020).

[239] Punit Mangla, Nupur Kumari, Abhishek Sinha, et al. 2020. Charting the right manifold: Manifold mixup for few-shot learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2218–2227.

[240] Sung Whan Yoon, Jun Seo, and Jaekyun Moon. 2019. TapNet: Neural network augmented with task-adaptive projection for few-shot learning. In Proceedings of the International Conference on Machine Learning. PMLR, 7115–7123.

[241] Liang Song, Jilin Liu, and Yongqiang Qin. 2019. Generalized adaptation for few-shot learning. arXiv preprint arXiv:1911.10807 (2019).

[242] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. 2020. Few-shot learning via embedding adaptation with set-to-set functions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8808–8817.

[243] Wenhui Li, Lei Wang, Jinglin Xu, Jing Huo, Yang Gao, and Jiebo Luo. 2019. Revisiting local descriptor based image-to-class measure for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7260–7268.

[244] Eunbyung Park and Junier B. Oliva. 2019. Meta-curvature. In Proceedings of the Annual Conference on Neural Information Processing Systems. 3309–3319. Retrieved from https://proceedings.neurips.cc/paper/2019/hash/57c0531e13f40b91b3b0f1a30b529a1d-Abstract.html.

[245] Ross Girshick. 2015. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision. 1440–1448.

[246] Xingkui Zhu, Shuchang Lyu, Xu Wang, and Qi Zhao. 2021. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2778–2788.

[247] XiaoSong Zhang, Feng Liu, Zhiliang Peng, Zonghao Guo, Fang Wan, Xiangyang Ji, and Qixiang Ye. 2022. Integral migrating pre-trained transformer encoder-decoders for visual object detection. arXiv preprint arXiv:2205.09613 (2022).

[248] Bo Sun, Banghui Li, Shengcai Cai, Ye Yuan, and Chi Zhang. 2021. FSCE: Few-shot object detection via contrastive proposal encoding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7352–7362.

[249] Chenchen Zhu, Fangyi Chen, Uzair Ahmed, Zhiqiang Shen, and Marios Savvides. 2021. Semantic relation reasoning for shot-stable few-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8782–8791.

[250] Yang Xiao and Renaud Marlet. 2020. Few-shot object detection and viewpoint estimation for objects in the wild. In Proceedings of the European Conference on Computer Vision. Springer, 192–210.

[251] Jiaxi Wu, Songtao Liu, Di Huang, and Yunhong Wang. 2020. Multi-scale positive sample refinement for few-shot object detection. In Proceedings of the European Conference on Computer Vision. Springer, 456–472.

[252] Xin Wang, Thomas E. Huang, Joseph Gonzalez, Trevor Darrell, and Fisher Yu. 2020. Frustratingly simple few-shot object detection. In Proceedings of the 37th International Conference on Machine Learning. PMLR, 9919–9928. Retrieved from http://proceedings.mlr.press/v119/wang20j.html.

[253] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. 2019. Meta R-CNN: Towards general solver for instance-level low-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 9577–9586.

[254] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. 2019. Meta-learning to detect rare objects. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 9925–9934.

[255] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. 2017. One-shot learning for semantic segmentation. In Proceedings of the British Machine Vision Conference. BMVA Press. Retrieved from https://www.dropbox.com/s/1odhw88465klz/0797.pdf?id=1.

[256] Zhihe Lu, Sen He, Xiatian Zhu, Yi Zhang, Yi-Zhe Song, and Tao Xiang. 2020. Prior guided feature enrichment network for few-shot segmentation. In Proceedings of the Annual Conference on Neural Information Processing Systems. 21984–21996. Retrieved from https://proceedings.neurips.cc/paper/2021/hash/bbb12f949378552c21f28de0fba8eb6-Abstract.html.

[257] Juhong Min, Dahyun Kang, and Minsu Cho. 2020. Hypercorrelation squeeze for few-shot segmentation. arXiv preprint arXiv:2104.01538 (2021).

[258] Boyu Yang, Chang Liu, Bofah Li, Jianbin Jiao, and Qixiang Ye. 2020. Prototype mixture models for few-shot semantic segmentation. In Proceedings of the European Conference on Computer Vision. Springer, 763–778.

[259] Zhoutao Tian, Hengshuang Zhao, Michelle Shu, Zhicheng Yang, Ruizhe Li, and Jiaya Jia. 2020. Prior guided feature enrichment network for few-shot segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 01 (2020), 1–1.

[260] Yongfei Liu, Xiangyi Zhang, Songyang Zhang, and Xuming He. 2020. Part-aware prototype network for few-shot semantic segmentation. In Proceedings of the European Conference on Computer Vision. Springer, 142–158.
[262] Khoi Nguyen and Sinisa Todorovic. 2019. Feature weighting and boosting for few-shot segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 622–631.

[263] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. 2019. CANet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5217–5226.

[264] Dan Andrei Ganea, Bas Boom, and Ronald Poppe. 2021. Incremental few-shot instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1185–1194.

[265] Khoi Nguyen and Sinisa Todorovic. 2021. FAPIS: A few-shot anchor-free part-based instance segmenter. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 11099–11108.

[266] Zhibo Fan, Jin-Gang Yu, Zhihao Liang, Jiarong Ou, Changxin Gao, Gui-Song Xia, and Yuanqing Li. 2020. FGN: Fully guided network for few-shot instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9172–9181.

[267] Xiaolin Zhang, Yunchao Wei, Yi Yang, and Thomas S. Huang. 2020. SG-One: Similarity guidance network for one-shot semantic segmentation. IEEE Trans. Cyber. 50, 9 (2020), 3855–3865.

[268] Claudio Michaelis, Ivan Ustyuzhaninov, Matthias Bethge, et al. 2018. One-shot instance segmentation. arXiv preprint arXiv:1811.11507 (2018).

Received 8 February 2022; revised 16 January 2023; accepted 23 January 2023