Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(1): 679–711.
DOI: 10.3934/mbe.2024029
Received: 19 November 2023
Revised: 05 December 2023
Accepted: 07 December 2023
Published: 19 December 2023

Review

Few-shot learning based on deep learning: A survey

Wu Zeng*and Zheng-ying Xiao

Engineering Training Center, Putian University, Putian 351100, China

* **Correspondence:** Email: wuzeng515@ptu.edu.cn.

Abstract: In recent years, with the development of science and technology, powerful computing devices have been constantly developing. As an important foundation, deep learning (DL) technology has achieved many successes in multiple fields. In addition, the success of deep learning also relies on the support of large-scale datasets, which can provide models with a variety of images. The rich information in these images can help the model learn more about various categories of images, thereby improving the classification performance and generalization ability of the model. However, in real application scenarios, it may be difficult for most tasks to collect a large number of images or enough images for model training, which also restricts the performance of the trained model to a certain extent. Therefore, how to use limited samples to train the model with high performance becomes key. In order to improve this problem, the few-shot learning (FSL) strategy is proposed, which aims to obtain a model with strong performance through a small amount of data. Therefore, FSL can play its advantages in some real scene tasks where a large number of training data cannot be obtained. In this review, we will mainly introduce the FSL methods for image classification based on DL, which are mainly divided into four categories: methods based on data enhancement, metric learning, meta-learning and adding other tasks. First, we introduce some classic and advanced FSL methods in the order of categories. Second, we introduce some datasets that are often used to test the performance of FSL methods and the performance of some classical and advanced FSL methods on two common datasets. Finally, we discuss the current challenges and future prospects in this field.

Keywords: few-shot learning; deep learning; image classification; metric learning; meta-learning; data enhancement

1. Introduction

In recent years, the rapid development of science and technology has helped the continuous iterative updating of various devices in the field of computing. For example, the performances of the central processing unit (CPU) and the graphic processing unit (GPU) become more and more power-

ful with iteration, effectively promoting the progress of deep learning (DL) technology. At present, DL has achieved great success in many application fields, such as image classification [1], object detection [2], deepfakes [3], generative adversarial networks (GANs) [4], natural language processing (NLP) [5], short video event detection [6], video summarization [7], dense video captioning [8], action detection [9], video object segmentation [10] and other fields. In addition, it also has a considerable degree of application, especially in some industrial application fields, such as intelligent evaluation of gear surface degradation [11], cooperative fault identification of rotating machinery [12] and network for bearing fault diagnosis [13]. There are three main reasons for the success of DL technology in these areas. First, computing devices (CPUs, GPUs) with powerful computing performance provide computing power support for the generation and calculation of these tasks. Second, various network architectures with strong feature extraction capabilities, such as AlexNet [14], ResNet [15], MobileNet [16], ShuffleNet [17] and DenseNet [18] have made great achievements in many fields of computation [19, 20, 21], which is also one of the important reasons for DL's achievements. Finally, it is supported by large datasets (for example, the ImageNet dataset [22] contains more than 14 million labeled images, and the COCO dataset [23] contains 2 million labeled images). Most models with high performance are generally trained based on large datasets. The more images each category in the dataset has, the more images about the pose, size ratio, etc. of the target object in that category will be included, and the model will learn more information from it. For example, in an image classification task, which contains the category "monke," if there are more than 10000 images in this category, these images may contain a large number of images about the target object "monkey" with different angles, different poses, different environmental backgrounds, different target size proportions, etc. To a large extent, this can cover as many poses of the target object that may be captured by the camera in the implementation. If the datasets containing a large number of images in these categories are introduced into the model for training, the trained model will have a high probability of strong performance and generalization ability in practical applications. In terms of probability, as long as the number of images in the training set is enough, the higher the probability that the image in the incoming model is similar to an image in the training set in the actual detection and the higher the accuracy of the model for its recognition. To sum up, large datasets are largely the key to whether a model can achieve strong performance. Unfortunately, in the actual application scenario, the vast majority of tasks cannot collect so many images for model training. In addition, even if more images can be collected to participate in model training, it is also a very time-consuming and laborious thing to label these large numbers of images. Obviously, this is not something that some individuals, small groups and small companies can afford. Therefore, in the real scene, most tasks can only collect limited (i.e., insufficient) image data to participate in the training of the final model. In the case of insufficient images in the dataset, using the traditional convolutional neural networks (CNNs) for training may lead to poor classification performance and generalization ability of the final generated model. Therefore, the traditional training strategies and models struggle to meet the needs of most tasks under the condition of small samples, and better learning strategies are needed. Unlike this, humans can learn some key information about the target object from fewer images. For example, a child who has never seen a monkey can learn some important characteristic information about the monkey with only one or a few photos about the monkey, and then the child can quickly recognize the monkey when he sees it again in the zoo or other places. Interestingly, this seems to be an inborn ability of human beings. Inspired by this, the few-shot learning (FSL) strategy [24] was proposed. As the name implies, FSL is the ability to learn the discrim-

inative features about the target object in the image through a small number of images. In addition, the FSL method needs to be strengthened in terms of understanding the internal mechanisms of the model and addressing cross domain learning issues. The goal of FSL is to acquire learning ability similar to human beings. FSL is applied in many fields, such as crop disease identification [25], food classification [26] and component life assessment [27]. Specifically, in the DL based FSL task, for most FSL strategies, the total images are usually divided into the training set, verification set and test set, and these data sets do not duplicate each other. In the training phase, most methods adjust the model parameters in the training set and validation set and then test and evaluate the model performance in the test set. Generally speaking, the training set has the most categories, and the number of categories in the test set is less. In order to better explain the basic process of the FSL, the basic framework of the FSL strategy is given in Figure 1.

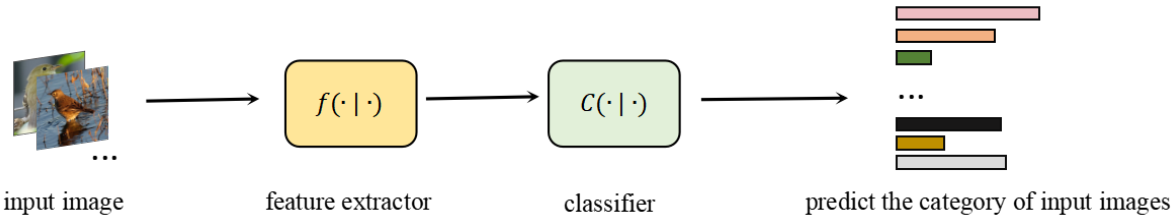


Figure 1. The basic framework of FSL strategy.

In order to optimize or improve some problems of FSL, researchers have proposed many methods about FSL. Wang et al. [28] believed that the core problem of FSL was the unreliability of minimizing the empirical risk. They divided FSL strategies into three major categories, namely, data, models and algorithms. Lu et al. [29] reviewed the research on FSL for a long time (from 2000 to 2019) and mainly emphasized the method based on meta-learning. In addition, the survey also introduced the application of FSL in many fields such as computer vision and natural language. In a relatively new survey (2023), Li et al. [30] mainly investigated deep metric learning in FSL and divided metric learning into three groups, namely, learning feature embeddings, learning class representations and learning distance measures. In this review, we will mainly investigate the small sample image classification methods. For a better introduction, we made Figure 2. We mainly divide them into four categories: FSL methods based on data enhancement, FSL methods based on metric learning, FSL methods based on meta-learning and FSL methods adding other auxiliary tasks. In addition, as shown in Figure 2, we have also subdivided each major category, so that readers can better understand the mechanisms and strategies mainly used in each method.

The general arrangement of the rest of this review is as follows: In section 2, we introduce the basic definition and main classifications of FSL. In section 3, the FSL method based on data enhancement is mainly introduced. In section 4, we introduce the FSL method based on metric learning in detail. In section 5, the FSL method based on meta-learning is introduced. In section 6, we introduce some FSL methods based on other strategies. In section 7, we give a summary of the FSL method at this stage. In section 8, the main datasets and evaluation indexes of small sample image classification are introduced, along with the performance of some classical and advanced methods in some datasets. In

section 9, we explain the current achievements, challenges and prospects for the future. In section 10, we summarize this review.

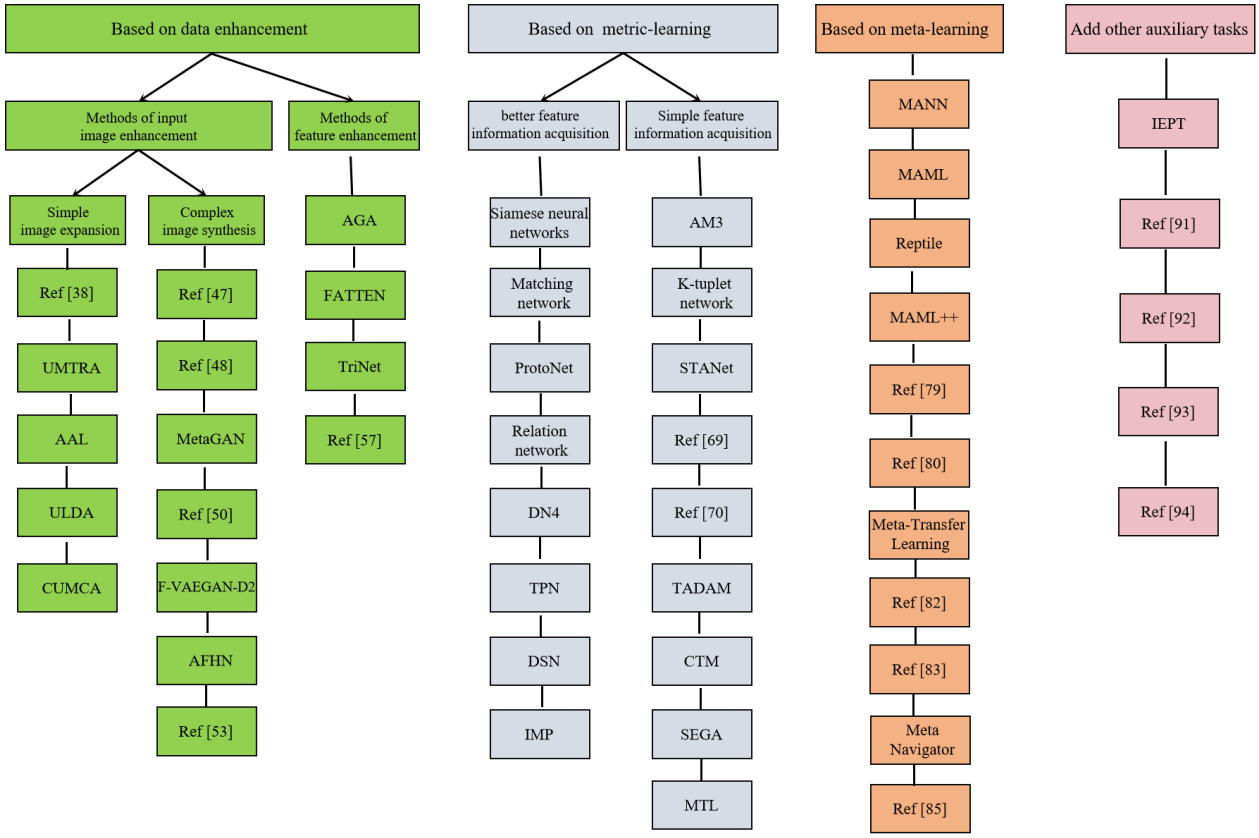


Figure 2. Various classic and advanced FSL methods.

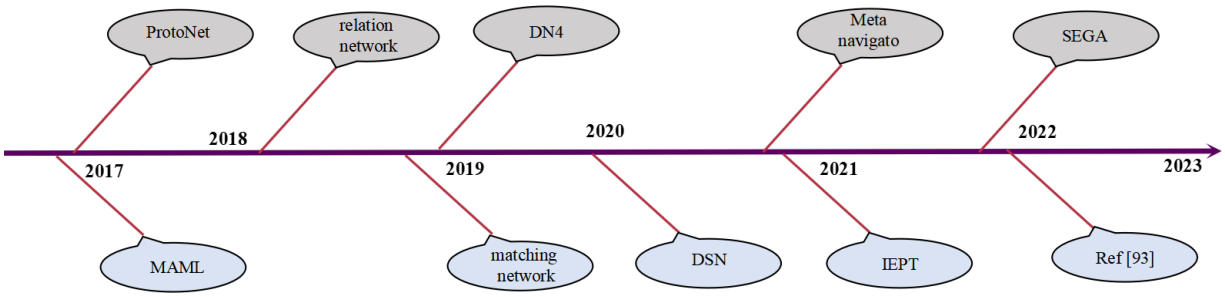


Figure 3. Some mainstream and advanced FSL methods in recent years.

In addition, in Figure 3, we also summarize some mainstream and advanced methods in the FSL field in recent years, which are given in the order of publication timeline, so as to better understand the development of FSL methods in recent years. These listed classical methods and mainstream

methods are either highly representative (their ideas are quoted and improved by most of the subsequent methods) or the methods with more powerful model performance.

2. Basic definition and main classification of FSL

In order to better introduce FSL, we will first explain the basic definition of FSL in Section 2.1. Subsequently, FSL strategies based on data augmentation, metric learning and meta-learning are introduced in section 2.2, section 2.3 and section 2.4, respectively.

2.1. Basic definition of FSL

Generally speaking, in an FSL task, meta-learning training is commonly used. Specifically, the datasets used for small sample image classification tasks are divided into the base set D_{base} and the new set D_{novel} (it should be noted that the images of these two datasets do not have overlapping parts, that is, their images do not have duplicate parts). Follow the same N-way K-shot training method in the same FSL task. Each meta-task T package contains a support set S (support set) and a query set Q (query set), and the meta-task aims to classify images in query set through the given support set. Specifically, in an N-way K-shot small sample task, N classes are first sampled from the training set and marked as C, and then K support samples and q query samples are extracted from these N classes respectively. Support set S, query set Q and meta-learning task T can be defined as:

$$S = \{(x_i, y_i)|y_i \in C, i = 1, 2, ..., N \times K\} \tag{2.1}$$

$$Q = \{(x_i, y_i)|y_i \in C, i = 1, 2, ..., N \times q\} \tag{2.2}$$

$$T = \{(S_i, Q_i)\}_{i=1}^{m} \tag{2.3}$$

Among them, x_i represents the image sample, and y_i represents its image label. m represents the number of meta-task learning T. In order to better illustrate its style, the FSL task description of 2-way 2-shot is shown in Figure 4. Figure 4 shows the random extraction of 2 classes from the dataset, with each class extracting 2 samples as a support set and an additional 2 samples as a query set.

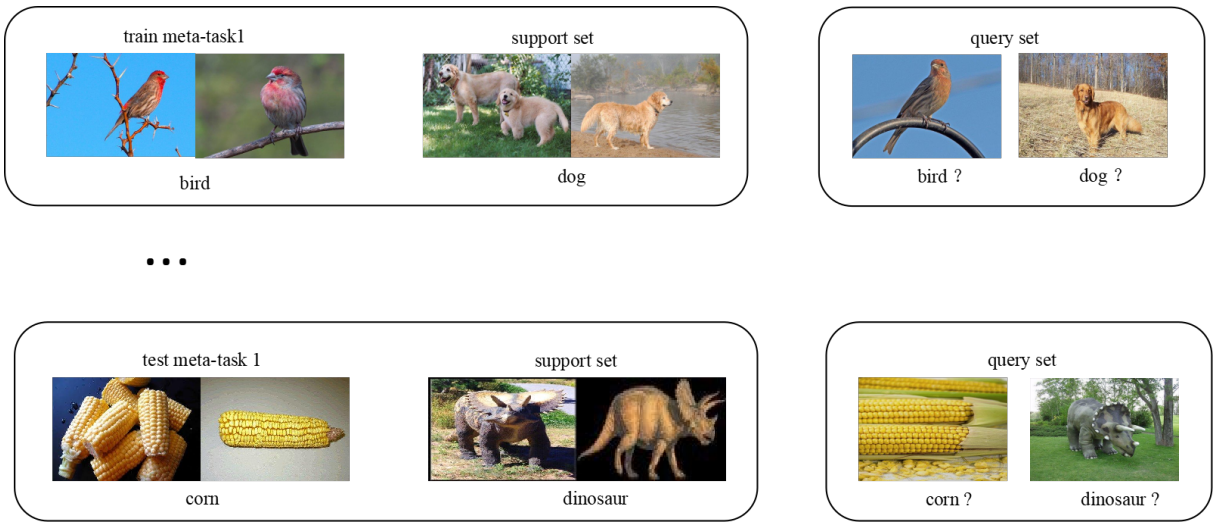


Figure 4. The FSL task description of 2-way 2-shot.

2.2. *Basic introduction to data enhancement strategies*

By using the name FSL, we can have a rough understanding of it, which is the task of lacking training samples (i.e., insufficient samples). Generally speaking, in many DL tasks, the best way to face such problems is to use data augmentation methods [31, 32, 33]. Image data augmentation technology can generate new samples with certain differences from the original image, expand the samples and thus improve model performance [34, 35]. Simple data augmentation methods include rotation, random cropping and translation. These methods are simple and practical, but their generated samples are too similar to the original image, so their effectiveness is limited. In order to better illustrate it, a rough flowchart of generating samples through simple data augmentation strategies in some methods is shown in Figure 5.

In addition, the popular GAN technology in recent years [36, 37] can generate realistic false samples by continuously playing games with generators and discriminators in the model. After continuous efforts by researchers in recent years, this type of method has introduced many high-performance variants that can generate richer backgrounds. The approximate process of generating new samples using the GAN method is shown in Figure 6. This type of method is also a powerful helper in promoting the progress of FSL based on data augmentation methods. Overall, the strategy of data augmentation mainly involves using data augmentation techniques to create new samples, expanding an original image into multiple ones and then obtaining new additional feature information from these images, thereby playing an auxiliary role. However, from the actual situation, although this type of method has played a certain role, its effectiveness is limited.

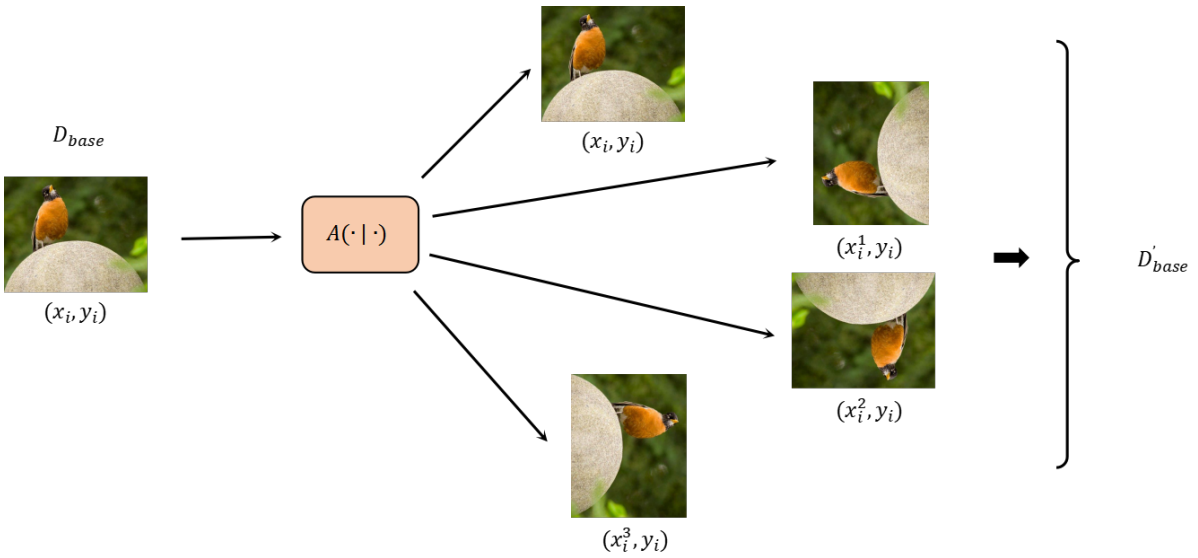


Figure 5. Simple data augmentation strategies.

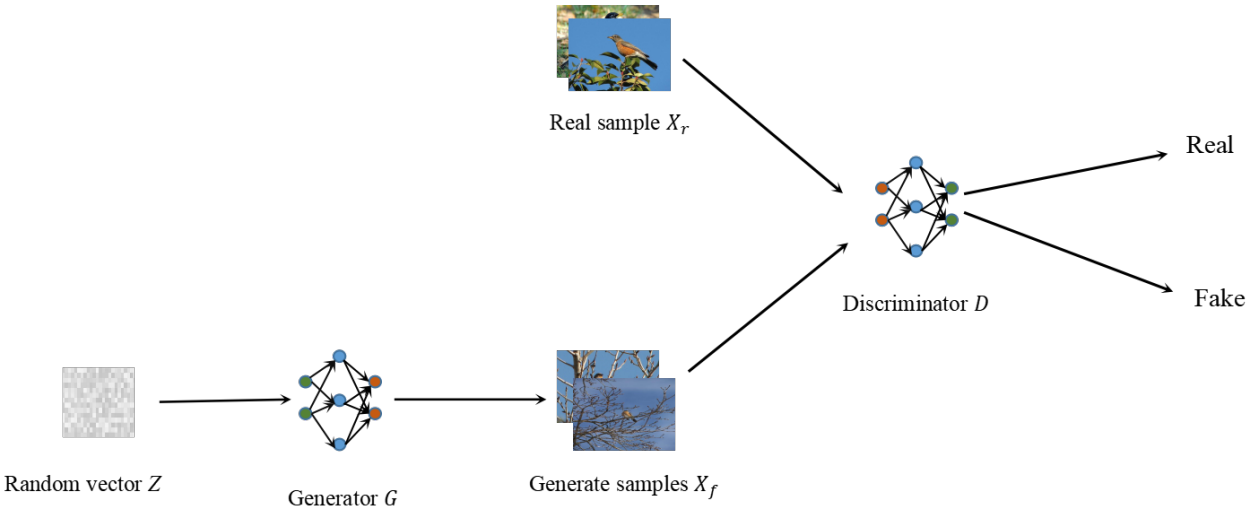


Figure 6. The approximate process of generating new samples using GANs method.

2.3. Basic introduction to metric learning strategies

For most FSL methods based on metric learning, the core idea is to determine the similarity between samples. The higher the similarity between samples, the closer the categories between the two are, and we use this as a benchmark for judgment. In order to better illustrate this type of strategy, in Figure 7, we provide the basic architecture of this type of strategy. From Figure 7, we can roughly observe that the feature extraction module will perform feature extraction processing on the input samples. Next, the model will input the feature vectors into an embedding space (it should be noted that the

purpose of model training is to make samples of the same categories closer to each other and samples of different categories farther apart). Afterwards, the distance between unlabeled query set samples and support set samples is calculated through the measurement module (where the distance function in the measurement module can be cosine function or Euclidean distance, etc.), and the unlabeled samples are ultimately classified as the support set class closest to their distance.

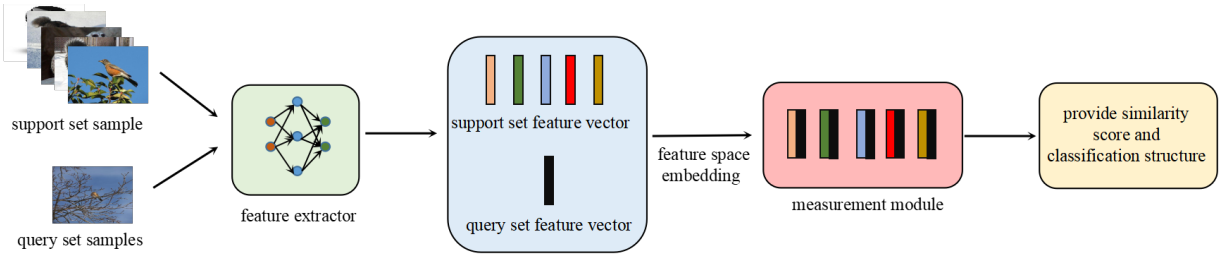


Figure 7. The basic model based on metric learning.

2.4. Basic introduction to meta-learning strategies

Generally speaking, the method based on meta-learning is learning to learn, and its main proposition is cross task learning. The general idea is shown in Figure 8. Based on different tasks, provide better learning parameters that are suitable for the current task, thereby guiding the model to learn quickly in new tasks.

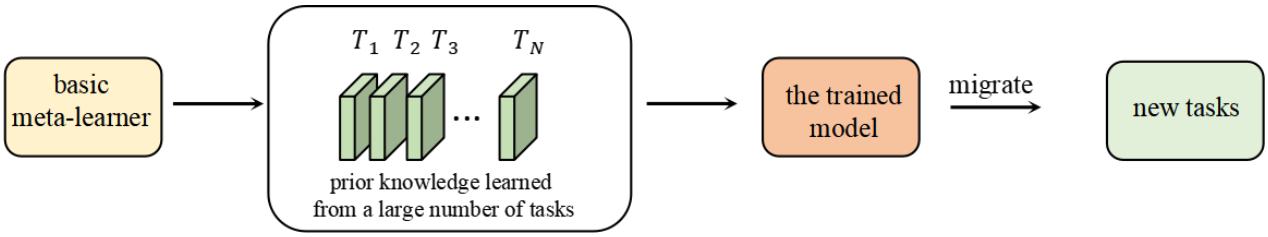


Figure 8. The basic model based in meta-learning.

3. FSL method based on data enhancement strategy

We explore the differences between different methods through extensive research. We first classify FSL methods based on data augmentation into two main categories, namely, "input image enhancement" and "feature enhancement," based on whether additional auxiliary images are generated using the original input samples as the primary difference. Second, based on the differences between the generated image and the original image (i.e., whether the generated sample has a significant difference from the original image), we further classify "input image enhancement" into two categories: "simple image amplification" and "complex image synthesis."

3.1. Methods based on input image enhancement

3.1.1. Simple image amplification

Chen et al. [38] proposed an image enhancement framework combining meta-learning to enhance the diversity of input samples. In order to increase the diversity of the input image, they use artifacts, cropping and replacing between paired images (similar to the strategy of cropping and pasting some blocks into other images in CutMix [39]) and random noise to generate new samples. At the same time, in order to not change the important content information in the original image, the authors use probe images to preserve important areas in the original image. Afterwards, in order to maintain the linearity of the generated samples, a simple parameterization method is used to linearly generate new samples. The experimental results indicate that the method achieved very competitive performance at that time. Afterwards, Khodadeh et al. [40] proposed the UMTRA (unsupervised meta-learning with tasks constructed by random sampling and augmentation) method, which is an unsupervised algorithm based on meta-learning, in the face of FSL challenges. In order to expand the learning samples during the training phase, UMTRA uses random sampling and random amplification methods to enhance the number of samples. This method only requires the use of labeled samples during the final feature learning and induction of the target object, and even the number of labeled samples can only be one. Compared to some traditional meta-learning algorithms, the number of labels required is greatly reduced.

Antonio et al. [41] proposed the AAL (assume, augment and learn) method to address the issue of insufficient sample data in FSL tasks. The core idea of this method is to expand the sample size of the support set through data augmentation methods to address the problem of insufficient image quantity in the support set. Specifically, the first step is to randomly generate a subset in an unlabeled dataset. Afterwards, data augmentation operations are used to expand the image samples in the subset and clustering operations are performed on the regenerated support set to obtain a target set. Finally, the meta-learning strategy FSL model can be used for training. Qin et al. [42] developed a framework called ULDA (unsupervised feed shot learning via distribution shift based data augmentation) that utilizes distributed and data augmentation strategies. The approximate architecture is shown in Figure 9. When implementing data augmentation strategies, this method always pays attention to the distribution of samples in each learning subtask, in order to improve the degree of overfitting that may occur in the model. The ULDA method can have a positive impact on the overall classification performance of the model by only using simple data augmentation methods such as image rotation and random cropping. Xu et al. [43] proposed the CUMCA (unsupervised meta-learning with tasks constructed by random sampling and augmentation) method to distinguish key categories in tasks. This method first constructs embedding tasks based on clustering methods and data augmentation. At the same time, in order to avoid shortcomings such as poor diversity of generated images caused by simple data augmentation methods, the author adopted different strategies in the training process of MAML's (model-agnostic meta-learning) inner and outer loops and provided a theoretical analysis as support. Then, in order to better improve the overall performance of the model, a data enhancement method called Prior-Mixup was proposed, which is better than simply flipping, cropping and rotating the image.

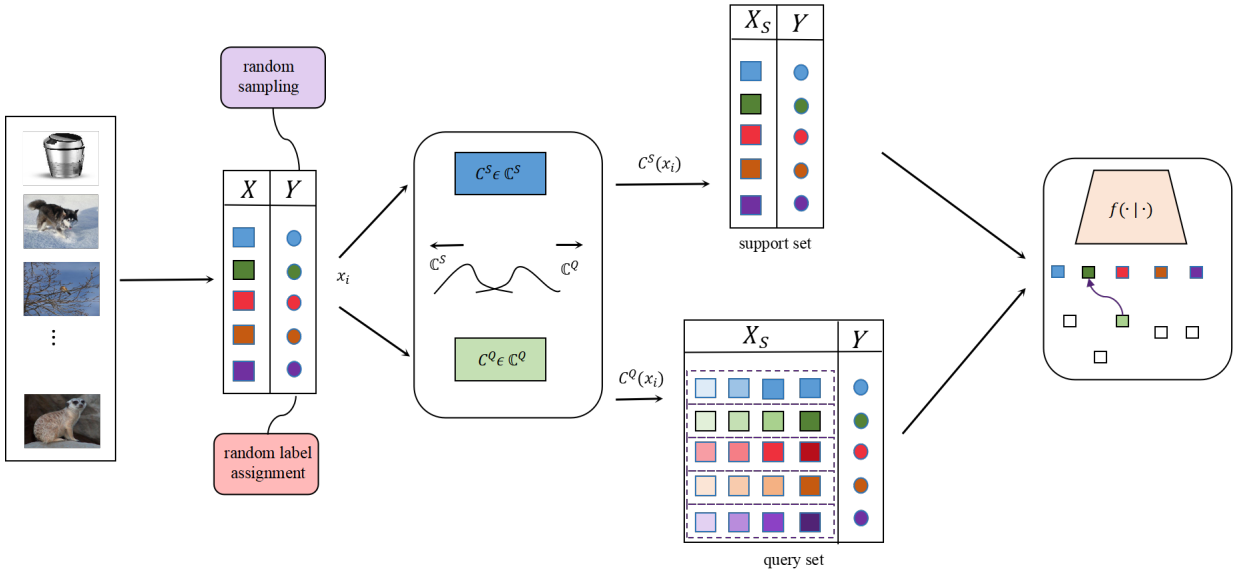


Figure 9. The architecture of the ULDA.

3.1.2. Complex image synthesis

Although the input image can be enhanced by a simple image data enhancement method, the effect is limited because it is similar to the original image. While some methods based on GANs [44, 45, 46] and others can generate images that are different from the original image, some researchers have introduced it into FSL and achieved good results.

Mehrotra and Dukkipati [47] proposed the method of combining residual network and GAN to improve the problems encountered in one shot learning. In addition, they also used the trainable distance measure in the one shot task and achieved competitive results in multiple data sets at that time. Wang et al. [48] designed a model using GAN to generate new samples. The model can use existing samples to create new virtual samples, and the background of the generated samples is quite different from the original samples. Specifically, the model generates new samples by introducing noise and existing samples into the generator. Then, the hallucination module is embedded into the meta-learning task to improve the performance of the model. In general, the background of the image generated by this method is relatively complex, which is quite different from the original image. Zhang et al. [49] introduced the MetaGAN method based on GAN. Slightly different from the above methods, the MetaGAN in this method not only needs to distinguish different categories of samples, but also needs to be used to distinguish the authenticity of the incoming image. In other words, this method can be regarded as an auxiliary task. If the model can accurately distinguish the authenticity of samples (that is, if it is a real sample or a false sample), then the model must learn to distinguish the differences between the details of these images. This also helps the model to better learn the content information of the image, so as to promote the performance of the classifier.

Schwartz et al. [50] proposed a sample synthesis method called Deltaencoder to solve the problem of object recognition with few samples. This method only needs a few samples and can synthesize new samples through unknown categories. Then, these newly synthesized samples are introduced into the

training model for learning. This method can extract the transferable intraclass deformation features between similar sample pairs and then transfer them to a new category. Xian et al. [51] proposed the F-VAEGAN-D2 method by combining the advantages of GAN and VAE, which can realize any shot learning. F-VAEGAN-D2 mainly synthesizes image feature vectors related to CNN from class features. In addition, this method also adds constraints to the feature generation model, enabling the learning model to synthesize the image feature vectors obtained from CNN in the category features. Moreover, the discriminator chosen by the authors can use samples that have not been learned (samples that have not been previously introduced into the model) to promote model performance improvement. Experiments on multiple datasets have shown that this method has excellent FSL performance and generalization ability. Li et al. [52] proposed the AFHN (advanced feature hallucination networks) method based on the conditional Wasserstein GAN to address the FSL challenge. Compared to most GAN methods, CWGAN can to some extent improve the stability of model training by improving the objective function in the model. In addition, the authors of this paper believe that simple data augmentation may not achieve the desired results, so the AFHN method introduces two novel regulators (i.e., classification regulators and the anti-collapse regulators) in GAN to improve the quality of virtual images generated by the model, mainly reflected in the richness of image features.

In addition to the above methods, Pahde et al. [53] used the idea of cross-modal hallucination to solve the FSL problem. They used other modes to add additional conditions, so as to generate more specific images. Specifically, text information is input as an additional control condition, and then a new virtual image is generated. The experimental results on the cub dataset show that it has strong robustness.

3.2. Methods based on feature enhancement

Most of the above two methods are to generate extended samples based on the original image and enhance the diversity of samples by increasing the number of image samples. In addition, another method is to enhance the diversity of samples by enhancing the eigenvector of input samples. Studies have shown that the performance of the model can also be improved by expanding the feature representation of input samples.

Dixit et al. [54] proposed the AGA (attribute guided augmentation) method to improve the problem of target detection in FSL. Specifically, the CNN feature extraction module in this method will first extract the features of the incoming image and then transfer it to a feature space, where the features are enhanced to make the attributes of these features meet our expectations. Liu et al. [55] proposed the FATTEN (feature transfer network) method by using the "encoder-decoder" strategy. Unlike other direct feature extraction methods for images, the predictor included in FATTEN can parameterize the attitude of input samples. Then, use the encoder to map the appearance and pose of the target object. Finally, we use the decoder to generate the feature vector of the object we need. Experimental results show that this method has achieved good results in the few-shot object recognition task.

Chen et al. [56] did not simply enhance and amplify the input image, but enhanced the input feature vector of the image through semantic enhancement. Specifically, the TriNet method they proposed first uses a multi-layer CNN extraction network to extract the basic features of the initial image. After that, the encoder is used to project the extracted basic feature vector into the semantic space for semantic enhancement. Then, it is projected back to the original feature space through the decoder. The method has been widely tested in multiple datasets of FSL, and the experimental results show that the

method has very excellent performance. In order to generate richer images, Zhang et al. [57] obtained the foreground and background of the input image by introducing the saliency detection technology. Important regions in saliency detection can be regarded as important target objects in the image and can be regarded as foreground, while non-important regions can be regarded as background. Then, the feature vectors of the foreground and background are extracted respectively by using the feature extraction network. Finally, the feature vectors of foreground and background are randomly combined to generate new virtual samples, so as to enhance the diversity of feature information.

4. FSL methods based on metric learning

Through the brief introduction in Section 2.3, we can find that the quality of the feature information in the incoming measurement module largely determines the accuracy of the classification module. Specifically, in the feature extraction phase of the image, it is extremely important to accurately obtain the important features in the image. In an image, there are many interferences (such as background interference in the image, interference with different proportions of the target object size, different poses of the target object, etc.). If these interferences can be alleviated, and the important information in the image can be extracted more accurately, it can undoubtedly obtain better feature vector information. Therefore, we mainly divide the metric learning methods into two categories, namely, "methods based on simple feature information acquisition" and "methods based on better feature information acquisition".

4.1. Methods based on simple feature information acquisition

One important purpose of FSL methods based on metric learning is to obtain more accurate information about input samples, and accurately extracting the features of input samples is an extremely important step. In 2015, Korch et al. [58] proposed siamese neural networks. The core strategy of this method is to use the basic CNN feature extraction network to extract the features of the input samples and then map them to the feature space. Then, the similarity between samples is detected by constructing a twin feature metric network architecture (which shares the same parameters). Specifically, the paired sample characteristic vectors are introduced into the metric network to obtain their similarity scores. It should be noted that in the one-shot task, each sample of the query set needs to be tested in pairs to determine the final similarity between samples. Vinyals et al. [59] proposed the matching networks to improve the problems encountered in FSL, which is also used to handle one shot problems in FSL tasks. After feature extraction of image samples, this method will be transmitted by long short-term memory (LSTM) network to a low-latitude feature space for similarity comparison between image samples. In addition, the attention mechanism is used in the similarity comparison of the method. The main function of the attention mechanism is to allocate different weight scores to regions with different importance, so that the method can better calculate the similarity between images. In 2017, Snell et al. [60] proposed the classic prototypical networks (ProtoNet). The authors believe that there is a prototype point in each category of input and then only need to calculate the distance between each query set sample and these prototype points to reflect the similarity between them. First, the model will extract the features of the incoming image. Then, the extracted features are mapped to a specific embedded space. The purpose of the learning module is to make the closer together samples in the same category and the farther apart samples in different categories. Then, calculate the prototype

point between each category, which can also be called the center point (which is generally shown in Figure 10). After that, the distance between the feature vector of the query set sample and the center point of each class is measured and calculated. The smaller the distance is, the closer the categories are to each other. Finally, the sample category is attributed to the nearest support set category. Once this method was introduced, it caused heated discussion and achieved certain results. However, because the link of feature extraction for samples is relatively simple, the link of feature representation for samples is not well done, but its idea of prototype points is extremely worth learning, and much subsequent work is based on it.

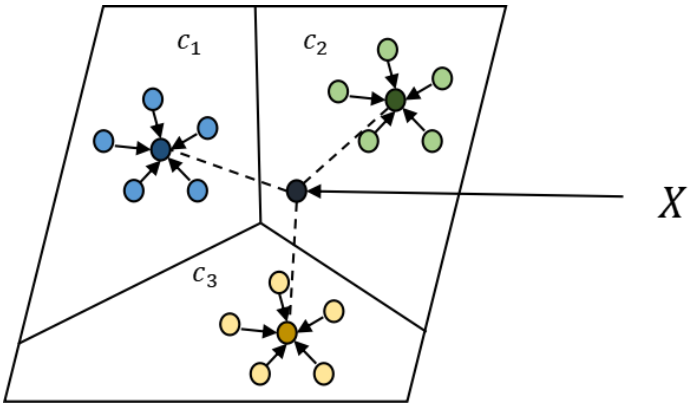


Figure 10. Schematic diagram of the rough strategy for prototype points in the ProtoNet method.

After that, Sung et al. [61] responded to the FSL task according to the improved similarity measurement. Unlike most strategies for similarity calculation based on metric functions (cosine distance and Euclidean distance), the relation network model proposed in this paper uses the network architecture to retrain a learnable nonlinear similarity metric function and use it to obtain the similarity score between samples. This method has achieved good classification accuracy in multiple data sets. Li et al. [62] believed that the local features in the image were extremely important information content, so they proposed a method called DN4 (deep nearest neighbor neural network). This method focuses on the description of the local feature information in the image. This method extracts the feature of each position in the input image and uses it as the input of the subsequent classifier. Then, use the module named deep local feature to describe the extracted features of each position. Finally, the "image-to-class" metric learning strategy is used to measure the similarity between samples, and the measured similarity score is used to determine the category of unknown samples. Liu et al. [63] proposed the TPN (transductive propagation network) method in order to improve the poor generalization of some FSL methods, and its general architecture is shown in Figure 11. This method first extracts the features of the input samples and maps them to a specific embedding space, which is similar to some metric learning methods. Then, in the class structure constructed by the support set and query set, obtain the manifold structure in these sample categories. Furthermore, the method will propagate the iterative labels of the support set to the query set samples according to the graph structure, so as to achieve the purpose of label propagation and measure the category similarity between the query set samples and

the support set samples. Simon et al. [64] proposed a subspace based FSL framework, namely, DSN (deep subspace network). This method first projects the feature vectors extracted from the feature extractor into a subspace, which can make the data samples with similar feature vectors closer. DSN can capture the complex relationship between samples and project it back to the low-latitude space after reducing the dimension. In addition, in order to adapt to tasks with different dimensions, this method can also adaptively shrink the dimension size of the subspace. After that, they proposed DSN-MR (DSN mean refinement) by evenly refining the central points between various classes.

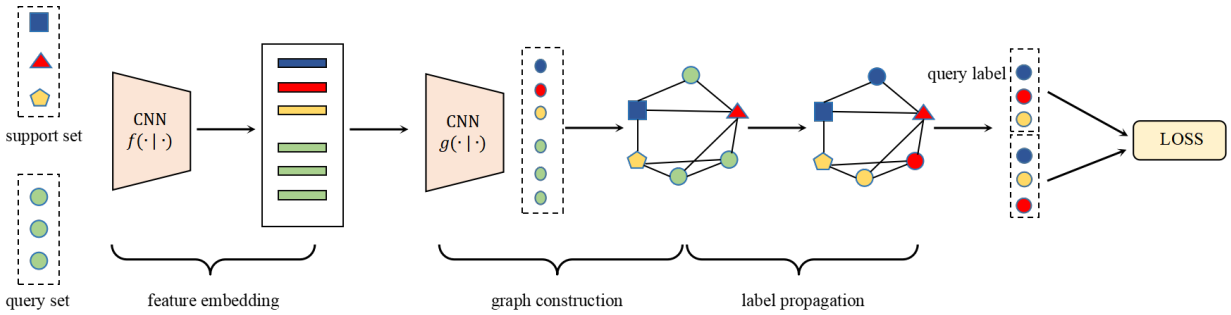


Figure 11. The architecture of the TPN.

Unlike the ProtoNet method, which only takes the average value between each class to get the class center point, the IMP (infinite mixture prototypes) method proposed by Allen et al. [65] makes a more detailed treatment. Specifically, the method is improved based on the ProtoNet method using metric learning. IMP first adds the strategy of infinite hybrid model to ProtoNet and strategically weights the distribution of samples among categories. Then, interpolation between the nearest neighbor and the prototype points is performed to obtain better performance and robustness of the model.

4.2. Methods based on better feature information acquisition

Although most of the methods in Section 4.1 have achieved some results, there are still some shortcomings. For example, in the feature extraction stage, it is impossible to obtain the feature information about the input samples as finely as possible, but it is extremely important to obtain more accurate feature schools about the input samples. In FSL, due to the lack of tag data, it is undoubtedly a pity that the input data cannot be extracted more accurately and optimally. Therefore, it is necessary and important to optimize the feature extraction of input samples. In order to improve this deficiency, some researchers have proposed a variety of methods to optimize the feature extraction of input samples. In the face of insufficient input samples, Xing et al. [66] proposed the AM3 (adaptive modality mixture mechanism) method to improve this problem. On the basis of ProtoNet, AM3 introduces the idea of cross-modality. In short, most methods only extract features from the visual information of input samples, while this method will extract additional semantic information in the image and design a module to combine them adaptively. Thus, more feature information about the sample is obtained from the limited sample, and the feature expression is enhanced. Experimental results show that this method has a greater performance improvement than the baseline method. Li et al. [67] proposed the K-tuplet network to improve the deficiencies in ProtoNet. The authors mainly improved it from three aspects.

First, a more effective similarity measurement strategy is obtained by improving the loss function in similarity measurement. Second, an additional category embedding vector module is added to obtain an additional feature vector about the input sample category information. Vectors of the same category can be introduced into the feature space of the same embedded vector, so as to achieve a better feature representation of the input sample. Finally, dynamic matching networks are used to adaptively adjust the measurement similarity according to the different task requirements, so that the model can reach a better state. Yan et al. [68] proposed a semantic embedded dual attention network model, STANet (spatial task attention network). By establishing an efficient dual attention network architecture, STANet uses an attention module to distinguish the important features of the local area of a single image sample, so that the model can better obtain the important representation information about the sample. Another attention module is used to process the features of multiple images to find other images with similar feature information to the current image. STANet achieves better performance by obtaining better feature representation of image samples. In the 5-way 1-shot task setting of the MiniImageNet dataset, conv64 is used as the feature extraction framework, and the classification accuracy is 53.11%. Li et al. [69] proposed a framework considering that the performance of the model may be affected when the background difference in the image is large. The framework can find important areas in the image for key learning. In addition, metric learning is also used as an auxiliary task to learn more discriminative feature information about the sample.

Later, Gao et al. [70] designed the instance-level and feature-level attention schemes to achieve a more refined and important representation of the samples. With the aid of these two attention mechanisms, the model can obtain more helpful feature information for model classification. The experimental results also show the effectiveness of this method. Oreshkin et al. [71] proposed the TADAM method (task dependent adaptive metric), which has a module responsible for scaling measurement. Specifically, in order to adapt to different tasks, the scale of distance in the metric function is adjusted adaptively. In addition, the method correlates and combines the task conditions with the feature extraction process of image samples through the task conditioning embedding module, so as to better capture and express the important features in the samples. This method significantly improves the performance of the baseline method. Li et al. [72] proposed the CTM (category traversal module) method to improve the performance of FSL. This method integrates the categories of support sets by traversing. Then, when a new learning task is performed, the most relevant data parameters are found to predict. The attention module in this method is used to capture the most relevant feature information of the current task, so as to improve the accuracy of prediction. Yang et al. [73] proposed an FSL method, SEGA (semantic guided attention), using semantic guidance and an attention mechanism. In order to make the model pay attention to some more critical feature information in the image, the method uses an attention mechanism and a semantic guidance strategy to select input features more finely. In this way, important regions in important images have greater weight coefficients, so as to suppress the influence of noise vectors on classification results. Hou et al. [74] proposed the MTL (meta-transfer learning) method by combining the cross attention mechanism with the MAML idea. This method searches the target region in the image through the attention mechanism and regards this region as an important region in the image sample, so that the model can better obtain more discriminative and important features for model learning.

Generally speaking, in the strategy of measurement learning, obtaining stronger semantic feature expression about samples can often make the model have higher performance. In order to make this

statement have a certain support, we select the task settings of 5-way 1-shot and 5-way 5-shot in the MiniImageNet dataset from the "simple feature information acquisition method" and "better feature information acquisition method." The experimental results are given in Table 1. From multiple groups of data in the table, we can find that the performance of the "method based on better feature information acquisition" strategy is better than the "method based on simple feature information acquisition" strategy in most cases. The experimental results also further confirm our statement that obtaining more accurate feature expression about the sample is one of the keys to the success of the metric learning method.

Table 1. Comparison of two subcategory methods in the MiniImageNet dataset.(%)

Method	MiniImageNet	
	5-way 1-shot	5-way 5-shot
Matching Network [59]	43.56 ± 0.84	55.31 ± 0.73
ProtoNet [60]	49.42 ± 0.78	68.20 ± 0.66
Relation Network [61]	50.44 ± 0.82	65.32 ± 0.70
DN4 [62]	51.24 ± 0.74	71.02 ± 0.64
IMP [65]	49.60 ± 0.80	68.10 ± 0.80
K-tuplet Network [67]	58.30 ± 0.84	72.37 ± 0.63
STANet [68]	53.11 ± 0.60	67.16 ± 0.66

5. FSL methods based on meta-learning

The method based on meta-learning strategy is a method of learning. The meta-learning strategy first learns the initial meta knowledge from a large number of prior tasks. Then, by transferring knowledge quickly, guide the new task to quickly learn how to learn and quickly learn appropriate learning parameters about the current task.

In 2016, Santoro et al. [75] proposed a meta-learning method called MANN based on memory-augmented neural networks to solve the one-shot problem in FSL tasks. Finn et al. [76] proposed the MAML method, which is a meta-learning method independent of specific models. The core goal of this method is to achieve good performance in new tasks by only slightly adjusting the model. Specifically, this method aims to make the model get good performance in different tasks, so it does not focus on specific task-specific training but focuses on how to make the model quickly adapt to the previous task and learn quickly when facing different types of tasks. In general, compared with some traditional learning strategies, this method has stronger generalization performance and robustness. Then, based on the MAML algorithm, the Reptile method proposed by Nichol et al. [77] only considers updating the meta-learning parameters using the first-order derivation algorithm. Better initialization parameters can be obtained by updating the first derivative parameters, which can be used in the learning of the meta-learning model to obtain better convergence results. Also based on the MAML method, Antoniou et al. [78] proposed the MAML++ method to make up for some deficiencies in MAML. In short, MAML++ is optimized to improve the lack of stability of MAML and the lack of generalization of network architecture, making the method more generalizable and stable and also improving the convergence speed. It is an excellent algorithm.

Ravi et al. [79] proposed a relatively new method for optimizing the FSL problem. The core idea

of this method is optimization. In short, in order to quickly adapt to the new task, the method first pre-trains the new task and optimizes the parameters by the difference between the prediction results and the pre-training tags. In order to better capture the more detailed features between different types of samples, LSTM is added. Gidaris et al. [80] improved the performance of the model through an attention mechanism. The basic framework of this method is shown in Figure 12. Specifically, consider that the traditional CNN needs to allocate independent weight coefficient vectors to each category in the data set when learning and training images, and it also needs to recalculate when facing new tasks, which is cumbersome and increases the training burden. In this paper, the attention mechanism is added to the classification weight of the basic categories. As a result, even in the new task, the method can also use the experience and knowledge learned in the past, so as to strengthen the rapid learning ability of the model in the new task.

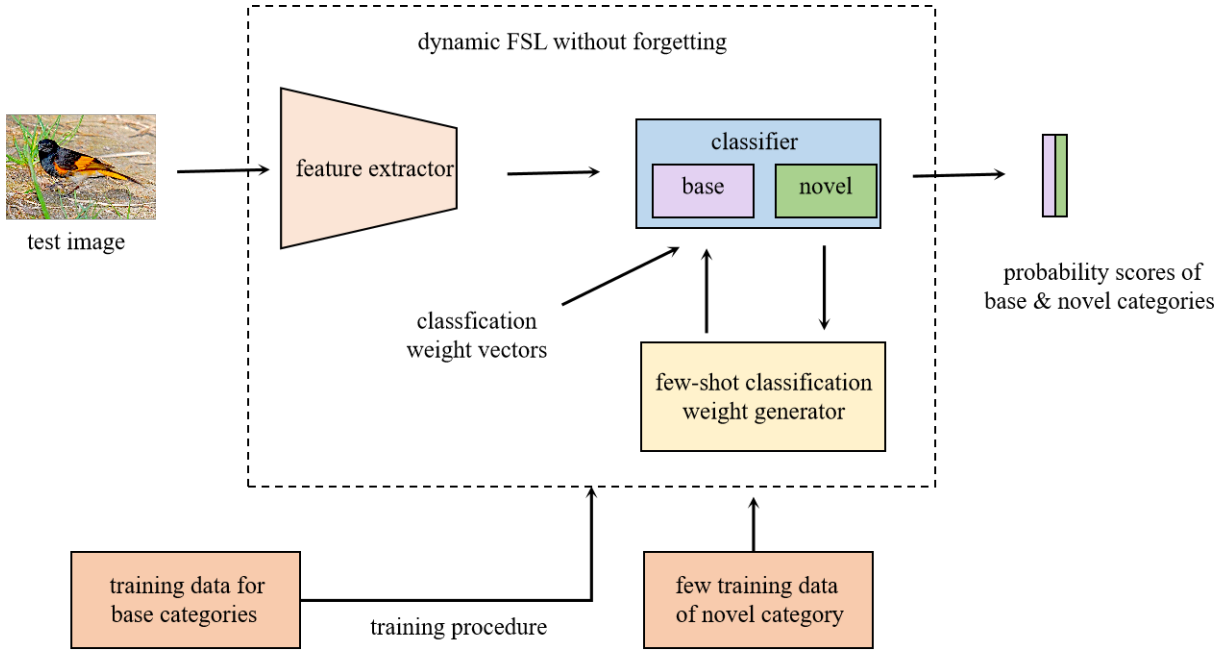


Figure 12. The architecture of Ref [80].

Sun et al. [81] proposed meta-transfer learning. The core idea of this method is to use the similarity between different tasks to transfer meta knowledge. Specifically, the method first uses a prior model to learn the similarity between different tasks. Then, the meta-learning loss function is used to predict and measure the performance of the model between different tasks. By minimizing this loss, the model can be quickly migrated to other new tasks for learning. Ye et al. [82] considered the irrelevance of aggregation order and added the set to set function to the FSL task. Specifically, embedding this function can make the model adapt to the feature mapping between tasks and improve its performance in the target task. Its core idea is to adaptively adjust the feature mapping of the model itself on the new feature space according to the different tasks given, so as to achieve the goal that the closer the feature points are among the same kind of image samples, the farther the distance between different kinds of samples. In order to make the model adapt to the new model more quickly, Lee et al. [83] combined

differentiable convex optimization with meta-learning. This method first takes the parameters in the model and the task itself as a variable and then uses the differentiable convex model to describe the relationship between the model parameters and the model itself, so that the model can release its best performance in the new task. In order to avoid the shortage that it takes more time to retrain the new model, the computational efficiency is optimized. Zhang et al. [84] proposed the meta-navigator method, the core idea of which is to search for adaptation strategies. In this strategy, a random search method is used. First, a group of parameters obtained by random sampling are trained. Then, it is optimized in the update iteration. By searching for a group of excellent or better parameters, the model can quickly adapt to new tasks and has better performance. Aimen et al. [85] proposed a new learning strategy based on meta-learning. The authors first introduced a batch episodic training task to improve the optimization of learning parameters in the model. In addition, they also made the assumption that in the same training batch task, different subtasks should have different learning weights due to the different difficulties of the tasks. In other words, the task attention module included in this method is designed to evaluate the importances of different subtasks, and the more important tasks will give them higher Weighting coefficients and vice versa. Based on the different weights of different tasks, the model can understand which subtask is more important, so that the model can better update the overall learning parameters. A large number of experimental results in multiple datasets show the effectiveness of the learning strategy.

6. FSL methods for adding other auxiliary tasks

In addition to the main strategies mentioned above, some researchers have adopted other strategies to improve the performance of small sample methods in recent years. Self supervised learning (SSL) [86], a popular learning method in recent years, is a new type of learning method. It is widely sought after because of its unique learning method, that is, the strategy of learning without labels and without labeling data, which can undoubtedly save a lot of human and financial resources. Specifically, SSL can be used as an auxiliary task to mine its own supervisory information from large-scale unsupervised information. This kind of training of the network by constructing the monitoring information task can make the model learn a lot of additional information useful to the downstream task and transfer the learned information to the downstream task. In the SSL task, we do not need to annotate the data, because these annotations are obtained by the model from the incoming source data. At present, self supervision has been added to a variety of downlink tasks and has achieved certain results. The main difference of existing SSL tasks lies in the design of auxiliary tasks. Common auxiliary tasks include 2D rotation of predicted images [87], image inpainting [88] and comparative learning tasks [89]. Therefore, in recent years, many researchers have added SSL tasks to FSL and achieved good results.

Zhang et al. [90] proposed the IEPT (instance-level and episode-level pretext task) method to improve the classification performance of FSL. IEPT first amplified the input samples, and amplified a single image into four image samples including the original image, through 90 degrees, 180 degrees, 270 degrees rotations respectively. After that, the authors added a self-supervised auxiliary task about predicting the rotation angle of the image. If the model can accurately predict the image amplification mode, it shows that the model can learn the subtle changes in the image. After that, the learned feature information is transferred to the FSL task, which will help the model learn more accurate detailed features about the sample, so as to improve the performance of the model.

In addition to the above methods, there are other learning tasks that can be added to the FSL task as auxiliary tasks. Contrastive learning (CL) can learn by comparing the differences between positive samples and negative samples. In this process, the model will receive two kinds of data input. The positive samples will be closer, and the negative samples will be farther away. In this way, the model can better learn some relevant detailed characteristic information between the data. Luo et al. [91] obtained better performance in FSL tasks and added CL tasks to FSL tasks. Specifically, in order to enable the model to better capture the subtle differences between different types of image samples, different perspectives of the same image sample are taken as positive samples, and different perspectives between different images are taken as negative samples. Then, by minimizing the loss between positive samples and negative samples. The experimental results show that it is beneficial to distinguish these subtle changes, which can effectively improve the classification performance of the FSL model. Lee and Yoo [92] combined CL to enhance the feature extractor in the model. Specifically, the authors use supervised comparative learning in the pre-training phase to enhance the generalization performance of the model. After the pre-training, a meta-learning loss function is used to optimize the prediction probability between different categories, so that the model can better sort out the characteristics between different categories. Yang et al. [93] also added CL to the FSL task. During the training, the parameters were optimized by minimizing the contrast loss of the classifier and the original loss function. In addition, a contrast loss function is proposed to minimize the distance between similar samples and maximize the distance between different samples. In addition, Lu et al. [94] proposed to use an unsupervised strategy instead of supervised strategy in the pre-training phase of FSL tasks. Using an unsupervised pre-training model can make the model obtain more information about the representation of samples. Experiments show that the performance of the model can be improved by adding unsupervised tasks in the pre-training phase.

Through the above introduction, we can find that adding some useful auxiliary tasks to FSL can improve the performance of the FSL model.

7. Summary of FSL method at present stage

In this review, the existing FSL methods are mainly divided into four categories: data enhancement based FSL methods, metric learning based FSL methods, meta-learning based FSL methods and other auxiliary tasks. Among them, according to the different main strategies, we divide the FSL methods based on data enhancement into the methods based on input image enhancement and the methods based on feature enhancement. Then, considering the complexity of image synthesis, the input image enhancement method is subdivided into simple image amplification and complex image synthesis methods. In the FSL method based on metric learning, we divide it into the method based on simple feature acquisition and the method based on better feature extraction according to whether the feature is extracted more carefully. These methods have made great contributions to the progress and development of FSL, but they also have their advantages and disadvantages. In Table 2, we give a general analysis of the advantages and disadvantages of these strategies.

Table 2. Advantages and disadvantages between different methods.

category			advantage	disadvantage
Methods based on data enhancement	Enhancement based on input image	Simple image amplification	Additional feature information can be obtained by amplification of sample data.	However, it may also introduce noise and have a negative impact.
		Complex image generation		
	Feature enhancement			
Based on metric learning	Simple feature information acquisition		The complex problem can be transformed into a simple feature measurement problem, and the performance of the model can be improved by optimizing its feature representation.	When the image sample is too small (for example, only one sample), it may cause measurement deviation.
	Better feature information acquisition			
Based on meta-learning			With just fine tuning, the model can quickly adapt to new tasks.	The model has high complexity and is relatively complex in actual training.
Add other auxiliary tasks			The performance of the model can be improved by adding additional auxiliary tasks to the FSL.	Relying more on the design of auxiliary tasks also increases computational consumption.

From Table 2, we can see that although FSL has achieved good performance, it also has some shortcomings. In general, we can roughly divide FSL methods into four categories, each of which has its advantages and disadvantages: Although the method based on data enhancement can achieve better performance by expanding the original samples, it also introduces noise, which may have a negative impact on the model. The method based on metric learning can change the complex problem into a simpler way of distance measurement and can also improve the performance of the model by optimizing its feature representation. However, when the image sample information is insufficient, it may cause measurement deviation (for example, in the case of 5-way 1-shot), that is, individual samples cannot represent the actual distribution of most samples. FSL based on meta-learning can quickly adapt the model to new tasks through better learning. However, in some cases, this method is difficult to adjust parameters, and needs more data and computational power. The method based on adding other tasks can obtain better performance by transferring the learning knowledge from other methods. However, this method is more dependent on the design of auxiliary tasks, and how to design auxiliary tasks is also a difficult problem. Moreover, the extended method will also increase the performance requirements of the model and bring computational burden.

8. FSL classification dataset and performance demonstration of some methods

8.1. Introduction to some datasets and their indicators

In this section, we will introduce in detail the datasets often used in the FSL classification task, including Omniglot [95], CIFAR-FS [96], CUB [97], Stanford Dogs [98], MiniImageNet [59], Tiered-ImageNet [99] and Stanford Cars. We will introduce them in detail below:

(1) Omniglot. The dataset is mainly composed of handwritten characters in 50 different languages. Specifically, there are 1623 types of handwritten characters in the dataset, and there are 20 characters in each category. It should be noted that each character in each category is handwritten by 20 different people, so as to ensure some differences between the same characters. When using this data set for experiments, 1200 characters are usually used as the training set of the experiment, and 423 classes

are taken as the verification set. If the experimenter thinks that the amount of data is not enough, he can also enhance and expand the data set by rotating the image 90 degrees, 180 degrees, 270 degrees respectively.

(2) CIFAR-FS. The images in this dataset are the same as those in the CIFAR-100 dataset, but the classification of images in this dataset is different. Specifically, the dataset contains 60000 images with a pixel size of 32×32 and 100 categories. Among them, 64 types of images are used for the training set, 16 types are used for the verification set, and 20 types are used for the test set.

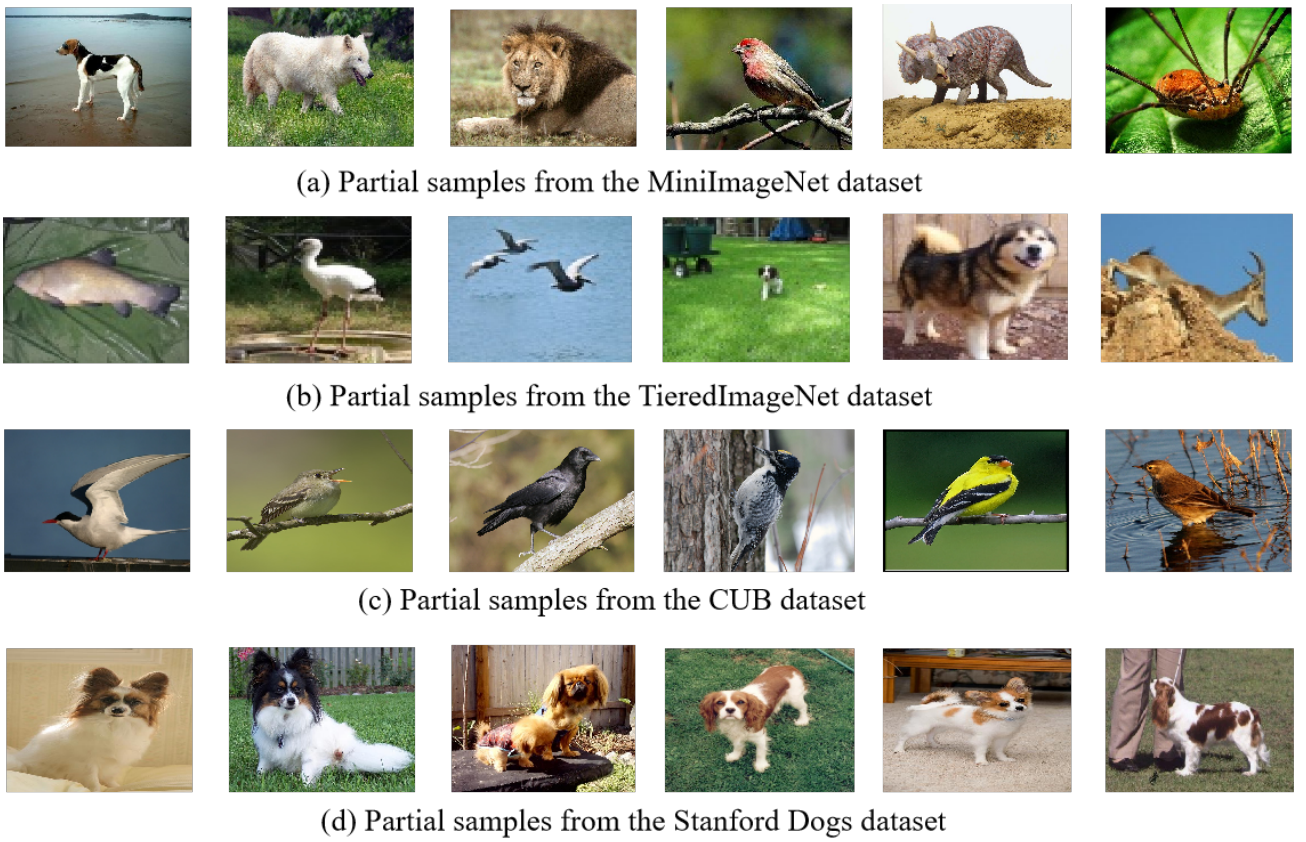
(3) CUB. CUB is also known as CUB-200-2011. The dataset contains images of 200 species of birds. The number of images in the dataset is 11788, and the average number of images of each species of birds is about 59. In this dataset, 130 categories are used for the training set, 20 categories are used for the verification set, and 50 categories are used for the test set.

(4) Stanford Dogs. This dataset is a subclass of the large dataset ImageNet. The dataset contains 120 types of dog images, with an average of 171 images in each category and a total of 20580 images in the dataset. In experiments, this data set is usually divided into 70 types for application model training, 20 types for the validation set and 30 types for the test set.

(5) MiniImageNet. Similar to the general division style in CIFAR-FS, the data set has 60000 images, 100 categories and 600 images in each category, including 64 categories in the training set, 16 categories in the validation set and 20 categories in the test set. The difference is that the pixel dimensions of the images in this dataset are much higher than those.

(6) TieredImageNet. TieredImageNet has 34 major categories and 608 minor categories. In normal FSL tasks, 608 categories are generally used, and each category has an average of 1282 images. The data set is divided as follows: The training set has 351 classes, the validation set has 97 classes, and the test set has 160 classes.

(7) Stanford Cars. There are 16185 images in this dataset, and the total categories are 196. During training, the datasets are usually divided as follows: 130 categories in the training set, 17 categories in the verification set and 49 categories in the test set.



(a) Partial samples from the MiniImageNet dataset

(b) Partial samples from the TieredImageNet dataset

(c) Partial samples from the CUB dataset

(d) Partial samples from the Stanford Dogs dataset

Figure 13. Examples of images from some datasets.

In order to have a fuller understanding of these datasets, we give a partial image display of some datasets in Figure 13. Moreover, based on the above introduction, we can find that these datasets are rich in types and can test and evaluate the performances of the models proposed by researchers to a certain extent. However, the data set used for FSL model testing still has some shortcomings. First, these datasets do not cover enough scenarios. In the actual application scenario, some situations we encounter are more complex. Second, the image quality in the existing data sets is generally high and clear. However, in practice, due to the uneven sampling equipment, the image we collect may have low photographic quality (blurred image and noise interference). In the case of lack of data, we may be reluctant to discard it. The existing data sets may not fully consider the occurrence of these scenarios. Finally, most of the existing data sets have relatively balanced data distribution. However, in practice, we may face these situations when collecting image samples. In order to enrich the richness of data in the dataset and better test the performance of various methods in different scenarios (including accuracy and generalization ability), and achieve more comprehensive performance testing of the proposed methods. The following work on creating new datasets may be improved from the following points: (1) Increase the distribution background of the image and collect the images of the target object in different backgrounds as much as possible, so as to increase the diversity of images in each category. (2) In order to solve the problem of uneven image quality in the actual scene, a certain proportion of noisy images are added to each category of the data set, so as to better detect the anti-interference ability

of the model. (3) In view of the uneven distribution of samples among different categories that may occur in the actual scene, the distribution ratio of samples among different categories can be artificially changed to better test the generalization ability of the trained samples.

8.2. Performance demonstration of some FSL methods

In this section, to demonstrate the performances of some classic and advanced methods, we present their performances with the MiniImageNet and TieredImageNet datasets, commonly used to test the performance of FSL models in Table 3. In the FSL task settings in the table, we have selected the commonly used 5-way 1-shot and 5-way 5-shot. In addition, the performance evaluation indicator is the Top-1 average accuracy commonly used in FSL classification. Through the information in the table, we can find that the more support set samples in each category, the stronger the performance of the final model. In the same method, the accuracy of 5-way 5-shot is always better than that of 5-way 1-shot.

Table 3. Performance demonstration of multiple methods in MiniImageNet and TieredImageNet datasets. (%)

Method	Backbone	MiniImageNet		TieredImageNet	
		5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Matching Network [59] (NIPS' 2019)	Conv4	43.56 ± 0.84	55.31 ± 0.73	-	-
MAML [76] (ICML' 2017)	Conv4	48.70 ± 1.84	63.10 ± 0.92	51.64 ± 1.81	70.30 ± 1.75
MAML++ [78] (arXiv)	Conv4	52.15 ± 0.26	68.32 ± 0.44	-	-
ProtoNet [60] (NIPS' 2017)	Conv4	49.42 ± 0.78	68.20 ± 0.66	53.31 ± 0.89	72.69 ± 0.74
Reptile [77] (arXiv)	Conv4	47.07 ± 0.26	62.74 ± 0.37	-	-
Relation Network [61] (CVPR' 2018)	Conv4	50.44 ± 0.82	65.32 ± 0.70	54.48 ± 0.93	71.32 ± 0.78
DSN [64] (CVPR' 2020)	Conv4	51.78 ± 0.96	68.99 ± 0.69	-	-
DN4 [62] (CVPR' 2019)	Conv4	51.24 ± 0.74	71.02 ± 0.64	-	-
IMP [65] (ICML' 2019)	Conv4	49.60 ± 0.80	68.10 ± 0.80	53.63 ± 0.51	71.89 ± 0.44
K-tuplet Network [67] (Neurocomputing' 2019)	Conv4	58.30 ± 0.84	72.37 ± 0.63	-	-
STANet [68] (AAAI' 2019)	Conv4	53.11 ± 0.60	67.16 ± 0.66	-	-
MetaOptNet [83] (CVPR' 2019)	ResNet-12	62.64 ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.53
Meta Navigator [84] (ICCV' 2021)	ResNet-12	65.91 ± 0.83	82.66 ± 0.55	73.52 ± 0.88	85.34 ± 0.62
DSN [64] (CVPR' 2020)	ResNet-12	62.64 ± 0.66	78.83 ± 0.45	66.22 ± 0.75	82.79 ± 0.48
CMS [100] (CVPR' 2021)	ResNet-12	66.64 ± 0.28	83.63 ± 0.18	73.48 ± 0.31	87.66 ± 0.20

"-" indicates that there is no such data.

9. Achievements, challenges and prospects for the future

At present, the FSL method has made considerable achievements. However, there are still some challenges/difficulties to overcome. In this section, we will introduce the achievements, challenges and future prospects of the FSL method in detail.

9.1. Achievements of FSL methods

In recent years, the FSL method has made good achievements, which can be summarized as follows:

(1) Transfer learning: Transfer the prior knowledge learned in the big data set to other new FSL tasks and quickly adapt the model to new tasks through existing knowledge.

(2) Meta-learning: By making the model quickly learn how to adapt to new tasks, the model can achieve higher classification performance with only a small amount of iterative training.

(3) Extended learning: In the process of implementing FSL tasks, the performance of FSL can be improved by integrating other tasks (such as adding SSL as an auxiliary task).

(4) Data enhancement: Due to the lack of training data for the FSL task, the input samples are enhanced to improve the detection performance of the model by expanding the amount of input data.

(5) Adding attention mechanism: Most methods improve the performance of the FSL model by adding an attention mechanism and then improve the acquisition of feature information about important target objects in limited images.

Through the benefits of the above methods, the FSL method has achieved good results in most public data sets and has been applied well in some actual application scenarios.

9.2. Challenges faced by FSL methods

Although the FSL method has made good achievements and made some progress in some areas, there are still some shortcomings, which can be summarized as follows:

(1) Lack of explicability: At present, it is difficult to explain how the DL based FSL method selects the important parameters considered by the model itself, the parameters are key quantities that play an important role in the model. If the interpretability of the FSL method can be improved, it will help researchers better study it.

(2) The perplexity of pre-training and actual tasks: The weight/model of pre-training can speed up the training of the model, and the new model only needs a few iterations to obtain a model with high performance. However, in practice, large-scale pre-training data sets that match the actual tasks are often difficult to obtain (because we assume there is a small amount of data to learn from). So, the possible lack of suitable large-scale data sets for pre-training is a problem to be solved.

(3) Noise in transfer learning: Transfer learning is mostly completed through the transfer of existing knowledge. There will be a problem when transferring prior knowledge to a new task, that is, the transferred knowledge may not be positive for the existing task, but may also have some negative parts. Learning these school level contents in the new model may reduce the performance of the model.

(4) Dependence on large data sets: Through partial transfer learning, the dependence on large data sets is relatively serious. In the pre-training stage, the more categories of samples there are, the better the quality and the more beneficial to the subsequent model learning. However, this also deepens the degree of dependence on large data sets. If such a large data set cannot be found in the actual situation, the performance of the model may be affected. This may also be one of the factors that transfer learning strategies should consider.

(5) The gap between prior knowledge and the current task: In the strategy of relying on transfer learning, the matching degree of prior knowledge and the current task can affect the performance of the final training model. If the prior knowledge is quite different from it, it may also make the performance of the final training model fail to achieve the expected effect.

9.3. Prospects for future work

Although FSL has made good achievements at this stage, there are still some difficulties and challenges. For future work, we have the following outlook:

(1) Dig deep into the information of the data itself: The reason why humans can learn about the target object through a small number of samples is essentially to learn more accurate details about the target object. The same is true of FSL. In the follow-up work, we hope to introduce an attention mechanism that can find more important positions in a small number of samples and give these eigenvectors larger weight coefficients.

(2) Propose better test data sets: As expected in section 8.1 of this review, we hope to introduce better data sets to better test the classification performance, generalization performance and robustness of the proposed FSL method.

(3) Propose a better model architecture: Although the current methods have done a better job in the classification performance of FSL, the generalization performance and robustness of most FSL methods may still be insufficient to meet the needs of most actual situations. In the face of the demand for model generalization performance in some practical tasks in the future, we hope to put forward a better model architecture.

(4) Combining more excellent learning strategies: Through the content of section 6 of this review, we can find that in the FSL task, the model can obtain better performance by combining learning mechanisms in other fields (such as SSL and comparative learning tasks in section 6). It can be seen that the performance of FSL can become better by combining it with excellent learning strategies. In future research, we hope to combine some strategies in other fields (such as reinforcement learning) with FSL to make it have stronger performance.

(5) Enhance the generalization performance of the model: In many cases, it is inevitable that we have such doubts. The performance of FSL methods that perform well in many public data sets may not be satisfactory in some other actual data sets, affecting the actual deployment of some FSL methods. In the future research, we hope to propose more generalization methods to adapt to more situations as much as possible.

(6) Combining multimodal data: In future research, FSL can combine some multimodal ideas, such as text, voice and other information. By promoting the combination of FSL and other fields, FSL thought can solve problems in more fields.

10. Conclusions

The success of the traditional CNN model is inseparable from the support of a large number of data. However, in order to train models with good performance in the case of insufficient data, researchers proposed the FSL method, which can obtain strong learning ability through a small number of samples. It has to be said that FSL is a challenging task. However, through the efforts of many researchers, many powerful methods have been proposed to solve this problem. In general, this paper divided most FSL methods into four categories, namely, methods based on data enhancement, methods based on metric learning, methods based on meta-learning and methods based on adding other tasks, and introduces some classical and advanced methods in detail. On this basis, we introduced the overall advantages and disadvantages of these four methods. Then, we introduced the datasets often used in FSL classification tasks in detail, introduced their basic parameters and showed the performances of some classical and

advanced methods based on the commonly used MiniImageNet and TieredImageNet datasets, so that readers can understand the performances of these methods. Finally, we introduced the achievements in the field of FSL, the challenges faced at present and the possible development direction in the future.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the Putian Science and Technology Project (2023SZ3001PTXY18).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, T. Ganslandt, Transfer learning for medical image classification: A literature review, *BMC Med. Imaging*, **22** (2022), 69. https://doi.org/10.1186/s12880-022-00793-7
2. Z. X. Zou, K. Y. Chen, Z. W. Shi, Y. H. Guo, J. P. Ye, Object detection in 20 years: A survey, *Proc. IEEE*, **111** (2023), 257–276. https://doi.org/10.1109/JPROC.2023.3238524
3. H. Q. Zhao, W. B. Zhou, D. D. Chen, T. Y. Wei, N. H. Yu, Multi-attentional deepfake detection, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE **8** (2021), 2185–2194. https://doi.org/10.1109/CVPR46437.2021.00222
4. I. Goodfellow, P. A. Jean, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, in *Advances in Neural Information Processing Systems*, **27** (2014), 1–9.
5. B. Pandey, D. K. Pandey, B. P. Mishra, W. Rhmann, A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions, *J. King Saud Univ. Comput. Inf. Sci.*, **34** (2022), 5083–5099. https://doi.org/10.1016/j.jksuci.2021.01.007
6. P. Li, X. H. Xu, Recurrent compressed convolutional networks for short video event detection, in *IEEE Access*, **8** (2020), 114162–114171. https://doi.org/10.1109/ACCESS.2020.3003939
7. P. Li, Q. H. Ye, L. M. Zhang, L.Yuan, X. H. Xu, L. Shao, Exploring global diverse attention via pairwise temporal relation for video summarization, *Pattern Recogn.*, **111** (2021), 107677. https://doi.org/10.1016/j.patcog.2020.107677
8. P. Li, P. Zhang, T. Wang, H. X. Xiao, Time–frequency recurrent transformer with diversity constraint for dense video captioning, *Inform. Process. Manag.*, **60** (2023), 103204. https://doi.org/10.1016/j.ipm.2022.103204

9. P. Li, J. C. Cao, L. Yuan, Q. H. Ye, X. H. Xu, Truncated attention-aware proposal networks with multi-scale dilation for temporal action detection, *Pattern Recogn.*, **142** (2023), 109684. https://doi.org/10.1016/j.patcog.2023.109684
10. P. Li, Y. Zhang a, L. Yuan, H. X. Xiao, B. B. Lin, X. H. Xu, Efficient long-short temporal attention network for unsupervised video object segmentation, *Pattern Recogn.*, **146** (2024), 110078. https://doi.org/10.1016/j.patcog.2023.110078
11. K. Feng, J. C. Ji , Y. C. Zhang, Q. Ni, Z. Liu, M. Beer, Digital twin-driven intelligent assessment of gear surface degradation, *Mechan. Syst. Signal Process.*, **186** (2023), 109896. https://doi.org/10.1016/j.ymssp.2022.109896
12. Y. D. Xu, K. Feng, X. A. Yan, R. Q. Yan, Q. Ni, B. B. Sun, et al., CFCNN: A novel convolutional fusion framework for collaborative fault identification of rotating machinery, *Inform. Fusion*, **95** (2023), 1–16. https://doi.org/10.1016/j.inffus.2023.02.012
13. K. Feng, Y. D. Xu, Y. L. Wang, S. Li, Q. B. Jiang, B. B. Sun, et al., Digital twin enabled domain adversarial graph networks for bearing fault diagnosis, in *IEEE Transactions on Industrial Cyber-Physical Systems*, **1** (2023), 113–122. https://doi.org/10.1109/TICPS.2023.3298879
14. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., ImageNet large scale visual recognition challenge, *Int J Comput Vis*, **115** (2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y
15. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2016), 770–778. https://doi.org/10.1109/CVPR.2016.90
16. A. G. Howard, M. L. Zhu, B. Chen, D. Kalenichenko, W. J. Wang, T. Weyand, et al., MobileNets: Efficient convolutional neural networks for mobile vision applications, preprint, arXiv:1704.04861.
17. X. Y. Zhang, X. Y. Zhou, M. X. Lin, J. Sun, ShuffleNet: An extremely efficient convolutional neural network for mobile devices, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2018), 6848–6856. https://doi.org/10.1109/CVPR.2018.00716
18. G. Huan, Z. Liu, L. V. D. Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2017), 2261–2269. https://doi.org/10.1109/CVPR.2017.243
19. W. H. Yu, M. Luo, P. Zhou, C. Y. Si, Y. C. Zhou, X. C. Wang, et al., MetaFormer is actually what you need for vision, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2022), 10809–10819. https://doi.org/10.1109/CVPR52688.2022.01055
20. Y. P. Chen, X. Y. Dai, D. D. Chen, M. C. Liu, X. Dong, L. Yuan, et al., Mobile-former: Bridging mobilenet and transforme, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2022), 5270–5279. https://doi.org/10.1109/CVPR52688.2022.00520
21. Y. T. Vuong, Q. M. Bui, H. Nguyen, T. Nguyen, V. Tran, X. Phan, et al., SM-BERT-CR: A deep learning approach for case law retrieval with supporting model, *Artif. Intell. Law*, **31** (2023), 601–628. https://doi.org/10.1007/s10506-022-09319-6

22. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, ImageNet: A large-scale hierarchical image database, in *2009 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2009), 248–255. https://doi.org/10.1109/CVPR.2009.5206848

23. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft COCO: Common objects in context, in *2014 European conference computer vision (ECCV)*, (2014), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

24. J. C. Yang, X. L. Guo, Y. Li, F. Marinello, S. Ercisli, Z. Zhang, A survey of few-shot learning in smart agriculture: developments, applications and challenges, *Plant Methods.*, **18** (2022), 28. https://doi.org/10.1186/s13007-022-00866-2

25. J. D. Chen, J. X. Chen, D .F. Zhang, Y. D. Sun, Y. A. Nanehkaran, Using deep transfer learning for image-based plant disease identification, *Comput. Electron. Agri.*, **173** (2020), 105393. https://doi.org/10.1016/j.compag.2020.105393

26. S. Q. Jiang, W. Q. Min, Y. Q. Lyu, L. H. Liu, Few-shot food recognition via multi-view representation learning, *ACM Transact. Multi. Comput. Commun. Appl.*, **16** (2020), 1–20. https://doi.org/10.1145/3391624

27. J. Yang, X. M. Wang, Z. P. Luo, Few-shot remaining useful life prediction based on meta-learning with deep sparse kernel network, *Inform. Sci.*, **653** (2024), 119795. https://doi.org/10.1016/j.ins.2023.119795

28. Y. Q. Wang, Q. M. Yao, J. T. Kwok, L. M. Ni, Generalizing from a few examples: A survey on few-shot learning, *ACM Comput. Surveys*, **53** (2020), 1–34. https://doi.org/10.1145/3386252

29. J. Lu, P. H. Gong, J. P. Ye, C. H. Zhang, Learning from very few samples: A survey, preprint, arXiv:2009.02653.

30. X. X. Li , X. C. Yang, Z. Y. Ma, J. H. Xue, Deep metric learning for few-shot image classification: A Review of recent developments, *Pattern Recogn.*, **138** (2023), 109381. https://doi.org/10.1016/j.patcog.2023.109381

31. A. Dabouei, S. Soleymani, F. Taherkhani, N. M. Nasrabadi, SuperMix: Supervising the mixing data augmentation, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2021), 13789–13798. https://doi.org/10.1109/CVPR46437.2021.01358

32. M. Hong, J. Choi, G. Kim, StyleMix: Separating content and style for enhanced data augmentation, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2021), 14857–14865. https://doi.org/10.1109/CVPR46437.2021.01462

33. N. E. Khalifa, M. Loey, S. Mirjalili, A comprehensive survey of recent trends in deep learning for digital images augmentation, *Artif. Intell. Rev.*, **55** (2022), 2351–2377. https://doi.org/10.1007/s10462-021-10066-4

34. E. D. Ubuk, B. Zoph, D. Mané, V. Vasudevan, Q. V. Le, AutoAugment: learning augmentation strategies from data, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE , (2021), 113–123. https://doi.org/10.1109/CVPR.2019.00020

35. T. DeVries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout, preprint, arXiv:1708.04552.

36. J. Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, (2017), 2242–2251. https://doi.org/10.1109/ICCV.2017.244

37. T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability and variation, preprint, arXiv:1710.10196.

38. Z. T. Chen, Y. W. Fu, Y. X. Wang, L. Ma, W. Liu, M. Hebert, Image deformation meta-networks for one-Shot learning, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE , (2019), 8672–8681. https://doi.org/10.1109/CVPR.2019.00888

39. S. Yun, D. Han, S. Chun, S. J. Oh, S. Chun, J. Choe, Y. Yoo, CutMix: Regularization strategy to train strong classifiers with localizable features, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE , (2019), 6022–6031. https://doi.org/10.1109/ICCV.2019.00612

40. S. Khodadadeh, L. Boloni, M. Shah, Unsupervised meta-learning for few-shot image classification, in *2019 Advances in Neural Information Processing Systems (NIPS)* , (2019).

41. A. Antoniou, A. Storkey, Assume, augment and learn: Unsupervised few-shot meta-learning via random labels and data augmentation, preprint, arXiv:1902.09884.

42. T. X. Qin, W. B. Li, Y. H. Shi, Y. Gao, Diversity helps: Unsupervised few-shot learning via distribution shift-based data augmentation, preprint, arXiv:2004.05805.

43. H. Xu, J. X. Wang, H. Li, D. Q. Ouyang, J. Shao, Unsupervised meta-learning for few-shot learning, *Pattern Recogn.*, **116** (2021), 107951. https://doi.org/10.1016/j.patcog.2021.107951

44. M. Tao, H. Tang, F. Wu, X. Y. Jing, B. K. Bao, C. S. Xu, DF-GAN: A simple and effective baseline for text-to-image synthesis, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2022), 16494–16504. https://doi.org/10.1109/CVPR52688.2022.01602

45. W. T. Liao, K. Hu, M. Y. Yang, B. Rosenhahn, Text to image generation with semantic-spatial aware GAN, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2022), 18166–18175. https://doi.org/10.1109/CVPR52688.2022.01765

46. X. T. Wu, H. B. Zhao, L. L. Zheng, S. H. Ding, X. Li, Adma-GAN: Attribute-driven memory augmented GANs for text-to-image generation, in *Proceedings of the 30th ACM International Conference on Multimedia*, ACM, (2022), 1593–1602. https://doi.org/10.1145/3503161.3547821

47. A. Mehrotra, A. Dukkipati, Generative adversarial residual pairwise networks for one shot learning, preprint, arXiv:1703.08033.

48. Y. X. Wang, R. Girshick, M. Hebert, B. Hariharan, Low-shot learning from imaginary data, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2018), 7278–7286. https://doi.org/10.1109/CVPR.2018.00760

49. R. X. Zhang, T. Che, Z. Ghahramani, Y. Bengio, Y. Q. Song, MetaGAN: An adversarial approach to few-Shot learning, in *2018 Advances in Neural Information Processing Systems (NIPS)*, (2018).

50. E. Schwartz, L. Karlinsky, J. Shtok, S. Harary, M. Marder, A. Kumar, et al., Delta-encoder: an effective sample synthesis method for few-shot object recognition, in *2018 Advances in Neural Information Processing Systems (NIPS)*, (2018).

51. Y. Q. Xian, S. Sharma, B. Schiele, Z. Akata, F-VAEGAN-D2: A Feature generating framework for any-shot learning, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 10267–102765. https://doi.org/10.1109/CVPR.2019.01052
52. K. Li, Y. L. Zhang, K. P. Li, Y. Fu, Adversarial feature hallucination networks for few-shot learning, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2020), 13467–13476. https://doi.org/10.1109/CVPR42600.2020.01348
53. F. Pahde, P. Jähnichen, T. Klein, M. Nabi, Cross-modal hallucination for few-shot fine-grained recognition, preprint, arXiv:1806.05147.
54. M. Dixit, R. Kwitt, M. Niethammer, N. Vasconcelos, AGA: Attribute-guided augmentation, in *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2017), 3328–3336. https://doi.org/10.1109/CVPR.2017.355
55. B. Liu, X. D. Wang, M. Dixit, R. Kwitt, N. Vasconcelos, Feature space transfer for data augmentation, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2018), 9090–9098. https://doi.org/10.1109/CVPR.2018.00947
56. Z. T. Chen, Y. W. Fu, Y. D. Zhang, Y. G. Jiang, X. Y. Xue, L. Sigal, Multi-level semantic feature augmentation in few-shot learning, preprint, arXiv:1804.05298.
57. H. G. Zhang, J. Zhang, P. Koniusz, Few-shot learning via saliency-guided hallucination of samples, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 2765–2774. https://doi.org/10.1109/CVPR.2019.00288
58. G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, in *2015 International Conference on Machine Leaning (ICML)*, (2015).
59. O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching networks for one shot learning, in *2019 Advances in Neural Information Processing Systems (NIPS)*, (2019).
60. J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in *2017 Advances in Neural Information Processing Systems (NIPS)* , (2017).
61. F. Sung, Y. X. Yang, Li, Zhang, T. Xiang, P. H.S. Torr, T. M. Hospedales, Learning to compare: Relation network for few-shot learning, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2018), 1199–1208. https://doi.org/10.1109/CVPR.2018.00131
62. W. B. Li, L. Wang, J. L. Xu, J. Huo, Y. Gao, J. B. Luo, Revisiting local descriptor based image-to-class measure for few-shot learning, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 7253–7260. https://doi.org/10.1109/CVPR.2019.00743
63. Y. B. Liu, J. H. Lee, M. Park, S. Kim, E. Yang, S. J. Hwang, et al., Learning to propagate labels: Transductive propagation network for few-shot learning, preprint, arXiv:1805.10002.
64. C. Simon, P. Koniusz, R. Nock, M. Harandi, Adaptive Subspaces for Few-Shot Learning, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2020), 4135–4144. https://doi.org/10.1109/CVPR42600.2020.00419
65. K. Allen, E. Shelhamer, H. Shin, J. Tenenbaum, Infinite mixture prototypes for few-shot learning, in *2019 International Conference on Machine Leaning (ICML)*, (2019), 232–241.
66. C. Xing, N. Rostamzadeh, B. Oreshkin, P. O. O. Pinheiro, Adaptive cross-modal few-shot learning, in *2019 Advances in Neural Information Processing Systems (NIPS)*, (2019).

67. X. M. Li, L. Q. Yu, C. W. Fu, M. Fang, P.-A. Heng, Revisiting metric learning for few-shot image classification, *Neurocomputing*, **406** (2020), 49–58. https://doi.org/10.1016/j.neucom.2020.04.040

68. S. P. Yan, S. Y. Zhang, X. M. He, A dual attention network with semantic embedding for few-shot learning, in *2019 Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, (2019), 9079–9086. https://doi.org/10.1609/aaai.v33i01.33019079

69. P. Li, G. P. Zhao, X. H. Xu, Coarse-to-fine few-shot classification with deep metric learning, *Inform.n Sci.*, **610** (2022), 592–604. https://doi.org/10.1016/j.ins.2022.08.048

70. T. Y. Gao, X. Han, Z. Y. Liu, M. S. Sun, Hybrid attention-based prototypical networks for noisy few-shot relation classification, in *2019 Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, (2019), 6407–6414. https://doi.org/10.1609/aaai.v33i01.33016407

71. B. Oreshkin, P. R. López, A. Lacoste, Tadam: Task dependent adaptive metric for improved few-shot learning, in *2018 Advances in Neural Information Processing Systems (NIPS)*, (2018)

72. H. Y. Li, D. Eigen, S. Dodge, M. Zeiler, X. G. Wang, Finding task-relevant features for few-shot learning by category traversal, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 1–10. https://doi.org/10.1109/CVPR.2019.00009

73. F. Y. Yang, R. P. Wang, X. L. Chen, SEGA: Semantic guided attention on visual prototype for few-shot learning, in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, (2022), 1586–1596. https://doi.org/10.1109/WACV51458.2022.00165

74. R. B. Hou, H. Chang, B. P. Ma, S. G. Shan, X. L. Chen, Cross attention network for few-shot classification, in *2019 Advances in Neural Information Processing Systems (NIPS)*, (2019).

75. A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, T. Lillicrap, One-shot with memory-augmented neural networks, preprint, arXiv:1605.06065.

76. C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in *2017 International Conference on Machine Leaning (ICML)*, (2017), 1126–1135.

77. A. Nichol, J. Achiam, J. Schulman, On first-order meta-learning algorithms, preprint, arXiv:1803.02999.

78. A. Antoniou, H. Edwards, A. Storkey, How to train your MAML, preprint, arXiv:1810.09502.

79. S. Ravi, H. Larochelle, Optimization as a model for few-shot learning, in *2017 International Conference on Learning Representations (ICLR)*, (2017)

80. S. Gidaris, N. Komodakis, Dynamic few-shot visual learning without forgetting, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2018), 4367–4375. https://doi.org/10.1109/CVPR.2018.00459

81. Q. R. Sun, Y. Y. Liu, T. S. Chua, B. Schiele, Meta-transfer learning for few-shot learning, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 403–412. https://doi.org/10.1109/CVPR.2019.00049

82. H. J. Ye, H. X. Hu, D. C. Zhan, F. Sha, Few-shot learning via embedding adaptation with set-to-set functions, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2020), 8805–8814. https://doi.org/10.1109/CVPR42600.2020.00883

83. K. Lee, S. Maji, A. Ravichandran, S. Soatto, Meta-learning with differentiable convex optimization, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 10649–10657. https://doi.org/10.1109/CVPR.2019.01091

84. C. Zhang, H. H. Ding, G. S. Lin, R. B. Li, C. H. Wang, C. H. Shen, Meta navigator: Search for a Good Adaptation Policy for Few-shot Learning, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2021), 9415–9424. https://doi.org/10.1109/ICCV48922.2021.00930

85. A. Aimen, S. Sidheekh, N. C. Krishnan, Task attended meta-learning for few-shot learning, preprint, arXiv:2106.10642.

86. R. Krishnan, P. Rajpurkar, E. J. Topol, Self-supervised learning in medicine and healthcare, *Nature Biomedical Engineering.*, **6** (2022), 1346–1352. https://doi.org/10.1038/s41551-022-00914-1

87. S. Gidaris, P. Singh, N. Komodakis, Unsupervised representation learning by predicting image rotations, preprint, arXiv:1803.07728.

88. W. X. Wang, J. Li, H. Ji, Self-supervised deep image restoration via adaptive stochastic gradient langevin dynamics, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2022), 1979–1988. https://doi.org/10.1109/CVPR52688.2022.00203

89. H. Q. Wang, X. Guo, Z. H. Deng, Y. Lu, Rethinking minimal sufficient representation in contrastive learning, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2022), 16020-16029. https://doi.org/10.1109/CVPR52688.2022.01557

90. M. L. Zhang, J. H. Zhang, Z. W. Lu, T. Xiang, M. Y. Ding, S. F. Huang, IEPT: Instance-Level and Episode-Level Pretext Tasks for Few-Shot Learning, in *2021 International Conference on Learning Representations (ICLR)*, (2021)

91. X. Luo, Y. X. Chen, L. J. Wen, L. L. Pan, Z. L. Xu, Boosting few-shot classification with view-learnable contrastive learning, in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, (2021), 1–6. https://doi.org/10.1109/ICME51207.2021.9428444

92. T. Lee, S. Yoo, Augmenting few-shot learning with supervised contrastive learning, *IEEE Access.*, **9** (2021), 61466-61474. https://doi.org/10.1109/ACCESS.2021.3074525

93. Z. Y. Yang, J. H. Wang, Y. Y. Zhu, Few-shot classification with contrastive learning, in *2022 European conference computer vision (ECCV)*, (2022), 293–309. https://doi.org/10.1007/978-3-031-20044-1_17

94. Y. N. Lu, L. J. Wen, J. Z. Liu, Self-supervision can be a good few-shot learner, in *2022 European conference computer vision (ECCV)*, (2022), 740–758. https://doi.org/10.1007/978-3-031-19800-7_43

95. S. Fort, Gaussian prototypical networks for few-shot learning on omniglot, preprint, arXiv:1708.02735.

96. L. Bertinetto, J. F. Henriques, P. H.S. Torr, A. Vedaldi, Meta-learning with differentiable closed-form solvers, preprint, arXiv:1805.08136.

97. C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie, The caltech-ucsd birds-200-2011 dataset: Technical report CNS-TR-2011-001, (2011), 1–8.

98. A. Khosla, N. Jayadevaprakash, B. P. Yao, F. F. Li, Novel dataset for fine-grained image categorization: stanford dogs, *CVPR Workshop on Fine-Grained Visual Categorization.*, **2** (2021).

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

100. G. Liu, L. L. Zhao, W. Li, D. S. Guo, X. Z. Fang, Class-wise Metric Scaling for Improved Few-Shot Classification, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2021), 586–595. https://doi.org/10.1109/WACV48630.2021.00063

AIMS Press