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

Yisheng Song, Ting Wang*, Subrota K Mondal, Jyoti Prakash Sahoo

Abstract—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 still remains a serious challenge. In this context, we extensively investigated 200+ latest papers on FSL published in the past three years, aiming to present a timely and comprehensive overview of the most recent advances in FSL along with impartial comparisons of strengths and weaknesses of the existing works. For the sake of avoiding conceptual confusion, we first elaborate and compare a set of similar concepts including few-shot learning, transfer learning, and meta-learning. Furthermore, we propose a novel taxonomy to classify the existing work according to the level of abstraction of knowledge in accordance with the challenges of FSL. To enrich this survey, in each subsection we provide in-depth analysis and insightful discussion about recent advances on these topics. Moreover, 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 the technology evolution trends together with potential future research opportunities in the hope of providing guidance to follow-up research.

Index Terms—Cross-domain, Few-shot Learning, Fine Tuning, Meta learning, Transfer Learning.

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, home automation, and many other justifying the Internet of Everything (IoE) frameworks. These massive number of terminal devices have been generating a huge amount of data, which need to be sent back to the server for central processing and storage. Although the total amount of generated data at the edge is very large, the volume of every dataset generated by a single device or single scene is extremely limited with very few samples. Traditional data-driven and single-domain algorithms do not perform well in these settings. To this end, numerous research has been conducted in exploring effective learning methods based on few samples and cross-domain scenes. Few-shot learning (FSL) as well as meta-learning have inevitably emerged as a promising way. However, how to effectively obtain valid information from small sample data set or even cross-domain still remains the greatest challenge faced by FSL today.

Besides, data distribution in real-world scenarios often has long-tail effects and it is difficult to generalize the same model across diverse domains. Taking the smart manufacturing industrial inspection as an example, such poor generalization issue has become one of the key challenges affecting the performance of its intelligent models. Specifically, current industrial quality inspection equipment requires certain specific lighting conditions, and the AI models trained under one lighting condition are difficult to ”generalize” to other lighting conditions. In addition, considering the high accuracy requirements of industrial quality inspection scenarios, current AI models, usually using supervised learning, require a large number of defective samples for training. However, it is difficult to collect a sufficient number of satisfactory samples as the proportion of defective products in actual production scenes is small. Moreover, it is not possible to transfer across domains either. For example, a model for PC appearance defect detection cannot be directly used to detect defects of mobile phone screens, refrigerators, washing machines, or even different models of PCs. Similarly, as another example, when recognizing character images of components and circuit boards, as there are many suppliers of components, many types of devices, and many different character styles, thus it is not possible to collect a sufficient number of all kinds of samples of character images for one supplier, resulting in few or no samples of each type. Table 1 provides a detailed summary of these challenges.

To address these challenges more effectively, FSL has produced some creative work on data, algorithms, and models. Up to now, as one of the most classical taxonomy, FSL is classified into meta-learning and metric-based learning. In this review, from the perspective of challenges, we divide the FSL into data augmentation, transfer learning, meta-learning, and multimodal learning. Data augmenta-

(Corresponding author: Ting Wang.)

- *Yisheng Song and Ting Wang are with the Engineering Research Center of Software/Hardware Co-design Technology and Application, Ministry of Education; the Shanghai Key Laboratory of Trustworthy Computing; East China Normal University, Shanghai 200062, China (email: 71205902054@stu.ecnu.edu.cn; twang@sei.ecnu.edu.cn).*
- *Subrota K Mondal is with the Faculty of Information Technology, Macau University of Science and Technology, Macao, China (email: skmondal@must.edu.mo).*
- *Jyoti Prakash Sahoo is with the Department of Computer Science & Information Technology, Institute of Technical Education and Research, Siksha 'O' Anusandhan University, India (email: jpsahoo@ieee.org).*

TABLE 1
Current challenges that industry urgently needs to address

Scenes	Challenges	Key Solutions
Quality Inspection Line	Susceptible to light Few sample data Unable to transfer across-domains	Few shot cross-domain transfer Robust model generalization
Electronic Component Identification	A variety of colors, sizes and brands Insufficient samples of each brand Existence of unseen new brands	Few-shot learning Transfer learning Unified feature representation

tion focuses on simulating data in different scenarios by metric or generative methods to maximize the actual data distribution. Transfer learning is mainly combined with pre-training and fine-tuning to extract prior knowledge from large-scale auxiliary data sets. When domain relevance is relatively uncommon or large auxiliary datasets are not available, transfer learning has definite limitations. Meta-learning is currently the mainstream approach to solve the FSL problem. In recent years, some scholars have questioned "Is such a kind of meta or episodic-training paradigm really responsible and optimal for the FSL problem?". This has led to extensive discussions [1], [2] on the necessity of meta-learning for FSL. As for multimodal learning, it integrates different dimensions of information, such as language, images, and audio. Multimodal learning is expected to break the dilemma of insufficient useful information for FSL in the real human information world.

Due to the specificity of FSL, each method of FSL is confronting multifaceted challenges to varying degrees. One of the most direct challenges in data augmentation is that the data samples are too limited and the model cannot evaluate the true data distribution by relying solely on a few samples. As a result, the model trained in this setting is biased and easily falls into over-fitting. In transfer learning, features can effectively alleviate the problem of FSL, where the volume of data is small and cannot be migrated across similar domains. Nevertheless, how to represent features effectively, how to reuse features between different tasks, and how to establish an effective mapping between data and labels are great challenges that exist in transfer learning. Moreover, in the meta-learning paradigm, when training the meta-learner with a set of tasks, it not only samples the data space but also the task space. By constantly adapting to each specific task, it makes the network have an abstract learning ability. When the training task and the target task are distinctly different, the effect of meta-learning is minimal. Furthermore, in the field of multimodal learning, extensive studies have been conducted to explore how to effectively integrate information from multiple modalities to assist the FSL.

Several existing survey papers have investigated the related work of FSL, for instance, the work [3] categorizes FSL approaches into experience learning and conception learning. The work [4] classifies FSL approaches into generative models and discriminative models according to probability distribution. Lately, the work [5] proposes a new taxonomy to classify the FSL approaches from the aspect of data, models, and algorithms. Nevertheless, to the best of our knowledge, no one paper has ever provided a taxonomy from the perspective of challenges in FSL. By summarizing the challenges of FSL, readers can better grasp the motivation and principle behind the FSL, rather than being limited to various models. A list of key acronyms used in this paper 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, introducing FSL, comparatively analyzing machine learning, meta-learning, and transfer learning along with summarizing the current variants of FSL and challenges. Furthermore, to tackle the obstacles systematically, in this section we demonstrate a new taxonomy to classify the existing FSL related works. Section 3 to Section 6 present a systematic investigation on the current mainstream researches 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, giving our insights into the respective fields accompanied by some potential research opportunities. Section 7 takes computer vision as an example, counting the latest progress of FSL in image classification, object detection, semantics segmentation, and instance segmentation in chronological order. Section 8 delves into exploring the current challenges faced by FSL and how to seek breakthroughs in each branch. The overall outline of this paper is shown in Fig. 1.

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

- We start with the edge computing scenario, in which the few-shot learning challenges arise, explaining and clarifying several similar concepts that are easily confused with FSL. 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 challenges through knowledge graphs and heat maps. With this taxonomy, we divide the FSL into several different levels, where the highest level is multimodal learning that mainly uses various semantic knowledge to assist judgment, and the second, third, and fourth levels are single-modal learning that addresses data level, feature level, and task level challenges, respectively. Notably, we also provide insightful discussions on FSL cross-domain research, which is currently the more challenging direction in the field of FSL.

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	Conventional Neural Network	CNN
K-NearestNeighbor	KNN	Support Vector Machine	SVM
Nearestcentroid Classifier	NCC	Graph Few-Shot learning	GFL
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

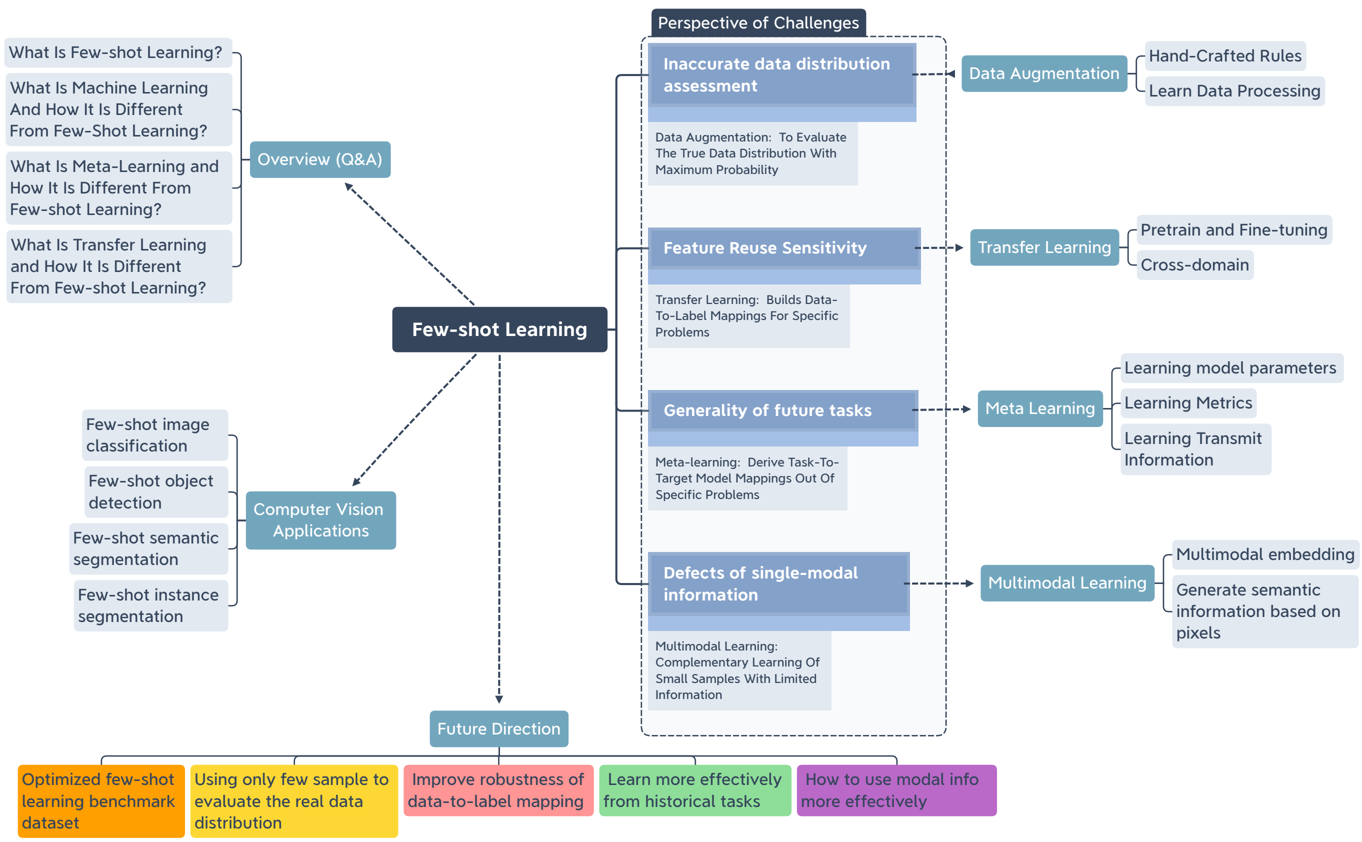

Fig. 1. The conceptual map of the survey.

- We investigate an adequate number of papers in recent three years and summarize the main achievements of FSL in the field of computer vision, including image classification, object detection, semantic segmentation, and instance segmentation.
- With these challenges mentioned at the end of the survey, combined with practical applications, we delve into the current challenges of FSL and explore how to find breakthrough points in each branch to jointly drive the research of FSL towards a more practical direction.
- We provide unique insights into the evolution of FSL and identify several future directions and potential research opportunities concerning each challenge.

2 Concepts and Preliminaries

As a branch of machine learning (ML), FSL is still a young field. What is FSL, and how does it relate to machine learning, transfer learning, and meta-learning? What are the variants of FSL that currently exist? What benchmark datasets frequently appear in research papers? In this section, we will address 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, and it is widely found in edge computing scenarios. In 2020, Wang et al. [5] give a detailed definition of FSL through experience, task, and performance of machine learning, which is one of the most recognized definitions so far: 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 here that E in FSL is very small. In recent years, relevant neural scientific evidence [6], [7] has shown 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.

To better understand FSL, it is necessary to introduce two concepts, one is N-way-K-shot problem and the other is cross-domain FSL. The N-way-K-shot problem is often used to describe the specific problems encountered by FSL. In this case, the support set represents the small dataset used in the training phase, which generates reference information for the second phase of testing. The query set is the task on which the model actually needs to predict. Notice that the query set classes never appear in the support set. Classical N-way-K-shot represents support set with N categories and K samples per category, then the whole task has only $N * K$ samples. As thus, N-way-1-shot represents one-shot learning and N-way-0-shot represents zero-shot learning. The concept of cross-domain originates from transfer learning, which refers to the transfer the knowledge from source domain to target domain. There usually exist domain gaps between these domains. Cross-domain FSL integrates the features of cross-domain and FSL, and is a challenging direction that has recently emerged.

At this stage, there still exist many challenges in FSL, which are generated from various aspects, including but not limited to data, models, and algorithms. In this context, the challenges can be generally summarized according to the degree of integration of knowledge as follows:

- `Inaccurate data distribution assessment`: FSL does not have access to large datasets due to costs, ethical, legal, or other reasons. Consequently, relying on few samples for learning produces biases in estimating the actual data distribution, which may be fatal for some tasks. To this end, maximizing the exploration of data distributions with limited information becomes the most significant challenge for FSL. Data augmentation is the direct way to address the inaccurate estimation of FSL. The primary efforts currently focus on exploring migratable intra-class or inter-class features and customizing specific images using generators.
- `Feature reuse sensitivity`: Continuous accumulation of a priori knowledge by sampling large-scale auxiliary datasets. Transfer learning can easily use it from the source domain to the similar target domain. Pre-training aims to extract high-dimensional feature vectors through a feature extractor, while the goal of fine-tuning is to make minor adjustments to the initial parameters of the pre-training. Transfer learning focuses on the data level and obtains more valuable features independent of the task by mapping data to labels. It has an outstanding performance in optimizing specific tasks, but it is generally limited by the characteristics of current tasks and has a poor generalization to future tasks. Especially when there is a large shift in the domain, without filtering and alignment of features may result in negative knowledge transfer.
- `Generality of future tasks`: Differing from transfer learning, meta-learning learns to quickly build mappings from known tasks to target models in previously unseen tasks by double sampling the task and data. In FSL, by exploring the task space, summarizing meta-knowledge in different tasks can result in fast aggregation of unseen tasks at a lower cost. As a general learning framework, meta-learning is independent of specific problems and more oriented to future tasks instead of optimizing the current one. However, meta-learning has proven effective only when the testing and training tasks are relatively similar, and it is highly depends on the network structure and lacks flexibility. When training meta-learners with a set of tasks at the same time, it is even difficult to adapt to the distribution of tasks, requiring a redesign of the network structure.
- `Defects of single-modal information`: It is difficult to learn features effectively because FSL is inherently information-limited. This situation is improved to a great extent when aided by getting information from other modalities. In this respect, semantic assistance [8], [9] is an excellent method to provide external prior knowledge, where through the introduction or generation of semantic information as weak supervision, adaptive classification can be accomplished in conjunction with the original model.

2.2 What Is Machine Learning And How It Is Different From Few-shot Learning?

The traditional Von Neumann computer architecture allows users to execute a series of instructions step by step in the form of a program [10]. However, this method does not work in ML. On the contrary, ML uses large-scale datasets as input. Its judgment on a new sample is based on statistical results extracted by historical data. Now, the burgeoning of 5G [11] provides massive connectivity for millions of end devices, enabling an interconnection of everything. The total amount of data generated by terminal devices is huge, but the amount of one single data set is extremely small. Hence, traditional ML, whose performance strongly depends on large data sets, cannot perform well in this setting with few samples. To this end, FSL emerges and provides a promising way to handle the data scarcity scenario.

In recent ten years, the research on FSL has been extensively conducted, and significant research progress has been achieved, e.g. the KGBert [12] proposed by Alibaba surpasses humans in the field of FSL for the first time. Fig. 2 exhibits the statistics of paper publications related to FSL in recent ten years based on the statistical results of the Web of Science. As revealed, there are relatively few related papers from 2011 to 2015 due to that the FSL theory is still incomplete. With the rise of deep learning since 2015, the number of FSL related research papers has increased linearly almost every year. In the past 2020, the number of relevant papers has reached as high as 239, and the number of citations has reached 2731 times, accordingly. Fig. 3 provides a knowledge map covering the hot research topics and cutting-edge developments in the field of FSL in recent years, including but not limited to zero-shot learning, one-shot learning, transfer learning, multi-task learning, and meta-learning. FSL related tasks include feature representation, visualization, robotics and cross-domain transfer. Among them, domain adaptation is a widely utilized method for few-shot cross-domain learning. Computer vision with predominant green color is the most active research field, including image classification, object detection, semantic segmentation, and instance segmentation.

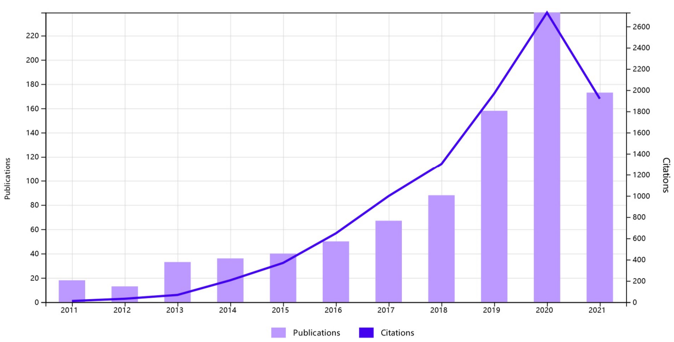

Fig. 2. The number of FSL-related papers published in prestigious journals from 2010 to the first half of 2021, excluding citations.

Fig. 3. The knowledge graph uses Few-shot learning, One-shot learning and Zero-shot learning as keywords to relate the main advances and research directions in the field of Few-shot learning in recent three years.

The most significant difference between FSL and traditional machine learning is that the set of classes of the support set and the query set are disjoint. In machine learning, the classes of the test set are included in the training set in advance. FSL combines the limited supervision information with prior knowledge to train the model. The input of the model is generally given in the form of tasks. Through continuously collecting tasks, the model can recognize the similarities and differences between data as well as task. When the model encounters an unseen task, knowledge transfer can be accomplished quickly with just a few iterative training steps with appropriate initialization parameters. In contrast, traditional machine learning requires optimization through the loss function generated by a large scale data sets in the model. In conclusion, FSL is only a very young branch of machine learning, which mainly addresses the issue of difficult access to quality data sets in machine learning scenarios.

2.3 What Is Transfer Learning and How It Is Different From Few-shot Learning?

Traditional transfer learning involves applying knowledge learned in the source domain to a different but related target domain. In FSL, the limited amount of training data, domain variations, and task modifications are the key factors that cause the model to fail to transfer well from the source domain to the target domain. For instance, a medical image dataset with low similarity to the natural image dataset imagenet is difficult to identify accurately without the help of relevant expertise, even for a human-guided by only a few images. Certainly, it is also effective when the source domain and target domain are relative similar. The end in FSL tasks, if the prior knowledge is obtained from other tasks or domains by pre-training, FSL can belong to transfer learning, which mainly learns the mapping of data to labels.

According to the taxonomy of transfer learning, there are many variants of FSL problems, including one-shot learning (OSL), zero-shot learning (ZSL), and cross-domain few-shot learning.

- `One-shot learning`: OSL has only one correct label for each sample in the support data set, which aims to find the most similar class as a match among the seen-classes. During the police interrogation, these two processes are incredibly similar. The witness just looked once at the suspect, and the photos given by the police can be regarded as the query image. The witness only needs to answer 'yes' or 'no' towards those photos. Similarly, one-shot learning does not classify the data specifically, but simply makes a cluster in order of similarity function. According to the existing work, one-shot learning can be divided into two main approaches. One is to use generative models to caste prior knowledge [13], [14], [15], where bayesian programming learning [16] is the most representative framework [17] in this field. Another method is to convert a OSL classification task into a verification task [18], [19].
- `Zero-shot learning`: ZSL was first proposed by Lampert et al. [20], which considers a more extreme case in FSL. In the absence of any query samples, the inference mechanism is solely relied on to identify samples that have not been seen before. ZSL is essentially done by using high-dimensional semantic features [21], [22], [23] to replace the low-dimensional raw data. Embedding representations and autoencoders are the most efficient ways to construct intermediate semantic spaces, which contains attributes that more comprehensively define the

categories. Up to now, zero-shot learning is one of the closest methods to human intelligence that discerns previously unobserved categories. One-shot learning and FSL can essentially be considered as special ZSL.

- `Cross-domian few-shot learning`: In transfer learning, each class in the target domain has a sufficiently large number of available samples. When a large domain shift occurs between the source domains and target domains, knowledge transfer tends to become very challenging. Cross-domain few-shot learning combines the challenges of transfer learning and FSL. In the existence of domain gaps, where the intersection of classes in the source and target domains is empty, and the available sample size for each class in the target domain is extremely small. The improvement of the model's generalization capability through source domain data alone brings very limited performance to the model. The present work mainly focuses on the shift transformation of features and the construction of auxiliary datasets. Cross-domian few-shot learning can be regarded as one of the most challenging setting in the field of FSL at present.

2.4 What Is Meta-Learning and How It Is Different From Few-shot Learning?

Meta-learning is a general learning paradigm that provides training on tasks in an episodic-training mechanism. Fig. 4 illustrates the three-steps involved in meta-learning [24] training. Meta-learning focuses on improving generalization for unseen tasks using prior knowledge. If prior knowledge is used to teach the model how to learn on a specific task, meta-learning can be regarded as a variant of FSL. It is emphasized that meta learning is not equivalent to FSL. FSL should be seen rather as an ultimate goal. It aims to achieve robust generations without relying on a large scale of datasets. By dual sampling of data and task space, meta-learning is enabled to construct a large number of auxiliary tasks related to the unseen task. Even if some papers do not use meta-learning, it is likely to improve the model's performance if episodic-training mechanism can be considered, such as meta reinforcement learning [25], [26], meta video detection [27], and so on.

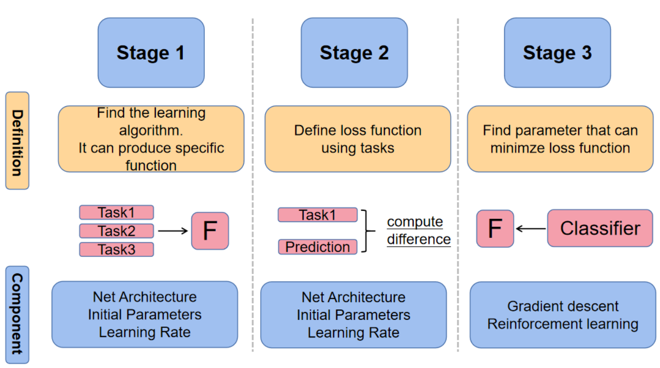

Fig. 4. Meta-learning training three-step approach includes: find the learning algorithm, define loss function using tasks, find parameter that can minimze loss function.

Nevertheless, meta-learning has its own limitations: when the training and testing tasks exist obviously domain gap, Meta-learning is rarely used to initialize parameter weights. It can easily lead to negative migration of the model. 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 solve FSL.

2.5 Datasets

Before the availability of FSL benchmark datasets, researchers regularly used tasks like manually constructing N-way-K-shots to evaluate the performance of models. However, these simple tasks cannot reflect the complexity of real-world problems. After 10 years of evolution, the FSL benchmark dataset has completed the transition from a single domain, single dataset to a cross-domain, multiple dataset.

As shown in Fig. 5, during 2017-2021, 898 papers used the CUB-200-2011 [28] dataset, accounting for 46.6% of the total number of statistics; 587 papers used the Mini-ImageNet [29] dataset, accounting for 30.5%; and 335 papers using the Omniglot [30] dataset, accounting for 17.4% ; 44 papers used the PASCAL-5i [31] dataset, and 46 papers used the Meta-DataSet [32]. The other specific datasets are Paris-Lille-3D [33], N-Digit MNIST [34], SUN397 [35], which are used in 15 papers in the past five years. In terms of quantity, the CUB-200-2011, Mini-ImageNet, and Omniglot benchmark datasets occupy a dominant position in the field of FSL. Table. 3 compares the datasets mentioned above from different dimensions. By the publication of the article, a more objective benchmark dataset [36] for evaluating the cross-domain ability of FSL was proposed. 1) CropDiseases [37], a plant diseases dataset , 2) EuroSAT [38], a dataset for satellite images, 3) ISIC [39] a medical skin image dataset, 4) ChestX [40], a dataset for X-ray chest images. The similarity comparing to MiniImageNet is decreas across these datasets.

2.6 Taxonomy

According to the degree of integration of knowledge, FSL is broaderly divided into a single-modal learning and a multimodal learning. In this survey, The single-modal learning can be further divided into data augmentation, transfer learning, and meta learning. It mainly focuses on abstracting or transferring limited information into higher-level feature vectors or meta-knowledge. Multimodal learning is more close to the real world of human intelligence, which no longer relies on the limited sample, and tries to find the space of other modalities to assist the FSL. With this taxonomy, we exhaustively review and discuss each method. Fig. 6 vividly demonstrates the FSL's taxonomy under the challenge perspective.

- `Evaluate The True Data Distribution`: The key to the difficulty of FSL is that limited samples cannot reflect the actual data distribution. The most intuitive idea of machine learning is to generate additional data based on a certain probability model or to extend the auxiliary data set using a large

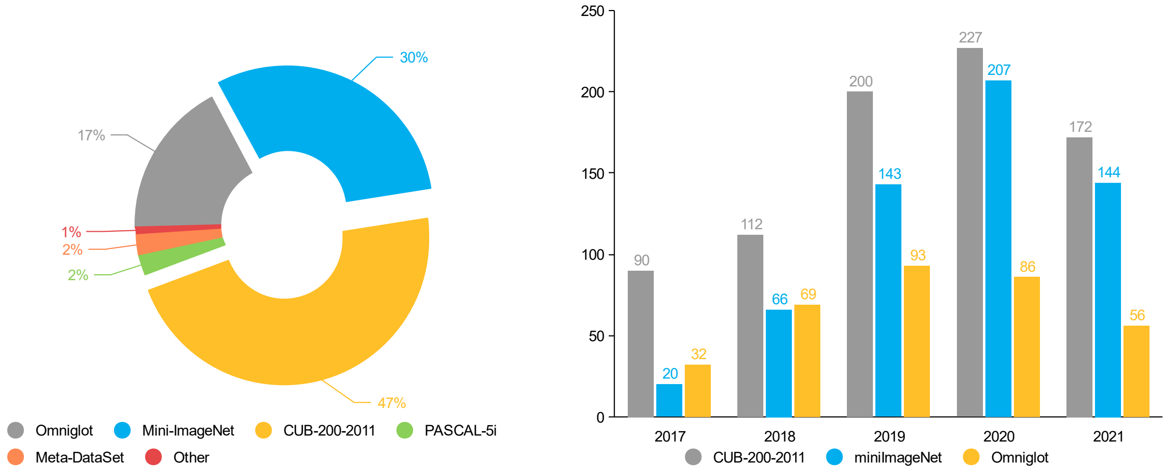

Fig. 5. There are eight most frequently used datasets in few-shot learning, including the number of papers on mainstream benchmark datasets (2017-2021). There may be one paper that tests all mainstream benchmark datasets. Data from "paperswithcode" platform.

TABLE 3
The latest performance of FSL in the main tasks of machine vision

Dataset Variant	Leader	Numbers/Classes	Train/Test	Content	Main FSL Task	License
CUB-200	Wah el al. [28]	11788/200	5994/5794	Birds	Few-shot Image classifition	Attribution 4.0
Mini-ImageNet	Vinyals et al. [29]	600/100	480/120	Real scene	Few-shot Image classifition	MIT
Omniglot	Lake et al. [30]	32460/50	4800/1692	Character	Few-shot Image classifition	MIT
PASCAL-5i	Shaban et al. [31]	576/20	11530/-	Really scence	Few-shot Image classifition	MIT
Meta-Dataset	Triantafillou et al. [32]	10 datasets	-/-	Really scence	Few-shot Image classifition	Multiple licenses
BSCD-FSL	Guo et al. [36]	4 datasets	-/-	Really scence, Satellite and medical image	Cross-domain few-shot learning	MIT
Paris-Lille-3D	Roynard et al. [33]	450000000/50	450000000/380000000	Point cloud	Few-shot Semantic Segmentation	CC BY-NC-ND 3.0
N-Digit MNIST	Oh et al. [34]	-/-	-/-	Character	Metric Learning	Apache License
SUN397	Oh et al. [35]	108,753/397	76128/21750	Really scence	Few-shot Image classifition	-

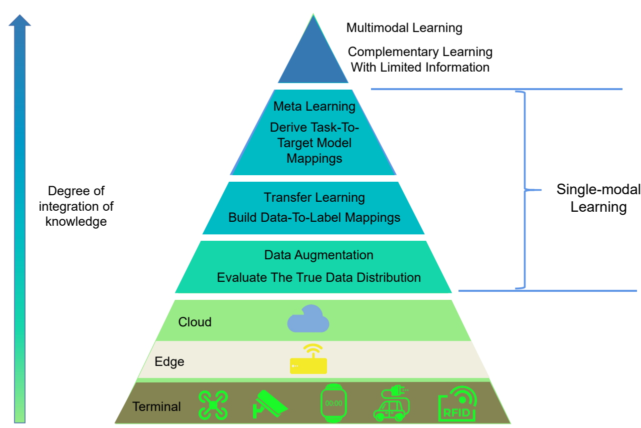

Fig. 6. The entire taxonomy is presented in the form of a pyramid. The bottom level represents the "cloud-edge-terminal" edge computing scenario, which is characterized by few-shot real-time computation under high traffic. Based on this, the challenges of FSL are classified into four levels according to the degree of integration of the required knowledge. Among them, the challenges represented by data augmentation, transfer learning, and meta-learning are single-modal challenges.

volume of unlabeled data from extending data. Existing work focuses on exploring feature differences that can be learned between classes or with external datasets at the semantic level. Handcrafted rules and automatic learning data processing are the two main approaches at this stage.

- `Build Data-To-Label Mappings`: Furthermore, if a large number of features from the benchmark dataset can be reused, which will significantly reduce the pressure of the model on the data. Pre-training and fine-tuning assist FSL by learning effective representation of data to labels, coupled with effective regularization of the underlying semantic features. In particular, the pre-training stage learns the optimal initialization parameters in a variety of different tasks, and the fine-tuning stage freezes most of the lower-level parameters and retrains only the parameters of the classification layer.
- `Derive Task-To-Target Model Mappings`: Fine-tuning already has a good performance in baseline models with small samples. Nevertheless, in multi-task learning, a large scale of tasks are learned just as one task, which leads to a terrible generalization of the model. In contrast, meta-learning dual-samples the data and task space using episodic-training mechanism, finding latent associations between different tasks and thus having a good description of the whole task space.
- `Complementary Learning With Limited Information`: learning has been proposed in deep learning for a long time, but it has only started to be integrated with FSL in recent years. Information in multimodal dimensions is closest to the real human information world, and it compensates to some extent for the inability of FSL to make accurate assessments of data distributions in a single modality. Pixels, semantics, and sounds can be supervised signals for FSL tasks, and even more recently unsupervised

learning has been used to explore more robust feature representations using contrast learning.

3 Data Augmentation To Evaluate The True Data Distribution With Maximum Probability

In real-world FSL tasks, the number of samples in the support and query sets is usually limited due to privacy, collection costs, and labeling costs. To mitigate this issue, data augmentation is recognized as the most direct way to increase the sample richness in FSL. Nevertheless, the core risk of the FSL data augmentation is how likely the augmented dataset can evaluate the distribution behind the real data. Based on whether the data augmentation techniques can be reused on other tasks, FSL data augmentation is divided into hand-crafted rules and automatic learning data processing.

3.1 Hand-Crafted Rules

Hand-Crafted rules require guidance from experts with specialized domain knowledge. A representative result is that Bouthillier et al. proposed to randomly discard pixels [41] on a random rectangular area to generate black rectangular blocks of simulated noise. Similar operations also include random erase [42] and fill [43], [44] in FSL. Nevertheless, briefly relying on the simple transformation of single-sample pixels cannot prevent the risk of overfitting. Further, the hand-crafted rules contains data level and feature level according to the dimension of information. Table. 4 summarizes the data augmentation methods for hand-crafted rules making.

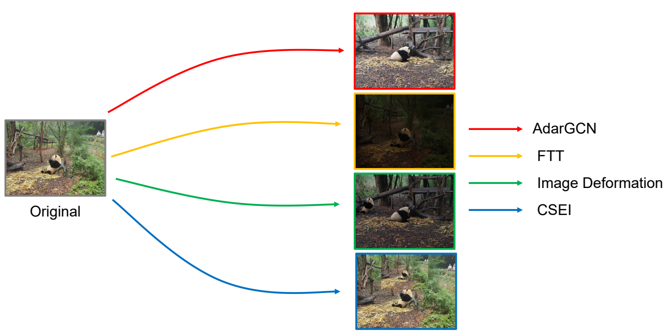

Fig. 7. FSL data augmentation based on data level mainly includes internet data collection, environment variation, difference transfer and random crop filling. Here a picture of a panda is used as an example to implement the above variation.

3.1.1 Data Level

The data level augmentation is mainly a transformation of the input data that aims to scale up existing data by making modifications to the data marginally for achieving diversity in model input. Random erasure [42] and random cropping [43], [53] are classical algorithms by simulating the images with different degrees of damage and thus improving the generalization of the model. Inspired by this, Li et al. [46] discarded the traditional approach based on a entire feature extractor for images and instead focused on local patch images. These methods require large-scale datasets as support. It is not easy to achieve in the FSL settings. Conversely, CSEI [46] does not require extra data sets. The specific operation is to erase most of the discriminatory area in the support set derived from the metric function and replace it with an image fill using the restoring operation. FTT [45] enriches the data set by linear interpolation of some transiently transformed attributes, such as different weather and lighting. Z Chen et al. [15] proposes an end-to-end approach to partitioning images as a whole inspired with the idea of MIXUP [54], which argues that images preserve important semantic information even after they have undergone various distortions. The most significant difference between image distortion and GNN is that image distortion simply stitches two images together in a linear pattern. This method is able to achieve maximum deformation without loss of classification. In addition, it is a good direction for data expansion by using the large amount of unlabelled data sets in the real world for supplementation. Finally, when both the source and target classes both have only a limited number of samples, AdarGCN's [47] implementation crawls data from internet resources and automatically removes irrelevant noise to achieve controllable data augmentation. At the same time, AdarGCN can automatically determine how far the information has been propagated in each graph node. In conclusion, data augmentation at the data level focuses on increasing the number of samples by means of pixel transformations and pixel generation. Fig. 7 shows the main methods based on data level under hand-crafted rules.

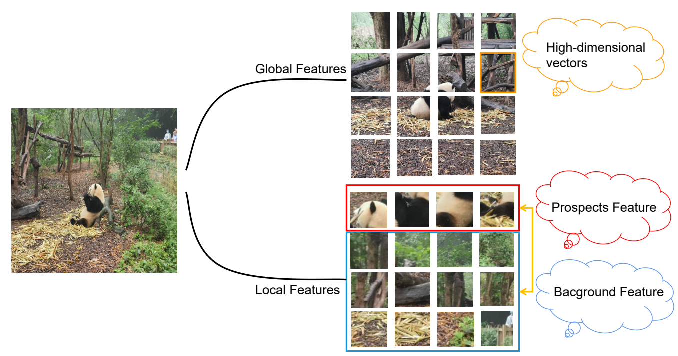

Fig. 8. Feature level-based data enhancement can be mainly divided into global features and local features. Global features focus on the whole image, including the foreground and background. Local features, on the other hand, selectively focus on the subject part in the foreground.

3.1.2 Feature Level

Feature level data augmentation mostly maps pixel information into a high-dimensional latent space. It carries more valid information than mere original pixels. Gao et al. [48] first explored the underlying distribution behind few-shot data and proposed an adversarial covariance augmentation network to overcome the limitations of FSL. Its experiments have shown that relying solely on learning the features of the entire image brings noise into the results. Chu et al. [49] tried to compute feature representations for each patch, rather than the entire image. Each small patch is connected by RNN and the features of the image are further fused. This heuristic algorithm is far superior to simple attentional models[55]. Zhang et al. [50] explain partial feature learning

TABLE 4
The latest performance of FSL in the hand-crafted rules.

Model	Core View	Key Approach	Experimental Dataset	Using External Dataset	Data Level	Feature Level
FTT [45]	Enriching Instant Attributes	Places-CNN	Transient Attributes Database	✔	✔	
CSEI [46]	Erase Repair	Metric based	miniImageNet	✗	✔	
Image Deformation [15]	Semantic Invariance	Meta-learning	miniImageNet	✔	✔	
AdarGCN [47]	Denoising the collected web images	GCN layer	-	✔	✔	-
Covariance-Preserving Adversarial Augmentation Network [48]	”Variability” of covariance information as base instances	Generative Adversarial Network	ImageNet	✔		✔
Spot and Learn [49]	extracts varying sequences of patches	reinforcement learning	miniImagenet	✗		✔
Saliency-guided Hallucination [50]	Background - Prospective Learning	Realation network	miniImagenet	✗		✔
Laso [51]	Explore the reliable differences between labels	Transfer learning	MS-COCO	✔		✔
Dual TriNet [52]	Semantic Synthesis Example	Auto Encoder	MS-COCO	✗		✔

from another perspective, proposing to use a pre-trained model to decompose visual features into three parts and then select the original, foreground, and background images to be re-stitched into new visual features. Similarly, Laso [51] explores the differences in features between different datasets in a high-dimensional space. Combining different labels through the intersection and complementation of sets allows images to contain key information from multiple classes at the feature level simultaneously. Training this part of the image as a support set can significantly improve the classification performance of small samples. Chen et al. [56] go further by extending the features to a high-dimensional semantic space. In FSL, feature-level augmentation is more effective than data-level augmentation by modeling the valid information in a compressed manner. Fig. 8 shows the main methods based on features level under hand-crafted rules.

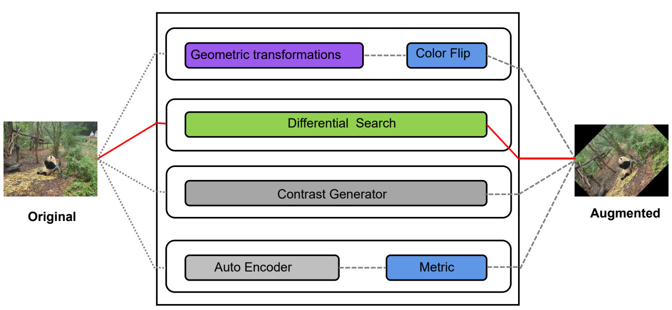

Fig. 9. Learned Data Processing aims to learn a policy generator in multiple task spaces so as to automatically match different tasks. Its biggest benefit over hand-crafted rule is that it can be reused.

3.2 Learning Data Processing

In 2018, data augmentation entered the area of auto augmentation with the maturation of meta-learning. Through the combination of meta-learning with other data augmentation methods, a large amount of excellent work emerged during this period. Hu et al. [57] was inspired by the DARTS algorithm to abstract the data augmentation into multiple sub-strategies, each with a certain probability of being selected according to the different few-shot tasks. In addition to the probability-based method, another approach is based on generation. Li et al. [58] proposed the adversarial feature phantom network-AFHN. The phantom diversity and discriminative features are conditional on a small number of labelled samples. Chen et al. [56] attempt to train a meta-learner and generate a network to learn similarities and differences between images end-to-end by fusing pairs of images. MetaGAN [59], on top of which an adversarial generator conditional on the task is introduced, which helps FSL tasks form generalizable decision boundaries between different classes. On the other hand, Zhang et al. [50] further demonstrate the usefulness of phantom data generation for FSL and propose a low-cost automated data generation method that uses a direct foreground-background combination to generate feature space-level data for training. In addition, it is also effective to explore the migratable differences between and within classes in support datasets. Delta-encoder [60] uses auto-encoder [61] to learn differences in the same class for transfer learning, which is different from metric-based computation [62] of visual similarity. Table. 5 summarizes the data augmentation methods for learning data process. Fig. 9 shows the main approaches involved in auto-learning data processing under FSL.

3.3 Discussion and Summary

To maximize the evaluation of the distribution of the real data in FSL setting, data augmentation has from the hand-crafted rules to the auto learned data processing stage. The watershed is the maturity of meta-learning in 2018. This section comprehensively investigates the emerging representative technologies in the field of data augmentation and reviews the evolution of few-shot data augmentation. Table. 5 summarizes the model in different dimensions clearly.

4 Transfer Learning Builds Data-To-Label Mappings For Specific Problems

Transfer learning [63] is a classical learning paradigm, which aims to solve the challenging problem that there are only a few or even no labelled samples [64] in the FSL [65]. Feature reuse is the core idea of transfer learning to solve FSL absence of data setting. The basic operation is to pre-train the model on an extensive dataset and then fine-tune on the limited support set. When source and target domains exist a large gap, knowledge transfer is invariably much less effective. This cross-domain setting brings a new challenge for FSL. In FSL, transfer learning can be broadly divided into pre-training and fine-tuning stage, which can also be referred to the baseline. Fig. 10 illustrates the general process.

4.1 Pre-training and Fine-Tuning

From 2012 to 2018, a large number of excellent works have emerged in the field of computer vision and natural

TABLE 5
The latest performance of FSL in the field of learn data process.

Model	Core View	Key Approach	Experimental Dataset	Using External Data
DADA [57]	Automatic generation of enhancement policies	Gradient Descent	CIFAR-10	✔
AFHN [58]	Condition-based generation	Generative Adversarial Network	MNIST	✔
MetaGAN [59]	Generate extral data	Generative Adversarial Network	MNIST	✔
Delta-encoder [60]	Differential transfer	Auto Encoder	Mini-ImageNet	✔
MSFN [62]	Calculate multi-scale features and the similarity of each class of labels	Metric Learning	Omniglot	✔

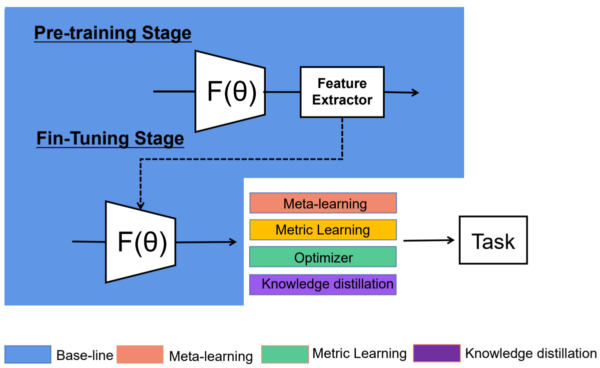

Fig. 10. Transfer learning can be divided into pre-training and fine-tuning stages, where the baseline model can be combined with other techniques to improve the model performance.

language processing, such as MobileNet [66], ResNet [67], ELMO [68], GPT [69], and BERT [70]. In particular, the area of natural language processing was slow to progress before the advent of pre-training models. It has grown considerably under the leadership of BERT as computing power has increased and excellent pre-training models have been proposed.

As a downstream task, how to use these excellent models to obtain features will largely alleviate the pressure on the data for FSL. Especially for few-shot image classification, as a pre-training model [71], [72], it needs to use an external large-scale label dataset to extract prior knowledge from similar task. The most common practice is to design a backbone model without classifier layer, which includes convolutional neural networks or auto-encoders. The input of the model is an array of images, and the output is the feature vector embedding in a high-dimensional space [73]. High-dimensional feature vectors obtain sufficient valid semantic information about the target image. After the pre-training was put forward, the researchers proposed fine-tuning later. Most parameters in the pre-training are frozen, and only the classification layer parameters are updated in the testing stage. Many recent works [74], [1] have proved that fine-tuning can improve the 5-way-1-shot tasks accuracy rate by 2%-7% compared with the baseline model. Although the number of samples in the support set and query set is small, pre-training and fine-tuning are still very helpful for improving the accuracy of FSL. The conclusions are analogous in natural language processing as well. The authors in [75], [76], [77] have also shown that fine-tuning can be embedded in state-of-the-art meta-learning or semi-supervised learning frameworks for optimizing model parameters.

Dhillon et al. [76] replaced the standard activation function with cosine similarity, and Nakamura et al. [74] replaced conventional gradient descent with an adaptive gradient optimizer, which both improve the fine-tuning process in the accuracy of the model. Currently, fine-tuning is usually combined with meta-learning. Cai et al. [78] attempted to integrate them to train networks with specific layers. However, experimental results suggest that since the support and query sets do not overlap in FSL setting, transferring whole knowledge from the source dataset is not the best solution for FSL. Shen et al. [79] suggest that knowledge should be transferred specifically for parts. The degree of transferability needs to be controlled by freezing or fine-tuning specific layers in the backbone model. Similarly, fine-tuning can also be used to prevent new classes of networks from polluting the feature space of the basic classes. Up to now, the FSL and fine-tuning have been widely used in tasks like plant disease and insect pest identification [80], road detection [81], and automatic question and answer [82].

4.2 Cross-Domain Few-shot Learning

The latest progress of FSL largely depends on the labelled data of the training stage. However, it is unrealistic to collect various forms of datasets for specific tasks in many practical applications, which results in challenging of FSL between intensely different domains. Cross-domain few shot learning integrates FSL and domain adaptive problems, which is a relatively comprehensive and challenging setting. For a long time, the benchmark datasets commonly used for FSL have suffered from a standardized dataset structure and large similarity of natural scenes, which leads to that models perform well on standard datasets but get unacceptable results in the real world task. Google first released a FSL cross-domain dataset named Meta-Dataset [32] in 2020, which includes a total of 10 public image datasets including ImageNet, CUB-200-2011, etc. Yet these datasets are still focused on natural scenarios and cannot be broadly regarded as cross-domain few-shot benchmark datasets. Until the availability of BSCD-FSL [36] datasets. According to the degree of similarity with the ImageNet, it is divided into CropDiseases [37], EuroSAT [38], ISIC [39], ChestX [40]. The authors extensively evaluate the performance of current FSL methods, and experiments show that the accuracy of all methods is correlated with the proposed natural image data similarity metric. Nowadays, cross-domain FSL focuses on distinguishing domain-irrelevant features and domain adaptive techniques with transfer learning.

The objective of domain adaptation is to transfer knowledge from the source domain to the target domain, which has the same set of classes but a different data distribution than the source domain. Recently, much work has used adaptive networks to align their features with a new domain or to select domain-irrelevant features from multiple backbone model. Dvornik et al. [83] obtained multiple domain representations separately by training a set of feature extractors with different domains. Setting the model to a dataset with multiple domains during training allows an attempt to migrate to other domains during the testing stage. Nevertheless, this approach may not be effective during the meta-training and meta-testing phases when the domains are orthogonal. Based on this, FRN [84] explores the potential space for few-shot image classification, using ridge regression to reconstruct and normalize the feature map without adding new learning parameters. FWT [85] utilizes only the source data for the affine transformation of features, as do LRP-GNN [86] and SBMTL [87]. FD-MIXUP [88] constructs auxiliary datasets by mixup and uses encoders to learn domain-irrelevant features to guide the network generalization to other tasks. STARTUP [89] takes advantage of not only the source data but also assumes that the model has access to a lot of unlabeled target data during training. A large amount of unlabeled data is used to enhance the generalizability of the model to other domains. Metric-based approaches are frequently used for semi-supervised and unsupervised cross-domain FSL. A recent paper by Lu et al. [90] uses attention as a metric strategy to reweight and combine domain-specific representations. Chen et al. [72] based on a meta-baseline by pre-training the classifier on all base classes and classifying a small number of samples based on the nearest centroid algorithms for meta-learning, which greatly surpasses the latest state-of-the-art methods. Li et al. [91] inspired by [92], [93], proposes to map domain-specific features to the same shared space, thus achieving a domain-irrelevant universal representation.

4.3 Discussion and Summary

While meta-learning methods have higher performance than transfer learning in standard FSL settings, the situation is reversed in cross-domain FSL settings. A newly published paper recently pointed out that the improvements from fine-tuning and pre-training are similarly very limited when the domains appear orthogonal. In the pre-trained feature space, the base classes form compact clusters, while the new classes are distributed in large difference groups. Currently, the actual deployment of trained models into production environments is often not adapted to rapidly changing environments. Pre-training can be seen as a task with many learning classes, but it is only a single learning task.

5 Meta-learning Derive Task-To-Target Model Mappings Independent Of Specific Problems

Meta-learning learns historical prior knowledge from a dual sampling of data and tasks, then extracts meta-knowledge to apply to future tasks. Meta-learning is independent of the specific problem, and exploring an optimal initialization parameter in task space, discarding the task-independent feature representation under traditional supervised learning. Up to now, most of the meta-learning models are updated with parameters using traditional gradient descent. Absolutely, there are also non-gradient descent methods based on reinforcement learning and metric methods. In FSL, meta-learning can be used to automate the learning of model parameters, metrics function, and the transfer of information

5.1 Learning Model Parameters

Most of the deep learning frameworks use different parameter initialization methods, such as uniform distribution, normal distribution, and so on. The biggest problem with this random initialization is that it easily falls into the local optimal position. The goal of meta-learning is to train a hyperparameter generator, the classical methods being MAML [94], Repital [95] even their derived variants. MAML identifies the global optimization direction by calculating the optimization direction for each task. Compared to MAML, Reptile can update fewer parameters at once. The biggest difference between meta learning and multi-task learning is that multi-task learning only focuses on the performance of the current task. Meta-learning was demonstrated to perform better than transfer learning with a standard FSL benchmark dataset. Nevertheless, meta-learning is more sensitive to network structure and requires fine tuning of hyperparameters. After that more versions have evolved to address these issues separately. Such as MAML++ [96], First-order MAML (FOMAML) [95], Meta-SGD [97], TAML [98], iMAML [99], iTMAML [100]. Of which Meta-SGD, expect MAML, finds the optimal learning rate and update the direction of the parameters at the same time, in addition to learning the initialization parameters. TAML [98] is a task-independent method, which overcomes the problem that MAML can only use an external model. Subsequently, IMAML [99] proposes a new loss function and a corresponding method for computing the gradient, making it possible to obtain the gradient of the parameters by calculating only the solution of the loss function, without caring its specific optimization method. iTMAML [100] based on TAML, which implements automatic task recognition. It can be quickly adapted to new tasks by updating when the data is in a continuous state. At present, MAML has been widely used in various tasks [101], [102], [103], producing different variants. Table. 6 distinguishes between MAML, Reptile and their variants in various perspective.

Learning optimizers are another important direction for learning model parameters. LSTM as the base optimizer [104], [105], which accepts the difference at time t and the hidden state of the meta-network at time $t-1$. The output of the original network is an updating of the model's weight and bias. In 2016, Xu et al. [106] proposed the BPTT to supervise LSTM training. It is notable that this is performed in the context of supervised learning. What update should be required to the optimization if it is in the setting of unsupervised and active learning? Inspired by this, there has been a long period of work focusing on

TABLE 6
Summary MAML, Reptile and their variants

Model	Directions for improvement	Key Approach	First order gradient	Two-step gradient
MAML [94]	Original	Inner-loop+outer-loop	✗	✔
Reptile [95]	Computational Complexity	Standard stochastic gradient descent	✔	✗
FOMAML [95]	Simplify secondary gradient updates	Using the gradient calculated from the previous task	✔	✗
Meta-SGD [97]	Increasing the Volume of the Model	Increase the learning rate vector parameter	✗	✔
TAML [98]	Task unbiased estimation	Introducing Entropy and Inequality Metrics	✗	✔
iMAML [99]	Gradient disappearance	Propose new loss functions and optimization methods	✗	✔
iTMAML [100]	Automatic task identification	Data is in continuous state	✗	✔

reinforcement learning [107], Bayesian inference [108] and evolutionary algorithms [109] in an attempt to automatically find optimization strategies through heuristic algorithms.

Finally, traditional Neural Architecture Search (NAS) also incorporated the idea of meta-learning and adapted it accordingly under FSL. To our best knowledge, the shared [110] and randomly selected supernet weights [111], [112] were early solutions for FSL. Recently, a large volume of work [112], [113], [114] has shown that performance differences still exist between one-shot NAS and traditional NAS. The one-shot NAS uses weight-sharing networks to train the supernetwork only once and then perform a single round of inference to get an accurate prediction, greatly reducing the amount of computation required for the experiment. Subsequently, Zhao et al. [115] proposed the few-shot NAS based on one-shot NAS. The core idea is to divide the supernets into multiple sub-supernets to search different regions of the search space. With a slight increase in the number of supernets, the accuracy of few-shot NAS is greatly improved. MetaNAS [116] is the first method that completely integrates meta-learning and traditional NAS. MetaNAS be capable of better initialization parameters with the help of meta-learning ideas. It completely replaces the weighted summation in the DARTS algorithm to reduce different operations, and the experimental results also show that it is more adaptable to more downstream learning tasks.

5.2 Learning Metric Algorithm

Metric learning [117] is different from classical meta-learning, metric learning no longer divides the model into training and testing stages. In many previous papers [4], [5], [3], [118], metric learning is always introduced separately. In our context, metric learning will be explained under the frame of meta-learning. Fig. 11 illustrates one of the most representative learning methods, which is based on a prototype network that has been improved to obtain substantial improvements on a benchmark dataset for classification tasks.

The siamese neural network [119] is a relatively early model in the metric learning. It can be simply regarded as a binary classification problem. The input to the model composes of a set of positive or negative sample pairs, and the model needs to evaluate the similarity of the images during inference stage. Triple loss [120] is another way to deal with more than pairs input in FSL metric learning. Contrary to the Siamese neural network, triple loss requires positive samples, negative samples, and anchor samples to be available at the same time. If training samples are easily distinguished from each other, this would not be beneficial for the model to better learn discriminative features. The hard sample selection technique [121] incorporates the absolute distance between positive sample pairs in addition to considering the relative distance between positive and negative samples. In addition to this, Li et al. [122] revisited the classical triplet network and extended it to a K-tuple network for FSL.

Compared to the Siamese neural network, the prototype network [123] realizes the true meaning of classification. The most significant difference is that the model allows for more data as input. By feature averaging it is feasible to find the most representative sample as a prototype. However, simple feature averaging is easily disturbed by noise. On this basis, many works [124], [125], [126], [127], [128] have explored how to make the distance between prototypes larger and larger. One of the most representative works is the proposal of positive and negative margins [129], which further reduce the over-fitting and enhance the generalization based on maximizing the discriminative ability of the model.

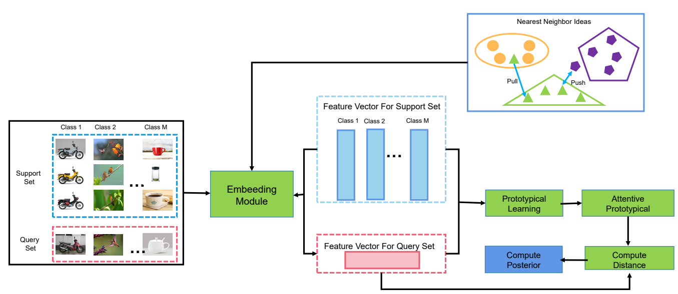

Fig. 11. The framework [128] performs end-to-end learning of the embedding model and prototype learning jointly, and the learned embedding features are used to compute the distance between the query image and the prototype, pushing the distance between different classes farther and bringing the distance between the same classes closer.

Matching Networks [29] is a more general network framework that maps few-shot datasets and unlabeled data to vectors in the embedding space. The matching network combines the best features of parametric and nonparametric models of the nearest neighbour algorithm to model the sample distance distribution by learning the embedding representation. Experiments have proved [130] that embedding propagation produces a smoother embedding manifold. How to learn high-quality embedding representations in a limited time is substantial for improving the model's accuracy. GVSE [131] fuses visual embeddings, semantic embeddings, and gating metrics automatically balances the relative importance of each metric dimension by the model.

Subsequently, Arvind Srinivasan et al. [132] proposed a new architecture to improve Inception-Net, U-Net, Attention U-Net, and Squeeze-Net, which takes the time to generate embedding quality as a cost. The processing based on the embedding representation plays a vital role in FSL.

The relational network [133] differs from the three models mentioned above in that its similarity is calculated by using a neural network. In contrast to the Siamese neural networks and prototype networks, relational networks can be seen as providing a learnable nonlinear classifier for determining relationships. The classifier can be a feature extractor of a pre-trained neural network [122] or a multiple embedded module [134]. The most significant contribution of the relational network is that it breaks away from a single linear metric function and explores the use of an alternative model to generate similarity. Table. 7 categorizes each of the representative metric learning algorithms, comparing their innovations on the original approach.

5.3 Learning To Transmit Information

It is proved that graph neural networks (GNNs) [137] have performed well on relational-based tasks in recent years [138]. Researchers have found that its classes-based transfer of information can work well to help FSL learn to identify new class, while avoiding these classes being dominated by proprietary features. Primarily, early graph neural networks simulate the propagation of weights between different nodes by creating full connections between support and query sets. The nodes can be represented by either a one-hot encoding or an embedding vector, and the connections between nodes can be passed through edges. Given the complexity of graph neural network algorithms, most graph neural networks currently have a shallow number of layers. In order to better accommodate FSL, graph neural networks have been uniquely designed with nodes and edges in recent developments. Fig. 12 shows a recent representative algorithm for FSL of graph neural networks from the perspective of exploring small sample distributions.

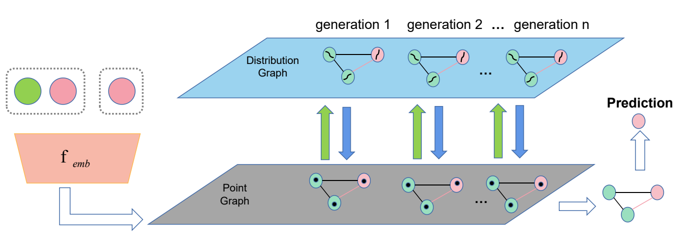

Fig. 12. DPGN [139] is concerned with the relationship between samples in GNN, in addition to the relationship between sample distributions. Where Point Graph is used to describe the samples and Distribution Graph is used to describe the distribution. The two GNNs fuse the instance-level and distribution-level relationships by passing information.

The EGNN uses vertex sets, edge sets, and task sets to encode the labels of nodes. When updates occur between nodes, both the similarity and the difference are considered, which greatly improves the generalization performance of graph neural networks to FSL. Meta-GCN [140] further incorporates the idea of meta-learning, which enables the updating of weights of graphs under FSL to also be optimized according to the gradient descent steps, the whole process requires very few gradient steps and can receive new data quickly. Subsequently, several models based on improvements in the graph structure itself have emerged. The prototype network was improved by GFL [141] network that focuses on learning small samples of data with a graph structure. DPGN uses a dual graph neural network that describes samples while modelling their distribution. Furthermore, GERN [142] uses embedding of graph neural network connections to achieve more robust intra-class weight transfer. Nevertheless, none of these approaches addressed the problem of shallow layers of graph neural networks until 2021, HGNN [143] designed three sections of bottom-to-top, and skip connections to remove the pitfall of ordinary GNNs losing the hierarchical association between nodes. Based on this, Frog-GNN [144] uses multidimensional information to synthesize information about the adjacency between nodes to form pairwise relational features of intra-class similarity and inter-class dissimilarity. At present, graph neural networks are widely used for tasks such as few-shot image classification [145], [146], semantic segmentation [147] and instance segmentation tasks.

5.4 Discussion and Summary

In FSL, meta-learning mainly explore the mapping from the task to the target model. It trains a super-tuning device that gives a good set of hyperparameters as it converges according to the different tasks. In contrast to multi-task learning, which learns only focus on single task. However, meta-learning is not universal for all conditions. The current idea of meta-learning is to have enough historical tasks. If there are not enough tasks on certain problems, then meta-learning may not be able to solve those problems. Similarly, if the domain gap between source and target is too large, the results will also become terrible.

6 Multimodal Complementary Learning Of Small Samples With Limited Information

Until now, FSL has made significant progress in the unimodal domain. Within unimodal learning, models are primarily responsible for representing information as feature vectors that can be processed by a computer or further abstracted into higher-level semantic vectors. Particularly, multimodal learning in FSL refers to learning better feature representations by exploiting complementarities between multiple modalities and removing redundancies between modalities. In real life, when parents teach their babies about things, they always include general information along with semantic descriptions. This is crucial for FSL, which inherently comes with little valid information to make a good evaluation of the data or feature distribution. Inspired by this, many research works [8], [148], [149] consider the introduction of other modal information when solving FSL. By fusing multimodal information, the ability of the model to perceive small sample data can be improved. Fig. 13 shows the main paths of FSL under multimodality,

6.1 Multimodal embedding

Recent works [9], [148], [149], [150], [160] proved the limitations of visual features for FSL of certain tasks. Semantic

TABLE 7
A Summary of Metric Learning by base approach.

Model	Key idea	Metric function	Improvement
Siamese Neural Network [119]	**A pair of inputs**	**Cosine**	**Original**
Triple loss [120]	Triple input	Cosine	Maximize intra-class distance and minimize inter-class distance
E-nagivate sample [121]	Difficult sample training	Cosine	Compare more samples at once
K-tuple network [122]	K inputs	Cosine	Compare more samples at once
Prototype Network [123]	**Prototype representation**	**European**	**Original**
Negative Margin Matters [129]	Negative margin loss	Cosine	Balancing discriminative and migratory
Attentive Prototype [128]	Consider spatial association between features	European	Weighted summation to obtain the prototype
SEN [135]	Feature normalization	European	The modal length of the constraint feature approximates the modal length of the prototype
Prototype Rectification [136]	Modifying prototypes with query sets	Cosine	Consistent distribution of query sets and support sets
Matching Network [29]	**Attention mechanism to access the memory matrix**	**Cosine**	**Original**
GVSE [131]	The relative importance of the automatic balancing model for each metric	Cosine	Fuses visual embeddings, semantic embeddings, and gating metrics
Optimization of image embeddings [132]	Monitoring embedded quality	Manhattan	Improved quality of embedding
Relational Network [133]	**Using models to compute similarity**	**Model**	**Original**
Revisiting metric learning [122]	Using a simple but powerful baseline	non-linear distance	Proposing a deep K-tuplet network
BSNet [134]	Simultaneously by two similarity measures	Euclidean and cosine	Learning feature maps based on the similarity of two different features

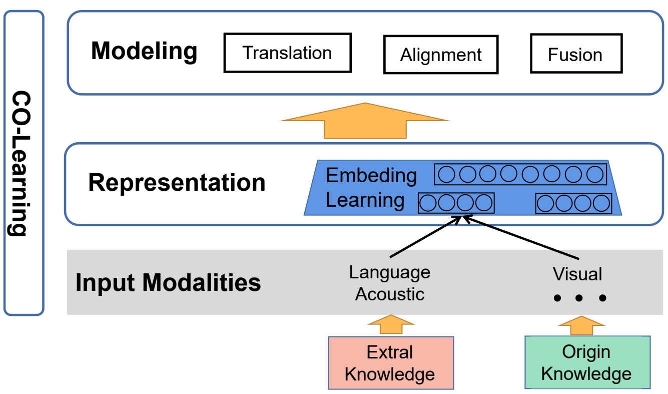

Fig. 13. Multimodal FSL scenarios, how to effectively model other modal information under the condition of feature representation by fusion, alignment, and assistance to compensate for the lack of valid information in itself.

TABLE 8
The challenge of learning from small samples in multimodality.

Approach	Representation	Alignment	Fusion	Co-Learning	Translation
Wang et al. [8]	✔				
Li et al. [148]				✔	
Eli et al. [150]			✔		
Peng et al. [151]	✔				
Schonfeld et al. [149]		✔			
Wang et al. [152]		✔			
Pade et al. [153]	✔				
Fortin et al. [154]				✔	
Zhang et al. [155]	✔				
Sharma et al. [156]					✔
Akata et al. [157]	✔				
Elhoseiny et al. [158]					✔
Zhu et al. [159]				✔	
Xian et al. [21]			✔		
Pahde et al. [153]					✔

space as auxiliary information can provide effective context for visual features and help FSL. Experiments have shown [8], [148], [149] that adaptive combinations of two or more modalities are much better than unimodal FSL. Wang et al. [8] constructed weak semantic supervision for each category by integrating multiple visual features. Schonfeld et al. [149] instead used variational autoencoders (VAEs) to model semantic features based on latent visual features. Subsequently, Schwartz et al. [150] and Peng et al. [151] further extended the semantic information by adding classes labels, attribute and natural language descriptions, and knowledge inference. The additional semantic information is aligned with visual features by embedding loss functions [152] to largely reduce the cost of knowledge transfer. Based on this, Karpathy et al. [160] used multimodal alignment to find potential correspondences that exist between image patches in the training set images and their descriptive utterances. Aoxue et al. [148] went further by using semantic information to model classes as hierarchy.

6.2 Generate semantic information from images

In addition to this, another related area using multimodal FSL is text-to-image generation. In few-shot visual classification tasks, the visual and semantic-based approaches [153], which try to use textual descriptions to generate additional training images, have a considerable advantage. Pade et al. [153] used generative adversarial networks as data generators to train the model, which can purposefully generate corresponding visual features based on semantic information, and enhanced visual features can be obtained by combining the original visual features. Zhu et al. [159] and Xian et al. [21] explored generative images and feature vectors, respectively, making promising progress in the field of ZSL.

Similarly, Fortin et al. [154] migrated text-to-image generation to the target detection task, which can be integrated with current FSL to implement a more general module in the contextual joint learning phase. Zhang et al. [155] improved the resolution of the generated images by concatenating two CGANs based on ordinary generative networks. The first subtask generates a relatively blurred image from the text, and the second subtask generates a high resolution image from the blurred image. Eventually, the model will use more details to generate images. Nevertheless, sometimes text descriptions contain multiple targets and a single text description does not capture all the details in an image. Sharma et al. [156] provided a dialogue interface that uses textual information from the dialogue to obtain more detailed information about the image.

Another set of text-to-image algorithms is based on the variant auto-encoder with embedding. Unlike the generative approach, the input to the encoder is a vector of attributes. Akata et al. [157] explored semantic features from different sources, such as WordNet and word embeddings. However, these methods were unable to recognize parts of an image without part-term annotation. Elhoseiny et al. [158] used a visual classifier to detect patches from bird image dataset by using only text terms and tests without partial annotation. The results show that visual text information and bird parts can be linked with zero samples.

6.3 Discussion and Summary

Multi-modal FSL is still in the developing stage and there are currently several challenges until now: how to combine data from heterogeneous domains, how to deal with the different levels of noise that occur during the combination of different modalities, and how to learn together. Table. 8 classifies FSL tasks in multimodality as representation, alignment, fusion, co-learning, and translation. In a multi-modal FSL, a good feature representation should be able to fill in the missing modalities based on the observed modal information. More approaches will emerge in the future, going well beyond modal embedding and generating semantic information from images.

7 FSL Applications In Computer Vision

In the past five years, we systematically combed and summarized FSL in the field of computer vision [161] and divided tasks into image classification, object detection, semantic segmentation, and instance segmentation. Following is a detailed summary in the form of graphs and tables based on the time dimension. By reading this section, the reader will be able to gain a comprehensive grasp of FSL in the field of computer vision.

7.1 Few-shot Image Classification

Except like Google and Facebook, most researchers in real life do not have access to a large dataset of good quality. In FSL computer vision classification tasks, each task may contain only one or a few samples. Solving few-shot image classification tasks is mainly addressed by data augmentation, transfer learning, meta-learning, and multimodal fusion learning. At present, the top three methods in terms of accuracy are all based on feature augmentation and feature transformation of the backbone model. In this section, we investigates all few-shot image classification models from 2016 to the present and counts the best performance of all models on the mini-ImageNet benchmark dataset. Here we use 5-way-1-shot and 5-way-5-shot as baseline tasks. Table. 9 and Fig. 14 illustrate our investigation results.

7.2 Few-shot Object Detection

Few-Shot Object Detection (FSOD) is the task of detecting rare objects from several samples. There has been a lot of progress in FSL for image classification, but rarely for object detection. At present the evolution of few-shot object detection can be divided into three main camps: Data augmentation, transfer learning, and meta-learning. Out of them, Attention mechanisms plays a pivotal role in small sample target detection. Equally, the issue of slow inference for a few-shot object detections to meet real-time requirements remains serious. The Table. 10 and Fig. 14 are used to show recent advances in object detection in FSL.

7.3 Few-shot Semantic Segmentation

Few-shot semantic segmentation was first proposed in [31] until 2017. And it has been widely used in scenes such as medical images and driverless cars. Unlike traditional semantic segmentation, few-shot semantic segmentation has less pixel annotation information in support data set. To our best knowledge, few-shot semantic segmentation can be broadly classified into supervised semantic segmentation, unsupervised semantic segmentation, and video semantic segmentation. In the machine learning stage, the more classical approach is to use probabilistic mappings as prior knowledge for derivation. In the deep learning phase, a large number of efficient algorithms for segmentation tools have emerged, but these models often require a large number of manual sample annotations. Recently, [201] has made significant improvements to few-shot semantic segmentation by proposing a more concise paradigm where only the classifier is meta-learned and the feature encoding decoder remains trained using a conventional segmentation model. Providing the Table. 16 and Fig. 14 for showing few-shot semantic segmentation.

7.4 Few-Shot Instance Segmentation

In contrast to semantic segmentation, instance segmentation involves identifying each pixel in an image and labelling it separately. Recently, few studies are dealing with the problem of segmenting few samples of instances. Current work still focuses on how to improve R-CNNs using some effective tools. The most recent work [210] proposes an incremental few-shot instance segmentation algorithm, which greatly improves the performance on benchmark data sets. In this section, we survey papers of recent three years on few-shot instance segmentation. Table. 12 and Fig. 14 show the research progress of the few-shot instance segmentation.

8 Future Direction and Opportunities of FSL

Considerable recent work has made promising progress on various task settings for FSL. Nonetheless, for more challenging scenes, both the training and validation data sets are minimal, where the distribution of other data neither helps real samples to be evaluated nor has extensive training data or validation datasets for transfer learning. Moreover, meta-learning also does not have enough tasks to initialize the parameters. With the taxonomy proposed in this survey, in this section we put forward several possible future

TABLE 9
The latest performance of FSL in the image classification tasks of computer vision

Ref	Model	5-way-1-shot accuary	5-way-5-shot accuary	Extra Training Data	Approach Features	Pubilsed Date	Available code
Lee et al. [162]	ESFR	76.84%	84.36%	✔	Early-Stage Feature Reconst -ruction	Jun 2021	✔
Wang et al. [163]	Multi-Task Learning	59.84%	77.72%	✗	Multi-Task Learning and Meta -Learning	Jun 2021	✔
Afham et al. [164]	RS-FSL	65.33%	-	✔	Semantic assistance	Apr 2021	✔
Rizve et al. [165]	Invariance-Equivariance	67.28%	84.78%	✗	Invariant and Equivariant Repres -entations	Apr 2021	✔
Esfandiarpoor et al. [166]	pseudo-shots	73.35%	82.51%	✗	Exploiting Existing Resources	Dec 2020	✔
Chen et al. [167]	MATANet	53.63%	72.67%	✗	Multi-scale Adaptive + Attention	Nov 2020	
Wang et al. [168]	MTUNet	55.03%	56.12%	✗	using backbone model and weight generated	Nov 2020	✔
Khacef et al. [169]	WRN + Self-Organizing Map	71.5%	82.2%	✗	Self-Organizing Maps	Sep 2020	✗
Xue et al. [170]	RCN - ResNet12	57.40%	75.19%	✗	Transfer learning	Sep 2020	✔
Zhong et al. [171]	MCRNET	62.53%	80.34%	✗	Meta-learning	Jul 2020	✔
Ziko et al. [172]	LaplacianShot	75.57%	84.72%	✗	Transductive Laplacian-regularized	Jul 2020	✔
Bateni et al. [173]	Transductive CNAPS + FETI	79.9%	91.5%	✔	Data Augmentation	Jun 2020	✗
Rajasegaran et al. [174]	SKD	67.04%	83.54%	✗	Knowledge Distillation	Jun 2020	✔
Hu et al. [175]	PT+MAP	82.92%	88.82%	✗	Feature Distribution	Jun 2020	✔
Simon et al. [176]	Adaptive Subspace Network	67.09%	81.65%	✗	central block of a dynamic classifier	Jun 2020	✔
Bateni et al. [173]	Transductive CNAPS	55.6%	73.1%	✗	Self-Organizing Maps	Jun 2020	✔
Li et al. [177]	TRAML	67.10%	79.54%	✗	Adaptive Margin Loss	May 2020	✔
Hu et al. [178]	SIB	70.0%	79.2%	✗	Empirical Bayes Transductive	May 2020	✔
Wang et al. [179]	ICI	69.66%	80.11%	✗	Instance Credibility Inference	Apr 2020	✔
Rodríguez et al. [130]	EPNet + SSL	-	88.05%	✔	Embedding Propagation	Apr 2020	✔
Nguyen et al. [180]	SImPa	52.11%	63.87%	✗	PAC-Bayes framework	Mar 2020	✗
Guan et al. [181]	DAPNA	71.88%	84.07%	✔	Domain Adaptation	Feb 2020	✗
Chen et al. [182]	AmdimNet	76.82%	90.98%	✔	Embedding network	Nov 2019	✔
Xu et al. [183]	MetaFun-Attention	64.13%	80.82%	✗	Use similar gradient descent to encode labeled data to predict unlabeled data.	Jau 2020	✔
Liu et al. [184]	MetaOptNet-SVM+Task Aug	65.38%	82.13 %	✔	Embedding Propagation	Nov 2019	✔
Song et al. [185]	ACC + Amphibian	62.21%	80.75%	✔	Uing pre-trained base model to generalize novel model	Nov 2019	✔
Rodríguez et al. [186]	DKT + BNCosSim	62.96%	64%	✗	Learn a kernel that transfers to new tasks	Dec 2019	✔
Mangla et al. [187]	S2M2R	64.93%	83.18%	✔	Embedding Propagation	Apr 2020	✔
Li et al. [188]	LST	70.1%	78.7%	✔	Train a small sample model to predict fake signatures on unlabeled data	Sep 2019	✔
Yoon et al. [189]	TapNet	61.65%	76.36%	✔	Learn the reference vector of each class in different tasks	May 2019	✔
Kim et al. [190]	EGNN + Transduction	-	76.37%	✗	Using graph neural network to model intra-class similarity	May 2019	✔
Ye et al.[191]	feat+	61.72%	78.38%	✗	Set-to-set applied to embedded functions	Jau 2019	✔
Li et al. [192]	DN4	51.24%	71.02%	✗	Use image local descriptors for measurement	Jun 2019	✔
Park et al. [193]	MC2+	55.73%	70.33%	✗	Factorization matrix	Jau 2019	✔

TABLE 10
The latest performance of FSL in the tasks of few-shot object detection during 2019-2021

Ref	Model	10-shot AP/ 30-shot AP	Extra Training Data	Core idea	Approach Taxonomy	Pubilsed Date	Available code
Zhang et al. [26]	Meta-DETR	17.8/22.9	✗	Use semantic alignment to perform specific encoding and feature-independent decoding of images	Feature Reconstruction	Jun 2021	✔
Sun et al. [194]	FSCE	11.1/15.3	✗	Compare proposal coding loss to improve intra-class compactness and inter-class variance	Feature embedding	Jun 2021	✔
Zhu et al. [195]	SSR-FSD	11.3/14.7	✔	Learning Semantic Embeddings Using the Invariance of Semantic Relations	Embedding learning	Mar 2021	✗
Xiao et al. [196]	FsDetView	12.5/14.7	✔	Share the features of the base class and the new class	Meta-learning	Jul 2020	✔
Wu et al. [197]	MPSR	9.8/14.1	✗	Refinement of samples using multi-scale techniques	Data augmentation	Jul 2020	✔
Wang et al. [198]	TFA w/ cos	10/13.7	✗	Fine-tuning the final layer of the detector	Fine-tuning	Mar 2020	✔
Yan et al. [199]	Meta R-CNN	-/12.4	✗	Meta-learning using partial features	Meta-learning	Sep 2019	✔
Wang et al. [200]	MetaDet	7.1/11.3	✔	Prediction of component-specific parameters from several samples	Meta-learning	Sep 2020	✗

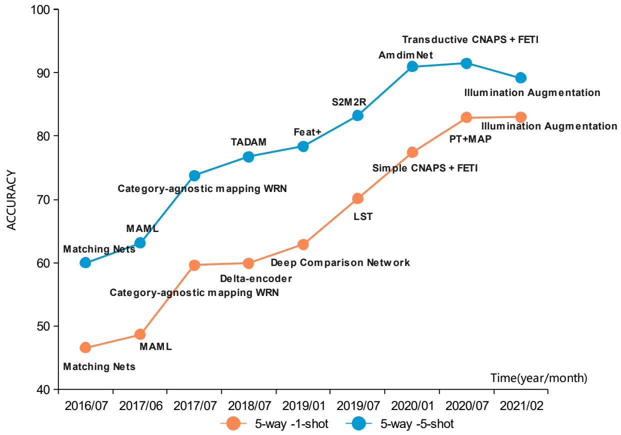

Fig. 14. Best performance of metric learning in image classification tasks during 2017-2021

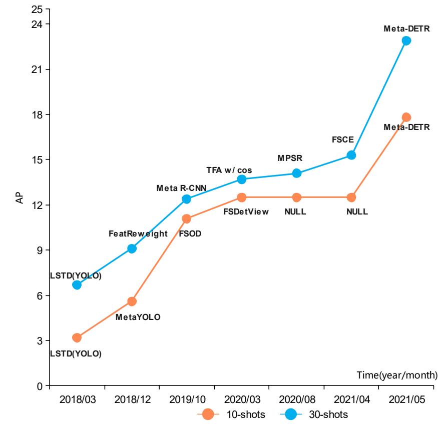

Fig. 15. Best performance of metric learning in object detection tasks during 2017-2021

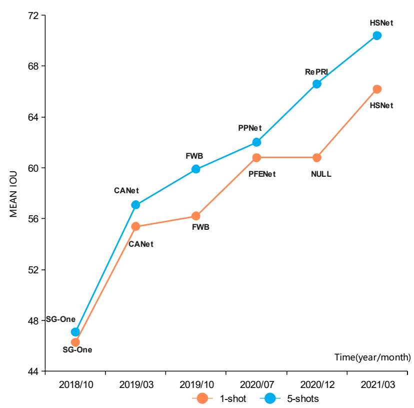

Fig. 16. Best performance of metric learning in semantic segmentation tasks during 2018-2021

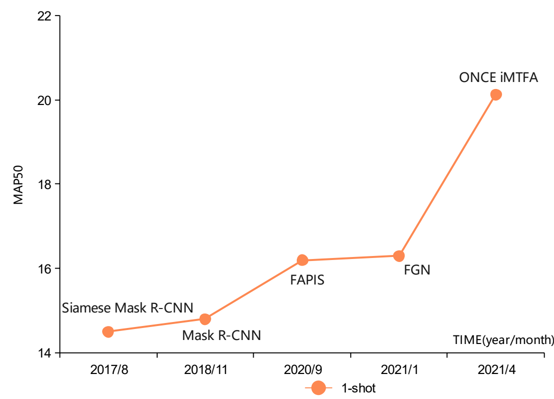

Fig. 17. Best performance of metric learning in instance segmentation tasks during 2017-2021

research directions in FSL. Furthermore, recent advances in applications and algorithms are also presented through this comprehensive survey of FSL.

8.1 Better evaluation of data distribution

The essence of FSL is that the support data sets are too small to evaluate the true data distribution. So what exactly can be done to maximize the evaluation of the true data distribution using a limited number of samples? The latest work [215] is making a useful attempt in this direction, proposing the idea of distribution correction where the mean and covariance of the base class are computationally corrected and then a linear classifier can be used directly to obtain good results. In fact, the difference between FSL and traditional deep learning is not big enough when the few samples are accurate enough to estimate the true data distribution. This is an exciting direction to explore. Similarly, in the field of computer vision, there are no task settings or datasets based on real application scenarios for FSL. Most of the work is still focused on leveraging and mining information from image data. The current mainstream benchmark datasets have more or less various problems: the mini-Imagenet dataset has some inappropriate samples or too difficult samples, such as solid occlusion, multiple objects in the same image, etc. The Omniglot dataset is far away from practical applications and is not easily inspired in real-world applications. BSCD-FSL [36] provides a more violent cross-domain FSL benchmark dataset involving satellite images, medical images. Until now there is no benchmark dataset to evaluate the generalization ability of a model at a fine grained detail. Developing and completing a benchmark dataset in the field of FSL will provide a more realistic evaluation of the current state-of-the-arts in FSL.

8.2 Improving the robustness of data-to-label mapping

A new challenge to FSL is posed by the emergence of BSCD-FSL. Its emergence explores and reveals the limitations of current FSL solutions for cross-domain learning. Recent research has produced some excellent results in this area, such as skillfully designed task tuning, more sophisticated hyperparameter tuning, formation of auxiliary data sets, and extraction of domain-irrelevant features. Currently, fine-tuning is already performing very robustly at the intersection of transfer learning and meta-learning. Nonetheless, both techniques are still very distinct. Pre-training can be seen as learning many categories of tasks, but it is single-task learning. Meta-learning, on the other hand, is a multi-task learning approach. 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

TABLE 11
The latest performance of FSL in the tasks of few-shot semantic segmentation during 2019-2021

Ref	Model	1-shot Mean IoU/ 5-shot Mean IoU	Extra Training Data	Core idea	Approach Taxonomy	Pubilsed Date	Available code
Zhang et al. [202]	CyCTR	64.3/66.6	✗	Aggregate supported and pixel features into query sets	Transfer learning	Jun 2021	✔
Min et al.[203]	HSNet	66.2/70.4	✔	Extracting different sets of feature composition from different levels of intermediate convolutional layers	Feature Engineering	Apr 2021	✔
Yang et al. [204]	RPMM	56.3/-	✗	Different image regions are associated with multiple prototypes to obtain a semantic representation	Metric learning	Sep 2020	✔
Tian et al. [205]	PFENet	60.8/-	✗	Feature enrichment + a priori mask	Feature Engineering	Aug 2020	✔
Liu et al. [206]	PPNet	51.5/62.0	✔	Refinement of samples using multi-scale techniques	GNN	Sep 2020	✔
Nguyen et al. [207]	FWB	56.2/59.9	✗	Measuring the cosine similarity of class feature vectors and query feature vectors	Metric learning	Sep 2019	✔
Wang et al. [208]	PANet	48.1/55.7	✗	Each pixel is compared to the prototype	Metric learning	Feb 2020	✔
Zhang et al. [209]	CANet	55.4/57.1	✗	Support for performing multi -level feature comparisons between images and query images	Metric learning	Mar 2019	✔

TABLE 12
The latest performance of FSL in the tasks of few-shot instance segmentation during 2018-2021.

Ref	Model	1-shot MAP50	Extra Training Data	Core idea	Approach Taxonomy	Pubilsed Date	Available code
Ganea et al. [210]	ONCE iMTFA	20.13	✗	Learning discriminative embedding vectors	Metric learning	Jun 2021	✔
Nguyen et al. [211]	FAPIS	16.3	✗	Modeling of shared parts	Metric learning	Apr 2021	✔
Fan et al. [212]	FGN	16.2	✗	Using attention and relationships to guide generalization	Based model	Sep 2020	✔
He et al. [213]	Mask R-CNN	14.8	✗	Adding masks to predicted objects	Data Augmentation	Aug 2017	✔
Michaelis et al. [214]	Siamese Mask R-CNN	14.5	✗	Encoding reference image subjects	Metric learning	Nov 2018	✔

complexity of meta-learning is a direction worthy of deeper discussion and exchange among researchers at the moment.

8.3 Learn meta-knowledge more effectively from historical tasks

Meta learning is still limited to performance in a specific task space under a defined network structure. In the case of classification tasks, only associations between classification tasks can currently be considered. Is it possible to have a framework that can take into account tasks such as classification, detection, prediction, and generation at the same time? This would enable meta-learning to be somewhat separated from the conception of tasks. Some recent work has attempted to optimize each small batch as a whole. In this case, how to optimize the inner loop will be an important direction of optimization for efficient applications. In the future, pre-training and fine-tuning will become the mainstream algorithms for FSL. At present, meta-learning is still exploring the correlation between tasks, and no relevant theory has yet emerged to explain the causal relationship behind meta-learning. As the causation theory framework evolves, meta-learning would probably tend to become a more general framework.

8.4 Full convergence of multimodal information

Multimodal learning is currently an emerging approach for solving FSL problems by automatically learning small sample tasks in edge scenarios without supervised information and quickly migrating to data from different domains. It is widely regarded as a path exploration from weak AI in limited domains towards general AI. The implementation of pre-training and fine-tuning in multimodal learning scenarios can largely enable the usage of a uniform feature representation across different tasks. For instance, cross-modal understanding, and cross-modal generation. The emergence of multimodal pre-training models can support multiple tasks, generalize across many scenarios, and have a substantial ability to generalize and replicate at scale. Extensive work has been done on fusing two or more types of information, including semantic information. Nonetheless, the main work is still focused on pixels and semantic information, with a relatively single function. In order to effectively address feature reuse under multiple modalities and reduce the cost of data annotation, there is an urgent need for the industry to materialize a powerful pre-trained model involving the fusion of three and more modalities.

9 CONCLUSION

As an important branch of deep learning, few-shot Learning does not require a large amount of data but chooses a softer approach to solve problems, where it can be perfectly integrated with techniques such as transfer learning, meta-learning and data augmentation. In this paper, we provide a comprehensive survey of FSL in the form of questions and answers that easily distinguish the confused concepts and summarize the rich baseline dataset under FSL. Besides, we provide unique insights into the challenges in the development of FSL following a new taxonomy. The evolution of relevant research methods is analyzed in depth according to the degree of integration of knowledge in each stage. 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. Overall, this paper 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] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. arXiv preprint arXiv:1904.04232, 2019.

[2] Carl Doersch, Ankush Gupta, and Andrew Zisserman. Crosstransformers: spatially-aware few-shot transfer. arXiv preprint arXiv:2007.11498, 2020.

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

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

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

[6] Endel Tulving. How many memory systems are there? American psychologist, 40(4):385, 1985.

[7] Endel Tulving. Episodic memory: From mind to brain. Annual review of psychology, 53(1):1–25, 2002.

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

[9] Chen Xing, Negar Rostamzadeh, Boris Oreshkin, and Pedro O O Pinheiro. Adaptive cross-modal few-shot learning. Advances in Neural Information Processing Systems, 32:4847–4857, 2019.

[10] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6):1–35, 2021.

[11] Sakshi Painuly, Sachin Sharma, and Priya Matta. Future trends and challenges in next generation smart application of 5g-iot. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pages 354–357. IEEE, 2021.

[12] Liang Yao, Chengsheng Mao, and Yuan Luo. Kg-bert: Bert for knowledge graph completion. arXiv preprint arXiv:1909.03193, 2019.

[13] Mark Woodward and Chelsea Finn. Active one-shot learning. arXiv preprint arXiv:1702.06559, 2017.

[14] Akshay Mehrotra and Ambedkar Dukkipati. Generative adversarial residual pairwise networks for one shot learning. arXiv preprint arXiv:1703.08033, 2017.

[15] Zitian Chen, Yanwei Fu, Yu-Xiong Wang, Lin Ma, Wei Liu, and Martial Hebert. Image deformation meta-networks for one-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8680–8689, 2019.

[16] Ruslan Salakhutdinov, Joshua Tenenbaum, and Antonio Torralba. One-shot learning with a hierarchical nonparametric bayesian model. In Proceedings of ICML Workshop on Unsupervised and Transfer Learning, pages 195–206. JMLR Workshop and Conference Proceedings, 2012.

[17] Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Sharathchandra Pankanti, Rogerio Feris, Abhishek Kumar, Raja Giries, and Alex M Bronstein. Repmet: Representative-based metric learning for classification and one-shot object detection. arXiv preprint arXiv:1806.04728, 4323, 2018.

[18] Kwangjin Yoon, Jeonghwan Gwak, Young-Min Song, Young-Chul Yoon, and Moon-Gu Jeon. Oneshotda: Online multi-object tracker with one-shot-learning-based data association. IEEE Access, 8:38060–38072, 2020.

[19] Shruti Jadon and Aditya Arcot Srinivasan. Improving siamese networks for one-shot learning using kernel-based activation functions. In Data Management, Analytics and Innovation, pages 353–367. Springer, 2021.

[20] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 951–958. IEEE, 2009.

[21] Yongqin Xian, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. Feature generating networks for zero-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5542–5551, 2018.

[22] Elyor Kodirov, Tao Xiang, and Shaogang Gong. Semantic autoencoder for zero-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3174–3183, 2017.

[23] Muhammad Ferjad Naeem, Yongqin Xian, Federico Tombari, and Zeynep Akata. Learning graph embeddings for compositional zero-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 953–962, 2021.

[24] Joaquin Vanschoren. Meta-learning: A survey. arXiv preprint arXiv:1810.03548, 2018.

[25] Kevin Li, Abhishek Gupta, Ashwin Reddy, Vitchyr H Pong, Aurick Zhou, Justin Yu, and Sergey Levine. Mural: Meta-learning uncertainty-aware rewards for outcome-driven reinforcement learning. In International Conference on Machine Learning, pages 6346–6356. PMLR, 2021.

[26] Gongjie Zhang, Zhipeng Luo, Kaiwen Cui, and Shijian Lu. Meta-detr: Few-shot object detection via unified image-level meta-learning. arXiv preprint arXiv:2103.11731, 2021.

[27] Meng Cheng, Hanli Wang, and Yu Long. Meta-learning based incremental few-shot object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2021.

[28] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

[29] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. Advances in neural information processing systems, 29:3630–3638, 2016.

[30] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. Science, 350(6266):1332–1338, 2015.

[31] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. One-shot learning for semantic segmentation. arXiv preprint arXiv:1709.03410, 2017.

[32] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. arXiv preprint arXiv:1903.03096, 2019.

[33] Xavier Roynard, Jean-Emmanuel Deschaud, and François Goulette. Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. The International Journal of Robotics Research, 37(6):545–557, 2018.

[34] Seong Joon Oh, Kevin Murphy, Jiyan Pan, Joseph Roth, Florian Schroff, and Andrew Gallagher. Modeling uncertainty with hedged instance embedding. arXiv preprint arXiv:1810.00319, 2018.

[35] Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In 2010 IEEE computer society conference on computer vision and pattern recognition, pages 3485–3492. IEEE, 2010.

[36] Yunhui Guo, Noel C Codella, Leonid Karlinsky, James V Codella, John R Smith, Kate Saenko, Tajana Rosing, and Rogerio Feris. A broader study of cross-domain few-shot learning. In European Conference on Computer Vision, pages 124–141. Springer, 2020.

[37] Sharada P Mohanty, David P Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. Frontiers in plant science, 7:1419, 2016.

[38] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7):2217–2226, 2019.

[39] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2097–2106, 2017.

[40] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368, 2019.

[41] Xavier Bouthillier, Kishore Konda, Pascal Vincent, and Roland Memisevic. Dropout as data augmentation. arXiv preprint arXiv:1506.08700, 2015.

[42] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 13001–13008, 2020.

[43] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552, 2017.

[44] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 6023–6032, 2019.

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

[46] Junjie Li, Zilei Wang, and Xiaoming Hu. Learning intact features by erasing-inpainting for few-shot classification. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 8401–8409, 2021.

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

[48] Hang Gao, Zheng Shou, Alireza Zareian, Hanwang Zhang, and Shih-Fu Chang. Low-shot learning via covariance-preserving adversarial augmentation networks. arXiv preprint arXiv:1810.11730, 2018.

[49] Wen-Hsuan Chu, Yu-Jhe Li, Jing-Cheng Chang, and Yu-Chiang Frank Wang. Spot and learn: A maximum-entropy patch sampler for few-shot image classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6251–6260, 2019.

[50] Hongguang Zhang, Jing Zhang, and Piotr Koniusz. Few-shot learning via saliency-guided hallucination of samples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2770–2779, 2019.

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

[52] Zitian Chen, Yanwei Fu, Yinda Zhang, Yu-Gang Jiang, Xiangyang Xue, and Leonid Sigal. Semantic feature augmentation in few-shot learning. arXiv preprint arXiv:1804.05298, 86:89, 2018.

[53] Hiroshi Inoue. Data augmentation by pairing samples for images classification. arXiv preprint arXiv:1801.02929, 2018.

[54] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017.

[55] Li-Qun Chen, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang, and He-Qin Zhou. A visual attention model for adapting images on small displays. Multimedia systems, 9(4):353–364, 2003.

[56] Zitian Chen, Yanwei Fu, Yinda Zhang, Yu-Gang Jiang, Xiangyang Xue, and Leonid Sigal. Multi-level semantic feature augmentation for one-shot learning. IEEE Transactions on Image Processing, 28(9):4594–4605, 2019.

[57] Yonggang Li, Guosheng Hu, Yongtao Wang, Timothy Hospedales, Neil M Robertson, and Yongxin Yang. Dada: Differentiable automatic data augmentation. arXiv preprint arXiv:2003.03780, 2020.

[58] Mingyu Kang, Yifan Yang, Duxin Chen, and Wenwu Yu. Cwgan: A graph vector based traffic missing data adversarial generation approach. In 2020 Chinese Automation Congress (CAC), pages 6234–6238. IEEE, 2020.

[59] Ying Ma, Guoqiang Zhong, Yanan Wang, and Wen Liu. Metacgan: A novel gan model for generating high quality and diversity images with few training data. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2020.

[60] Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Rogerio Feris, Abhishek Kumar, Raja Giryes, and Alex M Bronstein. Delta-encoder: an effective sample synthesis method for few-shot object recognition. arXiv preprint arXiv:1806.04734, 2018.

[61] Cheng-Yuan Liou, Wei-Chen Cheng, Jiun-Wei Liou, and Daw-Ran Liou. Autoencoder for words. Neurocomputing, 139:84–96, 2014.

[62] Junsik Kim, Tae-Hyun Oh, Seokju Lee, Fei Pan, and In So Kweon. Variational prototyping-encoder: One-shot learning with prototypical images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9462–9470, 2019.

[63] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1):43–76, 2020.

[64] I Kevin, Kai Wang, Xiaokang Zhou, Wei Liang, Zheng Yan, and Jinhua She. Federated transfer learning based cross-domain prediction for smart manufacturing. IEEE Transactions on Industrial Informatics, 2021.

[65] Chenjing Cai, Shiwei Wang, Youjun Xu, Weilin Zhang, Ke Tang, Qi Ouyang, Luhua Lai, and Jianfeng Pei. Transfer learning for drug discovery. Journal of Medicinal Chemistry, 63(16):8683–8694, 2020.

[66] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

[67] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[68] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. arXiv preprint arXiv:1802.05365, 2018.

[69] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

[70] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[71] Guneet S Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A baseline for few-shot image classification. arXiv preprint arXiv:1909.02729, 2019.

[72] Yinbo Chen, Xiaolong Wang, Zhuang Liu, Huijuan Xu, and Trevor Darrell. A new meta-baseline for few-shot learning. arXiv preprint arXiv:2003.04390, 2020.

[73] Mahbub Hussain, Jordan J Bird, and Diego R Faria. A study on cnn transfer learning for image classification. In UK Workshop on computational Intelligence, pages 191–202. Springer, 2018.

[74] Akihiro Nakamura and Tatsuya Harada. Revisiting fine-tuning for few-shot learning. arXiv preprint arXiv:1910.00216, 2019.

[75] Alexander R Fabbri, Simeng Han, Haoyuan Li, Haoran Li, Marjan Ghazvininejad, Shafiq Joty, Dragomir Radev, and Yashar Mehdad. Improving zero and few-shot abstractive summarization with intermediate fine-tuning and data augmentation. arXiv preprint arXiv:2010.12836, 2020.

[76] Yann Lifchitz, Yannis Avrithis, Sylvaine Picard, and Andrei Bursuc. Dense classification and implanting for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9258–9267, 2019.

[77] Zhongjie Yu, Lin Chen, Zhongwei Cheng, and Jiebo Luo. Transmatch: A transfer-learning scheme for semi-supervised few-shot learning. In Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition, pages 12856–12864, 2020.
[78] John Cai and Sheng Mei Shen. Cross-domain few-shot learning with meta fine-tuning. arXiv preprint arXiv:2005.10544, 2020.
[79] Zhiqiang Shen, Zechun Liu, Jie Qin, Marios Savvides, and Kwang-Ting Cheng. Partial is better than all: Revisiting fine-tuning strategy for few-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 9594–9602, 2021.
[80] David Argüeso, Artzai Picon, Unai Irusta, Alfonso Medela, Miguel G San-Emeterio, Arantza Bereciartua, and Aitor Alvarez-Gila. Few-shot learning approach for plant disease classification using images taken in the field. Computers and Electronics in Agriculture, 175:105542, 2020.
[81] Anay Majee, Kshitij Agrawal, and Anbumani Subramanian. Few-shot learning for road object detection. arXiv preprint arXiv:2101.12543, 2021.
[82] Ori Ram, Yuval Kirstain, Jonathan Berant, Amir Globerson, and Omer Levy. Few-shot question answering by pretraining span selection. arXiv preprint arXiv:2101.00438, 2021.
[83] Nikita Dvornik, Cordelia Schmid, and Julien Mairal. Selecting relevant features from a multi-domain representation for few-shot classification. In European Conference on Computer Vision, pages 769–786. Springer, 2020.
[84] Davis Wertheimer, Luming Tang, and Bharath Hariharan. Few-shot classification with feature map reconstruction networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8012–8021, 2021.
[85] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. arXiv preprint arXiv:2001.08735, 2020.
[86] Jiamei Sun, Sebastian Lapuschkin, Wojciech Samek, Yunqing Zhao, Ngai-Man Cheung, and Alexander Binder. Explanation-guided training for cross-domain few-shot classification. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 7609–7616. IEEE, 2021.
[87] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. arXiv preprint arXiv:1807.05960, 2018.
[88] Yuqian Fu, Yanwei Fu, and Yu-Gang Jiang. Meta-fdmixup: Cross-domain few-shot learning guided by labeled target data. In Proceedings of the 29th ACM International Conference on Multimedia, pages 5326–5334, 2021.
[89] Cheng Perng Phoo and Bharath Hariharan. Self-training for few-shot transfer across extreme task differences. arXiv preprint arXiv:2010.07734, 2020.
[90] Lu Liu, William Hamilton, Guodong Long, Jing Jiang, and Hugo Larochelle. A universal representation transformer layer for few-shot image classification. arXiv preprint arXiv:2006.11702, 2020.
[91] Wei-Hong Li, Xialei Liu, and Hakan Bilen. Universal representation learning from multiple domains for few-shot classification. arXiv preprint arXiv:2103.13841, 2021.
[92] Hakan Bilen and Andrea Vedaldi. Universal representations: The missing link between faces, text, planktons, and cat breeds. arXiv preprint arXiv:1701.07275, 2017.
[93] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. arXiv preprint arXiv:1705.08045, 2017.
[94] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning, pages 1126–1135. PMLR, 2017.
[95] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. arXiv preprint arXiv:1803.02999, 2018.
[96] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. arXiv preprint arXiv:1909.09157, 2019.
[97] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. arXiv preprint arXiv:1707.09835, 2017.
[98] Muhammad Abdullah Jamal and Guo-Jun Qi. Task agnostic meta-learning for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11719–11727, 2019.
[99] Aravind Rajeswaran, Chelsea Finn, Sham Kakade, and Sergey Levine. Meta-learning with implicit gradients, 2019.
[100] Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Mubarak Shah. itaml: An incremental task-agnostic meta-learning approach. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13588–13597, 2020.
[101] Joel Joseph and Alex Gu. Reproducibility report: La-maml: Look-ahead meta learning for continual learning. arXiv preprint arXiv:2102.05824, 2021.
[102] Anugunj Naman and Liliana Mancini. Fixed-maml for few shot classification in multilingual speech emotion recognition. arXiv preprint arXiv:2101.01356, 2021.
[103] Taewon Jeong and Heeyoung Kim. Ood-maml: Meta-learning for few-shot out-of-distribution detection and classification. Advances in Neural Information Processing Systems, 33, 2020.
[104] Olga Wichrowska, Niru Maheswaranathan, Matthew W Hoffman, Sergio Gomez Colmenarejo, Misha Denil, Nando Freitas, and Jascha Sohl-Dickstein. Learned optimizers that scale and generalize. In International Conference on Machine Learning, pages 3751–3760. PMLR, 2017.
[105] Kiran Kumar Chandriah and Raghavendra V Naraganahalli. Rnn/lstm with modified adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimedia Tools and Applications, pages 1–15, 2021.
[106] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. In Advances in neural information processing systems, pages 3981–3989, 2016.
[107] Rein Houthooft, Richard Y Chen, Phillip Isola, Bradly C Stadie, Filip Wolski, Jonathan Ho, and Pieter Abbeel. Evolved policy gradients. arXiv preprint arXiv:1802.04821, 2018.
[108] Luisa Zintgraf, Kyriacos Shiarlis, Maximilian Igl, Sebastian Schulze, Yarin Gal, Katja Hofmann, and Shimon Whiteson. Varibad: A very good method for bayes-adaptive deep rl via meta-learning. arXiv preprint arXiv:1910.08348, 2019.
[109] AJ Piergiovanni, Anelia Angelova, and Michael S Ryoo. Evolving losses for unsupervised video representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 133–142, 2020.
[110] Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, and Jeff Dean. Efficient neural architecture search via parameters sharing. In International Conference on Machine Learning, pages 4095–4104. PMLR, 2018.
[111] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. arXiv preprint arXiv:1806.09055, 2018.
[112] Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, and Quoc Le. Understanding and simplifying one-shot architecture search. In International Conference on Machine Learning, pages 550–559. PMLR, 2018.
[113] Kaicheng Yu, Christian Sciuto, Martin Jaggi, Claudiu Musat, and Mathieu Salzmann. Evaluating the search phase of neural architecture search. arXiv preprint arXiv:1902.08142, 2019.
[114] Xuanyi Dong and Yi Yang. Nas-bench-201: Extending the scope of reproducible neural architecture search. arXiv preprint arXiv:2001.00326, 2020.
[115] Yiyang Zhao, Linnan Wang, Yuandong Tian, Rodrigo Fonseca, and Tian Guo. Few-shot neural architecture search. In International Conference on Machine Learning, pages 12707–12718. PMLR, 2021.
[116] Thomas Elsken, Benedikt Staffler, Jan Hendrik Metzen, and Frank Hutter. Meta-learning of neural architectures for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12365–12375, 2020.
[117] Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check. In European Conference on Computer Vision, pages 681–699. Springer, 2020.
[118] Nihar Bendre, Hugo Terashima Marín, and Peyman Najafirad. Learning from few samples: A survey. arXiv preprint arXiv:2007.15484, 2020.
[119] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In ICML deep learning workshop, volume 2. Lille, 2015.
[120] Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In International workshop on similarity-based pattern recognition, pages 84–92. Springer, 2015.

[121] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: a deep quadruplet network for person re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 403–412, 2017.

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

[123] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. arXiv preprint arXiv:1703.05175, 2017.

[124] Liu Yang and Rong Jin. Distance metric learning: A comprehensive survey. Michigan State Universiy, 2(2):4, 2006.

[125] Matthew Schultz and Thorsten Joachims. Learning a distance metric from relative comparisons. Advances in neural information processing systems, 16:41–48, 2004.

[126] Shihyen Chen, Bin Ma, and Kaizhong Zhang. On the similarity metric and the distance metric. Theoretical Computer Science, 410(24-25):2365–2376, 2009.

[127] Kilian Q Weinberger, John Blitzer, and Lawrence K Saul. Distance metric learning for large margin nearest neighbor classification. In Advances in neural information processing systems, pages 1473–1480, 2006.

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

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

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

[131] Yan Huang, Yang Long, and Liang Wang. Few-shot image and sentence matching via gated visual-semantic embedding. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 8489–8496, 2019.

[132] Arvind Srinivasan, Aprameya Bharadwaj, Manasa Sathyan, and S Natarajan. Optimization of image embeddings for few shot learning. arXiv preprint arXiv:2004.02034, 2020.

[133] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1199–1208, 2018.

[134] Xiaoxu Li, Jijie Wu, Zhuo Sun, Zhanyu Ma, Jie Cao, and Jing-Hao Xue. Bsnet: Bi-similarity network for few-shot fine-grained image classification. IEEE Transactions on Image Processing, 30:1318–1331, 2020.

[135] Van Nhan Nguyen, Sigurd Løkse, Kristoffer Wickstrøm, Michael Kampffmeyer, Davide Roverso, and Robert Jenssen. Sen: A novel feature normalization dissimilarity measure for prototypical few-shot learning networks. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16, pages 118–134. Springer, 2020.

[136] Jinlu Liu, Liang Song, and Yongqiang Qin. Prototype rectification for few-shot learning. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16, pages 741–756. Springer, 2020.

[137] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1):4–24, 2020.

[138] Victor Garcia and Joan Bruna. Few-shot learning with graph neural networks. arXiv preprint arXiv:1711.04043, 2017.

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

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

[141] Huaxiu Yao, Chuxu Zhang, Ying Wei, Meng Jiang, Suhang Wang, Junzhou Huang, Nitesh Chawla, and Zhenhui Li. Graph few-shot learning via knowledge transfer. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 6656–6663, 2020.

[142] Zhen Liu, Yitong Xia, and Baochang Zhang. Graph embedding relation network for few-shot learning. In 2020 39th Chinese Control Conference (CCC), pages 7328–7334. IEEE, 2020.

[143] Cen Chen, Kenli Li, Wei Wei, Joey Tianyi Zhou, and Zeng Zeng. Hierarchical graph neural networks for few-shot learning. IEEE Transactions on Circuits and Systems for Video Technology, 2021.

[144] Shiyao Xu and Yang Xiang. Frog-gnn: Multi-perspective aggregation based graph neural network for few-shot text classification. Expert Systems with Applications, 176:114795, 2021.

[145] Spyros Gidaris and Nikos Komodakis. Generating classification weights with gnn denoising autoencoders for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21–30, 2019.

[146] Chao Xiong, Wen Li, Yun Liu, and Minghui Wang. Multi-dimensional edge features graph neural network on few-shot image classification. IEEE Signal Processing Letters, 28:573–577, 2021.

[147] Guo-Sen Xie, Jie Liu, Huan Xiong, and Ling Shao. Scale-aware graph neural network for few-shot semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5475–5484, 2021.

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

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

[150] Eli Schwartz, Leonid Karlinsky, Rogerio Feris, Raja Giryes, and Alex M Bronstein. Baby steps towards few-shot learning with multiple semantics. arXiv preprint arXiv:1906.01905, 2019.

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

[152] Xin Wang, Fisher Yu, Ruth Wang, Trevor Darrell, and Joseph E Gonzalez. Tafe-net: Task-aware feature embeddings for low shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1831–1840, 2019.

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

[154] Mathieu Pagé Fortin and Brahim Chaib-Draa. Multimodal sentiment analysis: A multitask learning approach. In ICPRAM, pages 368–376, 2019.

[155] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 5907–5915, 2017.

[156] Shikhar Sharma, Dendi Suhubdy, Vincent Michalski, Samira Ebrahimi Kahou, and Yoshua Bengio. Chatpainter: Improving text to image generation using dialogue. arXiv preprint arXiv:1802.08216, 2018.

[157] Scott Reed, Zeynep Akata, Honglak Lee, and Bernt Schiele. Learning deep representations of fine-grained visual descriptions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 49–58, 2016.

[158] Mohamed Elhoseiny, Yizhe Zhu, Han Zhang, and Ahmed Elgammal. Link the head to the" beak": Zero shot learning from noisy text description at part precision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5640–5649, 2017.

[159] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. A generative adversarial approach for zero-shot learning from noisy texts. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1004–1013, 2018.

[160] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the

IEEE conference on computer vision and pattern recognition, pages 3128–3137, 2015.
[161] Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal. Improved few-shot visual classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14493–14502, 2020.
[162] Dong Hoon Lee and Sae-Young Chung. Unsupervised embedding adaptation via early-stage feature reconstruction for few-shot classification. arXiv preprint arXiv:2106.11486, 2021.
[163] Haoxiang Wang, Han Zhao, and Bo Li. Bridging multi-task learning and meta-learning: Towards efficient training and effective adaptation. arXiv preprint arXiv:2106.09017, 2021.
[164] Mohamed Afham, Salman Khan, Muhammad Haris Khan, Muzammal Naseer, and Fahad Shahbaz Khan. Rich semantics improve few-shot learning. arXiv preprint arXiv:2104.12709, 2021.
[165] Mamshad Nayeem Rizve, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Exploring complementary strengths of invariant and equivariant representations for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10836–10846, 2021.
[166] Reza Esfandiarpoor, Amy Pu, Mohsen Hajabdollahi, and Stephen H Bach. Extended few-shot learning: Exploiting existing resources for novel tasks. arXiv preprint arXiv:2012.07176, 2020.
[167] Haoxing Chen, Huaxiong Li, Yaohui Li, and Chunlin Chen. Multi-scale adaptive task attention network for few-shot learning. arXiv preprint arXiv:2011.14479, 2020.
[168] Bowen Wang, Liangzhi Li, Manisha Verma, Yuta Nakashima, Ryo Kawasaki, and Hajime Nagahara. Match them up: Visually explainable few-shot image classification. arXiv preprint arXiv:2011.12527, 2020.
[169] Lyes Khacef, Vincent Gripon, and Benoît Miramond. Gpu-based self-organizing maps for post-labeled few-shot unsupervised learning. In International Conference on Neural Information Processing, pages 404–416. Springer, 2020.
[170] Zhiyu Xue, Lixin Duan, Wen Li, Lin Chen, and Jiebo Luo. Region comparison network for interpretable few-shot image classification. arXiv preprint arXiv:2009.03558, 2020.
[171] Xian Zhong, Cheng Gu, Wenxin Huang, Lin Li, Shuqin Chen, and Chia-Wen Lin. Complementing representation deficiency in few-shot image classification: A meta-learning approach. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 2677–2684. IEEE, 2021.
[172] Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed. Laplacian regularized few-shot learning. In International Conference on Machine Learning, pages 11660–11670. PMLR, 2020.
[173] Peyman Bateni, Jarred Barber, Jan-Willem van de Meent, and Frank Wood. Enhancing few-shot image classification with unlabelled examples. arXiv preprint arXiv:2006.12245, 2020.
[174] Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Mubarak Shah. Self-supervised knowledge distillation for few-shot learning. arXiv preprint arXiv:2006.09785, 2020.
[175] Yuqing Hu, Vincent Gripon, and Stéphane Pateux. Leveraging the feature distribution in transfer-based few-shot learning. arXiv preprint arXiv:2006.03806, 2020.
[176] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive subspaces for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4136–4145, 2020.
[177] Aoxue Li, Weiran Huang, Xu Lan, Jiashi Feng, Zhenguo Li, and Liwei Wang. Boosting few-shot learning with adaptive margin loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12576–12584, 2020.
[178] Shell Xu Hu, Pablo G Moreno, Yang Xiao, Xi Shen, Guillaume Obozinski, Neil D Lawrence, and Andreas Damianou. Empirical bayes transductive meta-learning with synthetic gradients. arXiv preprint arXiv:2004.12696, 2020.
[179] Yikai Wang, Chengming Xu, Chen Liu, Li Zhang, and Yanwei Fu. Instance credibility inference for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12836–12845, 2020.
[180] Cuong Nguyen, Thanh-Toan Do, and Gustavo Carneiro. Pac-bayesian meta-learning with implicit prior and posterior. arXiv preprint arXiv:2003.02455, 2020.
[181] Jiechao Guan, Zhiwu Lu, Tao Xiang, and Ji-Rong Wen. Few-shot learning as domain adaptation: Algorithm and analysis. arXiv preprint arXiv:2002.02050, 2020.
[182] Da Chen, Yuefeng Chen, Yuhong Li, Feng Mao, Yuan He, and Hui Xue. Self-supervised learning for few-shot image classification. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1745–1749. IEEE, 2021.
[183] Jin Xu, Jean-Francois Ton, Hyunjik Kim, Adam Kosiorek, and Yee Whye Teh. Metafun: Meta-learning with iterative functional updates. In International Conference on Machine Learning, pages 10617–10627. PMLR, 2020.
[184] Jialin Liu, Fei Chao, and Chih-Min Lin. Task augmentation by rotating for meta-learning. arXiv preprint arXiv:2003.00804, 2020.
[185] Liang Song, Jinlu Liu, and Yongqiang Qin. Generalized adaptation for few-shot learning. arXiv preprint arXiv:1911.10807, 2019.
[186] Massimiliano Patacchiola, Jack Turner, Elliot J Crowley, Michael O'Boyle, and Amos Storkey. Bayesian meta-learning for the few-shot setting via deep kernels. 2020.
[187] Puneet Mangla, Nupur Kumari, Abhishek Sinha, Mayank Singh, Balaji Krishnamurthy, and Vineeth N Balasubramanian. Charting the right manifold: Manifold mixup for few-shot learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 2218–2227, 2020.
[188] Xinzhe Li, Qianru Sun, Yaoyao Liu, Qin Zhou, Shibao Zheng, Tat-Seng Chua, and Bernt Schiele. Learning to self-train for semi-supervised few-shot classification. Advances in Neural Information Processing Systems, 32:10276–10286, 2019.
[189] Sung Whan Yoon, Jun Seo, and Jaekyun Moon. Tapnet: Neural network augmented with task-adaptive projection for few-shot learning. In International Conference on Machine Learning, pages 7115–7123. PMLR, 2019.
[190] Jongmin Kim, Taesup Kim, Sungwoong Kim, and Chang D Yoo. Edge-labeling graph neural network for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11–20, 2019.
[191] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8808–8817, 2020.
[192] Wenbin Li, Lei Wang, Jinglin Xu, Jing Huo, Yang Gao, and Jiebo Luo. 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, pages 7260–7268, 2019.
[193] Eunbyung Park and Junier B Oliva. Meta-curvature. arXiv preprint arXiv:1902.03356, 2019.
[194] Bo Sun, Banghuai Li, Shengcai Cai, Ye Yuan, and Chi Zhang. Fsce: Few-shot object detection via contrastive proposal encoding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7352–7362, 2021.
[195] Chenchen Zhu, Fangyi Chen, Uzair Ahmed, Zhiqiang Shen, and Marios Savvides. Semantic relation reasoning for shot-stable few-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8782–8791, 2021.
[196] Yang Xiao and Renaud Marlet. Few-shot object detection and viewpoint estimation for objects in the wild. In European Conference on Computer Vision, pages 192–210. Springer, 2020.
[197] Jiaxi Wu, Songtao Liu, Di Huang, and Yunhong Wang. Multi-scale positive sample refinement for few-shot object detection. In European Conference on Computer Vision, pages 456–472. Springer, 2020.
[198] Xin Wang, Thomas E Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection. arXiv preprint arXiv:2003.06957, 2020.
[199] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. Meta r-cnn: Towards general solver for instance-level low-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9577–9586, 2019.
[200] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Meta-learning to detect rare objects. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9925–9934, 2019.
[201] Zhihe Lu, Sen He, Xiatian Zhu, Li Zhang, Yi-Zhe Song, and Tao Xiang. Simpler is better: Few-shot semantic segmentation with classifier weight transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 8741–8750, 2021.

[202] Gengwei Zhang, Guoliang Kang, Yunchao Wei, and Yi Yang. Few-shot segmentation via cycle-consistent transformer. arXiv preprint arXiv:2106.02320, 2021.

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

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

[205] Zhuotao Tian, Hengshuang Zhao, Michelle Shu, Zhicheng Yang, Ruiyu Li, and Jiaya Jia. Prior guided feature enrichment network for few-shot segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (01):1–1, 2020.

[206] Yongfei Liu, Xiangyi Zhang, Songyang Zhang, and Xuming He. Part-aware prototype network for few-shot semantic segmentation. In European Conference on Computer Vision, pages 142–158. Springer, 2020.

[207] Khoi Nguyen and Sinisa Todorovic. Feature weighting and boosting for few-shot segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 622–631, 2019.

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

[209] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. 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, pages 5217–5226, 2019.

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

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

[212] Zhibo Fan, Jin-Gang Yu, Zhihao Liang, Jiarong Ou, Changxin Gao, Gui-Song Xia, and Yuanqing Li. Fgn: Fully guided network for few-shot instance segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 9172–9181, 2020.

[213] Xiaolin Zhang, Yunchao Wei, Yi Yang, and Thomas S Huang. Sg-one: Similarity guidance network for one-shot semantic segmentation. IEEE transactions on cybernetics, 50(9):3855–3865, 2020.

[214] Claudio Michaelis, Ivan Ustyuzhaninov, Matthias Bethge, and Alexander S Ecker. One-shot instance segmentation. arXiv preprint arXiv:1811.11507, 2018.

[215] Shuo Yang, Lu Liu, and Min Xu. Free lunch for few-shot learning: Distribution calibration. arXiv preprint arXiv:2101.06395, 2021.