Applied Intelligence (2022) 52:15916–15928
https://doi.org/10.1007/s10489-021-03131-6

Sample extraction and expansion method with feature reconstruction and deformation information

Zhengchao Zhang[1] · Hongbin Wang[1] · Nianbin Wang[1]

Accepted: 18 December 2021 / Published online: 19 March 2022

Abstract
Neural networks often need a large number of data to complete effective training. In the low-data regime, networks perform poor in training effect and generalization ability. Recently,most few-shot learning methods based on the data expansion generate data by utilizing generative adversarial idea,or fulfill data augmentation by directly adopting difference information between similar datasets. The data after expansion will still be lack of important data features without considering whether the features of few-shot are complete before expansion. Nor does it examine whether the adoption of difference information is reasonable, which will generate wrong samples. Therefore, this paper puts forward an adversarial data augmentation model based on feature reconstruction and deformation information. Firstly, it proposes sample extraction method based on feature reconstruction, which is used to improve the feature loss of few-shot, and it adopts feature reconstruction to extract typical few-shot for sample expansion. Moreover, it puts forward the sample expansion method based on deformation information, and it adopts deformation information of different clusters under the same class to fulfill the data expansion. The above mentioned methods are applied to the character datasets and some popular few-shot datasets. The typical few-shot after reconstruction and the dataset after expansion have good effects. Furthermore, the experiment results demonstrate the state-of-the-art performance and effectiveness of the proposed methods.

Keywords Few-shot learning · Sample extraction · Data augmentation · Image recognition

1 Introduction

With the development of machine learning, deep learning [1] is applied in all fields, among which great achievements are made in image recognition [2], natural language processing [3, 4], video recognition [5] and other fields. However, in order to get a model with good effects, the training process of deep learning relies on hundreds of labeled data and multiple epochs of iteration training. There are not so many labeled data for training in some fields, such as underwater target recognition, security treatments, medical treatments and so on. Under the condition of less labeled data, if only applying deep learning method and a few labeled data to train the model, the overfitting situation may occur in the training dataset, and the generalization effect of the model isn't good in the testing dataset. Therefore, under the condition of having limited labeled data, how to utilize limited samples to continuously promote the learning ability of networks remains a challenge.

Currently, the best deep learning system still needs hundreds of samples to learn about features in it. In contrast, human beings only need a few supervised information when learning a new concept. For instance, children can summarize the general features of several pictures that are under the same class. Human learning process motivates our interest in few-shot learning [6–10], that is, how to learn a better model from a few labeled samples.

Few-shot learning based on data augmentation [11–15] generates more data through conversion and generation. Conversion usually includes random translation [16], rotating, noise adding and other methods. Generation is usually a method of learning some transitions between similar datasets. These methods essentially expand the number of

✉ Hongbin Wang
wanghongbin@hrbeu.edu.cn

Zhengchao Zhang
zzc13234993193@163.com

Nianbin Wang
wangnianbin@hrbeu.edu.cn

[1] College of Computer Science and Technology, Harbin Engineering University, 150001 Harbin, China

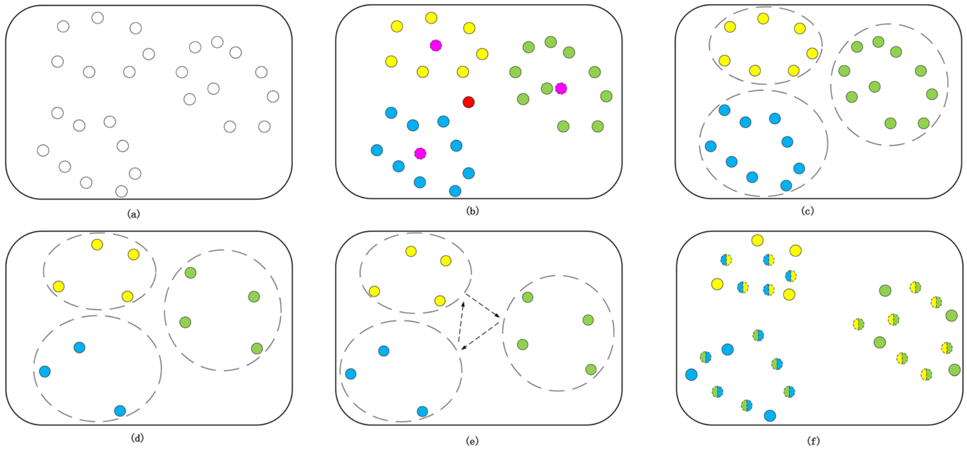


Fig. 1 Schematic diagram of the Sample Extraction Method Based on Feature Reconstruction and the Sample Expansion Method Based on Deformation Information. (a) Homogeneous datasets (b) Datasets after division through homogeneous cluster error and centroid error. Each pink spot represents the centroid of each cluster, and the red spot represents the centroid of this class. (c) Datasets after optimal division. (d) The typical few-shot extracted by utilizing feature reconstruction method. (e) The arrow refers to the deformation information among clusters of few-shot. (f) The new dataset after expansion based on deformation information, and double-colored spots represent the newly generated samples

few-shot, but do not consider data features of original few-shot and how to utilize the learned conversion methods. Therefore, the expanded datasets do not reach our anticipated effect. In order to solve the above mentioned problems, our goal is to expand data based on few-shot that with comprehensive features.

In this paper, we try to solve the problem of a few samples in few-shot learning by means of feature reconstruction and inter-class deformation information. Starting from the deep learning of datasets, we provide a new idea for few-shot learning, and our methods and motivations are shown in Fig. 1. First of all, we analyze the quantification relationship between dataset and few-shot. Then, we extract a typical few-shot from dataset by utilizing the feature reconstruction method of homocluster error and centroid error. Finally, the extracted typical few-shot is expanded into a new dataset through the inter-class deformation information. Experiments show that the proposed extraction method and expansion method are very effective. Our contributions are summarized as follows:

- This paper proposes a sample extraction method based on feature reconstruction, which can extract typical few-shot from dataset from the perspective of data feature. The method takes the centroid of dataset as the standard of extraction metric, which makes the extracted typical few-shot have more comprehensive features and more stable effects. In the meantime, it also provides a typical few-shot with relatively complete features for data expansion.
- In this paper, a sample expansion method based on inter-class deformation information is proposed. It expands the extracted typical few-shot into a new dataset by utilizing the deformation information between data of different clusters under the same class in optimal division. The method extracts and uses the deformation information more reasonably, which makes the generated data more accurate.
- Moreover, we verify the method by applying it to characters datasets and popular few-shot recognition datasets. The features are extracted through convolution network and ResNet network, and then implementing and training through CNNs and MLP. The results show that the extracted typical few-shot and the expanded new dataset are effective. The extraction method and expansion method outperform previous state-of-the-art baselines.

2 Related works

Few-Shot Learning based on Data Augmentation Few-shot learning based on data augmentation refers to expanding the sample quantity of training dataset through scaling, cropping, rotating or generating. Its idea is to expand the prior knowledge by augmenting the minimum available samples and generating more diversified samples, with the goal of making the model better generalised and avoiding overfitting. Hariharan and Girshick et al. [11] proposed low-shot visual recognition by shrinking and hallucinating features. Schwartz et al. [17]put forward a method to classify few-shot by learning the synthesis of new-class samples when only one sample or several samples are available.Antoniou et al. [18] put forward the DAGAN network, and utilized the generative adversarial [19] idea to implement sample expansion at the data level. Alfassy et al. [13] performed intersection, union, set-difference and other operations on data samples, and processed samples with more than one label for few-shot classification settings. However, some data expansion methods directly apply the existing few-shot to sample expansion without considering whether data features involved in few-shot are comprehensive or not, which may cause that the model trained by the expanded datasets can't achieve the anticipated effect.Meanwhile,as for the conversion between samples, some methods directly apply the conversion under a base class to few-shot without considering whether the conversion is reasonable for the new class (Fig. 2).

Few-Shot Learning based on Embedding The idea of few-shot learning based on embedding [20, 21] is to map data into the representation space, and then adopts corresponding distance functions of metric to do the clustering or comparison. Sung et al. [22] put forward the relation network, a flexible, simple and effective network model for few-shot learning. Snell et al. [23] proposed prototypical network, which provides each class a prototype, and the Euclidean distance between samples and prototypes is equal to the distance between samples and its classes. Vinyals et al. [24] put forward matching network, a new-style matching network that introduces attention mechanism. However, some few-shot learning methods based on metric take the centroid of few-shot as the measure standard of that class, but there is a certain error between it and the real centroid, and its stability is not strong.

The relationship between few-shot and database In accordance with several cross validations, the bootstrapping resampling method [25] is adopted to quantify the relationship between few-shot and database. Meanwhile, a hyperparameter n can be obtained, which is the minimum number of samples corresponding to each class. In other words, from the perspective of sample feature, few-shot covers most features of database when the number of few-shot reaches

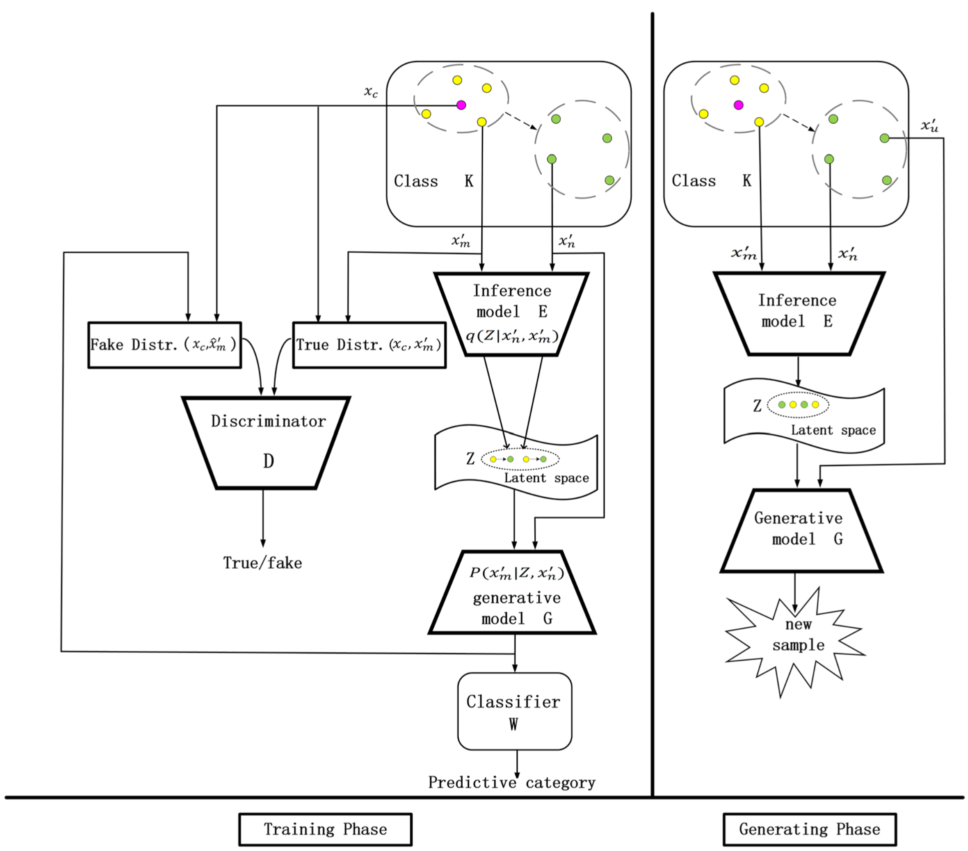


Fig. 2 Adversarial Data Augmentation Model Based on Feature Reconstruction and Deformation Information. During the training phase (left), the inference model encodes the sample pairs (x'_m, x'_n) of different clusters under the same class, and forms deformation information z in the hidden space. Generative model reconstructs the sample $\hat{x}'_m$ through the deformation information z and the sample x'_n. Discriminator is used to discriminate between the true sample pairs (x_c ,x'_m) and the reconstructing sample pairs (x_c,$\hat{x}'_m$). Classifier is used to classify the reconstructing sample $\hat{x}'_m$. During the generating phase (right), the sample pairs (x'_m, x'_n) of different clusters under the same class are still taken as the input of inference model, but the difference is that we take deformation information Z and sample x'_u(x'_u is of homogeneous cluster under the same class with x'_n) as the input of generative model, and finally outputting a new sample for that class

about 30% of the database and the distribution of few-shot is similar to that of database.

When the number of the selected samples in each class reaches a certain proportion, and the distribution of few-shot fits that of dataset , we can use typical few-shot to do few-shot learning. From the perspective of sample feature, the few-shot can't replace dataset when the samples are too limited or the distribution of few-shot is seriously inconsistent with that of dataset. At this time, we should make some adjustments to few-shot from the perspective of feature, which is also the direction and content that we want to study in the next step.

3 Method

In this section, we firstly explain the proposed new definitions and concepts about the extraction and expansion problem of few-shot (section 3.1). Then, in order to solve the problem of sample extraction, we describe the method that extracting typical few-shot from optimal divided clusters of dataset by utilizing feature reconstruction(section 3.2). Finally, in order to expand the extracted typical few-shot, we describe the model structure and its training and generation method. They promote the expansion from typical few-shot to a new dataset by utilizing the inter-class deformation information based on optimal divided clusters of typical few-shot(section 3.3).

3.1 Problem definition

Feature reconstruction It means to extract a few-shot from dataset from the perspective of sample feature, and to reconstruct the features contained in dataset, making the features contained in few-shot can be used to completely describe and summarize the features contained in dataset.

Optimal divided cluster It means the result of dataset's dynamic division when the sum of homogeneous cluster loss and centroid loss is minimal.

Typical few-shot It is extracted from dataset. From the perspective of sample feature, its features can better summarize most features of dataset. Moreover, the extraction process can be called feature reconstruction.

Inter-class deformation information Z=E(x'_m, x'_n)where $E(\cdot)$ denotes the encoder, x'_m and x'_n are the data of different clusters under the same class. Z refers to the required deformation information during the conversion from the sample x'_n to the sample x'_m.

3.2 A sample extraction method based on feature reconstruction

Among the few-shot learning methods based on sample expansion, some methods directly utilize supplied few-shot for data expansion, and these methods have problems from the perspective of data feature. That is because supplied few-shot directly influence the training effect of themselves and expanded dataset. In the previous subsection, the bootstrapping resampling method is adopted to extract different few-shot from the same dataset to train networks, and the different few-shot generate different training effects. In dataset, there must be some data with typical features, and some data with inconspicuous features. Therefore, the supplied few-shot are of great significance, which will influence the learning effect of networks. We can analogize that to human beings' learning process that people usually acquire new knowledge through several typical examples, which can also be used to provide the neural network with effective and typical data. In deep learning, the reason why we can learn a model with a better effect through abundant data is because abundant data can better summarize and describe the features of their classes. For the few-shot, our goal is to summarize the feature of dataset more completely with a small amount of data. Therefore, our motivation is to explore how to extract a typical few-shot from dataset. That is to say, the typical few-shot should cover the features of dataset as many as possible.

In order to solve the above problems, this paper puts forward a sample extraction method based on feature reconstruction(SEFR). SEFR is an extraction method based on metric, and its main idea is to make optimal division of dataset through the method of unsupervised ambiguity clustering, and then to reconstruct the typical few-shot. To be specific, the first step is to compute the central support points C_k of all classes in dataset. Secondly, the sample data under the same class are divided into clusters with dynamic quantity from the perspective of sample feature, and the centroid of each cluster is computed in accordance with the division condition. Then, according to the centroid of each cluster, a new center of that class is computed under the certain division condition, and the sum of homogeneous cluster error and centroid error is computed as the total error of that class under the certain division condition. The condition with minimum error is selected as the optimal division of that class. Finally, the typical few-shot is reconstructed among optimal divided clusters of each class.

3.2.1 The reconstruction process

The average value (centroid) computation of each class The mean vector of all feature vectors of each class in dataset is computed, as the central support point C_k of the class.

$$C_k = \frac{1}{|S_k|} \sum_{(x_i,y_i)\in S_k} f_\varphi(x_i) \tag{1}$$

Where S_k is the sample set of the class k.x_i is the feature vector that belongs to S_k and $f_\varphi(\cdot)$ is an embedding function.

The optimal division of dataset under the same class The sample feature vectors of each class in dataset are clustered into a cluster with dynamic quantity (assuming the number of cluster is m).

At first, computing the new centroid$C'_{k_m_n}$ of each cluster, ($C'_{k\_m\_n}$ means the centroid of the cluster n after dividing data of the class k into m clusters), and describing all samples of that cluster as $X \in S_{k\_m\_n}$($S_{k_m_n}$means the sample set of the cluster n after dividing samples of the class k into m clusters).

Then, under the certain division condition, a new centroid C'_{k_m} of that class is computed through the centroid $C'_{k\_m\_n}$ of each new cluster.

$$C'_{k\_m} = \frac{1}{m} \sum_{n=1}^{m} C'_{k\_m\_n} \tag{2}$$

Where C'_{k_m} is the new central support point of the class k and is formed with the centroid of each cluster after dividing the samples of the class k into m clusters.

Finally, the sum of homogeneous cluster error (L_{hce}) and centroid error (L_{ce}) is regarded as the total error (L_s) of that class under the certain division condition. The condition with minimum error is selected as the optimal division of that class data.

The homogeneous cluster error(L_{hce}) is the sum of distances between all samples of each cluster and the centroid of that cluster.

$$L_{hce} = \sum_{\substack{n=1 \\ X\in S_{k\_m\_n}}}^{m} |X - C'_{k\_m\_n}| \tag{3}$$

Centroid error (L_{ce}) is the distance between the new centroid of that class that constructed by the centroid of each cluster and the centroid of that class in dataset.

$$L_{ce} = C'_{k\_m} - C_k \tag{4}$$

The total error:

$$L_s = L_{hce} + L_{ce} \tag{5}$$

Few-shot reconstruction In accordance with the sample distribution, samples are extracted from each optimal division cluster in a well-distributed and quantitative way, taking as the few-shot for feature reconstruction.

In the optimal division of data under the same class, the sum of homogeneous cluster error and centroid error is regarded as the total error of that class data under that division condition. The goal is that the smaller the homogeneous cluster error is, the denser the homogeneous cluster samples are. And then, the denser the homogeneous cluster data is, the more similar data features are. From the overall point of view, the smaller centroid error indicates the division of each cluster is more reasonable.

Algorithm 1 The SEFR Method

Input: Dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ class center$\{C_k\}_{k=0}^{c-1}$ number of class K SAMPLE(D,k) denotes dividing the dataset D by label k K_Means(S_k,c) means clustering S_k into c categories Select(Q) is uniform extracting from dataset Q
Output: D' denotes typical few-shot

for $k = 0$ to $c-1$ **do**
  $S_k \leftarrow$SAMPLE(D,k)
  $C_k \leftarrow \frac{1}{|S_k|}\sum_{(x_i,y_i)\in S_k} f(x_i)$
end for
for $k = 0$ to K **do**
  for $m = 2$ to 10 **do**
    for $n = 1$ to m **do**
      $Q_{k\_m}, C'_{k\_m\_n} \leftarrow$ K_Means(S_k,m)
      Compute L_{hce} ,L_{ce} ,L_s through C_k ,C'_{k_m} ,$C'_{k\_m\_n}$
    end for
    $\mathbf{C'_{k\_m}} \leftarrow \frac{1}{m}\sum_{n=1}^{m} C'_{k\_m\_n}$
  end for
  Select the optimal partition value m for the kth category
  $D' \leftarrow$Select(Q_{k_m})
end for

We obtain a typical few-shot through the sample extraction method based on feature reconstruction of dataset. From the perspective of sample feature, the features of typical few-shot are more comprehensive. From the perspective of sample distribution, the distribution of typical few-shot is similar to that of dataset. From the perspective of class centroid, the support point computed in typical few-shot is much closed to that computed in dataset, and we try to guarantee that each cluster is more dense in optimal division. Therefore, the typical few-shot can better summarize the feature of dataset.

In this subsection, we make optimal divided cluster for data of each class. The process is also a full preparation for learning deformation information of different clusters under the same class in the next subsection.

3.3 A sample expansion method based on inter-class deformation information

Few-shot learning based on data expansion is to learn the conversion method between data of base class, and then to apply that acquired conversion method to other few-shot that similar to the base class, so as to expand few-shot into dataset. In the task of few-shot recognition, there may be several problems when using the above-mentioned methods to synthesize new data. First, it does not consider whether the conversion method learned on the base class is suitable for few-shot datasets. For instance, we apply the conversion method of the items learned by the base class to the few-shot datasets of animals. This must be unreasonable. Meanwhile, some sample synthesis methods do not consider the similarity of features between the base class and the few-shot datasets. It may generate wrong samples if this conversion method of the base class is imposed on few-shot without considering whether it's suitable for few-shot or not. Wrong samples influence not only the training effect of network, but also the distribution of dataset, making the effect of expanded dataset not ideal. Moreover, in the absence of prior knowledge, it is usually a good method to randomly select few-shot dataset for initial training, but since there are only a small number of samples, we should select as much as possible the typical ones that cover more features. Because different few-shot datasets have different training effects on the network.

With the purpose of solving the problem of few-shot expansion, this paper puts forward a generative and adversarial sample expansion method, also called sample expansion method based on inter-class deformation information. Its main idea is to learn deformation information between data under the same class, and then to apply the deformation information to other samples under same class to generate and expand new data for that class. In order to learn deformation information between data under the same class better, we adopt sample extraction method based on feature reconstruction to provide a typical few-shot with more complete features for data expansion. Each class of typical few-shot is divided into clusters with different features. Specifically, there are two phases involved in sample expansion: the training phase and the generating phase.

Training phase In accordance with the optimal division of each class in the previous subsection, we divide typical few-shot into sample pair (x'_m, x'_n),where x'_m and x'_n are sample data of different clusters under the same class. For instance $x'_m \in Q_{k\_i\_s}$,$x'_n \in Q_{k\_i\_t}$,$s \neq t$, $Q_{k\_i\_s}$ and $Q_{k\_i\_t}$ represent the cluster s and t after optimally dividing data under the class k into i clusters. In this way, the deformation information between the cluster s and cluster t can be learned under the condition of guaranteeing optimal division. The advantage of this way is the acquired deformation information is relevant to that class, and can be applied to that class. The method we propose is to train a network structure that is formed with inference model, generative model, discriminator and classifier.

Inference model learns about the deformation information between x'_m and x'_n, the two samples are of different clusters under the same class. In other words, deformation information is the required additional conversion information during the conversion from $x'_n$to x'_m. Existing in latent space, it is described as $Z = E(x'_m, x'_n)$. The generative model, describing as $\hat{x}'_m = G(Z, x'_n)$, reconstructs $\hat{x}'_m$ by utilizing the deformation information Z in latent space and the input sample x'_n.

$$\boldsymbol{L_{mse}}\left(\hat{\mathbf{x}}'_m, \mathbf{x}'_m\right) = \mathbb{E}\left(\hat{\mathbf{x}}'_m - \mathbf{x}'_m\right)^2 \tag{6}$$

Where $L_{mse}(\hat{x}'_m, x'_m)$ is the mean square error between the newly generated sample $\hat{x}'_m$ and the true sample x'_m.

Discriminator is trained to distinguish between the true sample pair (x_c,x'_m) and the reconstructing sample pair $(x_c ,\hat{x}'_m)$.It can make networks reconstruct the sample more accurately through adversarial training.

$$L_D = -\mathbb{E}\left[\log D\left(\mathbf{x}_c, \mathbf{x}'_m\right)\right] - \mathbb{E}\left[\log\left(\mathbf{1} - \boldsymbol{D}\left(\boldsymbol{G}\left(\boldsymbol{Z}, \mathbf{x}'_{\boldsymbol{n}}\right), \mathbf{x}_{\boldsymbol{c}}\right)\right)\right] \tag{7}$$

Where L_D is adversarial loss.$D(\cdot,\cdot)$ is discriminator and $G(\cdot,\cdot)$ is generator.

Classifier is used to classify the reconstructing sample $\hat{x}'_m$.

$$\boldsymbol{L_{cls}}\left(\boldsymbol{W}, \hat{\mathbf{x}}'_m, \boldsymbol{y}\right) = \mathbb{E}\left[\boldsymbol{l}\left(\boldsymbol{W}\left(\hat{\mathbf{x}}'_m\right), \boldsymbol{y}\right)\right] \tag{8}$$

Total error:

$$L = L_{mse} + L_D + L_{cls} \tag{9}$$

Generating phase We still take the sample pair (x'_m, x'_n) of different clusters under the same class as the input of inference model. Generative model takes the deformation information Z and the sample $x'_u \in Q_{k\_i\_t}$(that of homogeneous cluster under the same class with x'_n) as the input and finally generates a new sample for that class. To generate more new samples for the class by constantly changing the sample pairs (x'_m, x'_n) inputted by inference model (the sample pairs should be guaranteed to be trained in the training process), or changing the sample x'_u inputted by generative model, so as to guarantee that the acquired deformation information is utilized to the greatest extent. During the training, the purpose of adopting adversarial training is to enable networks to generate a new sample based on the existing samples. As long as the new sample looks different from the input samples, it can become a new sample of that class.

We can obtain a new dataset after expanding typical few-shot through inter-class deformation information. From the perspective of sample feature, typical few-shot has relatively complete features, and we use this method to generate sample data that different from typical few-shot, so as to make the training model of expanded dataset more generalized. From the perspective of sample distribution, during the training phase, we utilize adversarial training to discriminate whether it is from true distribution or fake distribution, and try to guarantee that the distribution of generated dataset is closed to that of typical few-shot. From the perspective of deformation information, it is more reasonable to apply the deformation information between different clusters under the same class to the other samples under the same class, and it is more available.

4 Experiment

In this section, we firstly introduce several standard benchmark datasets (MNIST, EMNIST [26],miniImageNet and CIFAR-100) that are applied to our experiment(section 4.1). Then, we conduct three experiments, the quantification experiment to dataset and few-shot(section 4.2), the extraction experiment to typical few-shot(section 4.3), and the expansion experiment to typical few-shot(section 4.4). All the three experiments are used to evaluate SEFI method and SEIDI method.

4.1 Setup

In this subsection, we evaluate methods in some common datasets and popular few-shot datasets. They are the typical few-shot selected by utilizing sample extraction method based on feature reconstruction and the new dataset is generated by utilizing sample expansion method based on inter-class deformation information. In order to prove their effectiveness, they are compared with some state-of-the-art few-shot learning methods. Then, we utilize the visualization method to show the centroid of optimal divided cluster, and the new sample or feature vector synthesized through the SEIDI method.

For the MNIST dataset, there are 10 classes, with 60,000 pieces of data in training dataset, and 10,000 pieces of data in testing dataset. We select the first 2000 pieces of data in training dataset as the dataset, and extract 30% of the dataset (600 pieces of data in total) as few-shot. In the few-shot, there are about 60 samples in each class.

For the balanced dataset in the EMNIST dataset, there are 47 classes (including 10 digit classes, 37 character classes), with 112,800 pieces of data in training dataset, and 18,800 pieces of data in testing dataset. We only select 6 classes (42-47) as the character recognition task, so as to compare with other experiments. Similarly, we select 200 pieces of data from each class (1200 pieces of data in total) as dataset.

For the miniImageNet dataset and the CIFAR-100 dataset, both are the popular datasets in few-shot image recognition. The miniImageNet dataset is 100 categories selected from Imagenet [27], and each category contains 600 samples. In the data preprocessing, the size of all images is adjusted to 84*84 , and the feature vector of the image is extracted with the ResNet network. The CIFAR-100 [28] dataset also has 100 categories and each category contains 600 samples.The size of the image is 32*32. For the above two datasets, in order to compare with some methods, we use two types of sample size: 1shot/5shot-5way.

In the process of experiments, we take the extracted data as few-shot, and take the not extracted data as remaining dataset, also using for testing.

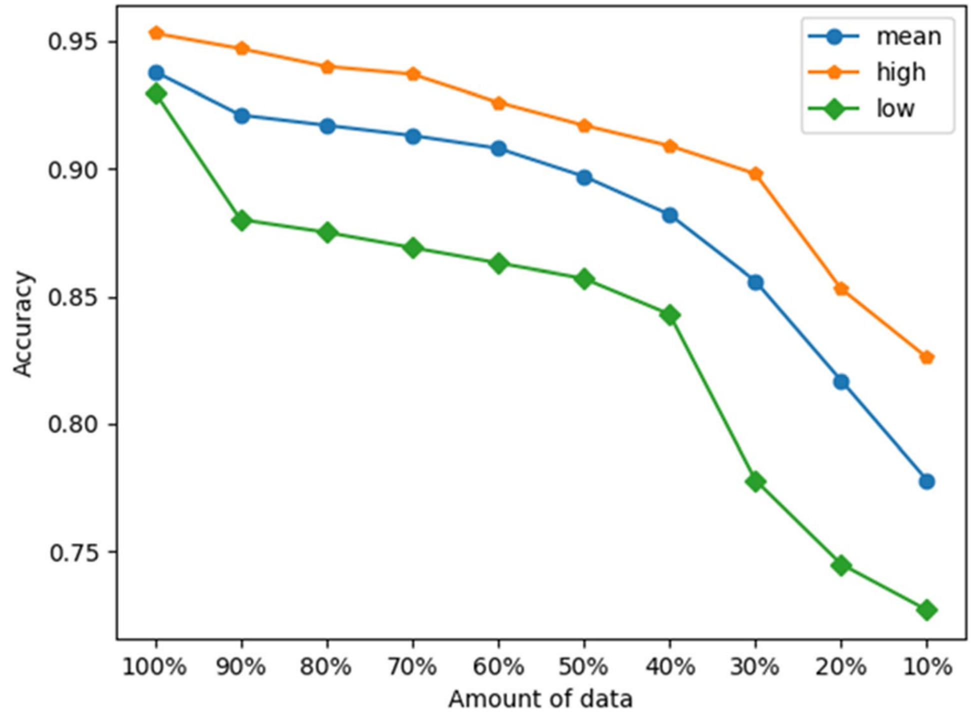
4.2 Quantitative experiment of few-shot and database

First of all, we should make it clear that the number of dataset is on the basis of the number of few-shot. Assuming that there are 60,000 pieces of data in training dataset of MNIST, and even if we extract 30% of that, there will be 18,000 pieces of data. For the few-shot mentioned in this paper, tens of thousands pieces of data are too many, so it can't become our few-shot. In this paper, the number of few-shot should be about 60 pieces, and be no more than 100 pieces.

Secondly, for MNIST dataset, its accuracy rate can reach 98.33% when the model is trained by utilizing 60,000 samples of training dataset, and tested by utilizing 10,000 samples of testing dataset. Its accuracy rate can reach 93.4% when the model is trained by utilizing the first 2000 samples in training dataset, and tested by utilizing 10,000 samples in testing dataset at the same time. In order to explore the quantification relationship between dataset and few-shot, we take the first 2000 pieces of data in training dataset as dataset, and their corresponding model testing accuracy (93.4%) as the benchmark value.

Finally, we use bootstrapping resampling method to extract sample data, from 10% of dataset (200 samples) to 90% of dataset (1800 samples). We use the same CNNs model structure to train the extracted data, and use the same testing dataset to test, the results concluded are shown in the Fig. 3.

Fig. 3 The Figure about Relationship Between Data Number and Model Accuracy



Analysis of Experiment Results (1)From Fig. 3, it's clear that the model accuracy declines with the decrease of data number, and the model accuracy declines more seriously when the extracted samples are less than 30% of dataset. (2)In the dataset, when extracting few-shot with the same number, models trained by different samples have obvious differences during testing, and the differences become more obvious with the decrease of extraction number. Moreover, it also shows that it's significant for us to provide typical few-shot for model. (3)According to the quantification experiment of dataset and few-shot, we confirm taking 30% of dataset as the few-shot (50-60 samples in each class). When sampling through bootstrapping, the effect may be closed to the training effect of dataset when extracting 30% of the dataset. However, if the number reduces again, it's difficult to achieve the training effect of dataset at the optimal condition.

4.3 Few-shot extraction experiment

First of all, we make optimal division to data of each class in dataset, and it increases incrementally after dividing into two clusters during division. When $L_{n+1} > \alpha L_n$(L_n is the total loss when it is divided into n clusters, α refers to hyperparameter, setting $\alpha = 0.95$ during experiments), we make sure the optimal cluster number of that class is n. Then, data of each class are divided after making sure the optimal division number. The division results of MNIST dataset are shown in the following Table 1.

Table 1 The optimal division of datasets

Label	The optimal division number
0	4
1	6
2	3
3	4
4	4
5	4
6	4
7	4
8	3
9	5

Secondly, samples are extracted from optimal division of each class in an well-distributed and quantitative way to construct typical few-shot, and the not extracted data are taken as remaining dataset. Models are trained by utilizing the extracted typical few-shot, and tested by utilizing testing dataset and remaining dataset.

Finally, we compare four methods: dataset, bootstrapping resampling, uniform sampling, and SEFR sampling (tested by testing dataset and its corresponding remaining dataset respectively). 50 experiments are conducted, and the best, average, and worst conditions are summarized and shown in Table 2:

Table 2 The Experiment Results of Digits Datasets

Methods	Maximum	Average	Minimum
Datasets	95.3	93.8	93.0
Bootstrapping	89.8	85.6	77.8
Uniform sampling	90.3	86.5	81.8
SEFR-Test	**91.7**	**89.3**	**86.2**
SEFR-Remaining	92.2	90.1	86.4

Bold entries font is to reflect the best results in this group in the experiment

Analysis of Experiment Results

(1) It's obvious that the effect of SEFR method is better than that of bootstrapping resampling and uniform sampling. Firstly, the average classification accuracy of SEFR method is higher than that of bootstrapping resampling and uniform sampling. Secondly, the fluctuating range of SEFR method is obviously smaller than that of bootstrapping resampling, so its effect is more stable, and the worst condition of SEFR method is obviously better than that of bootstrapping resampling. SEFR method can avoid to extract extreme few-shot, and its worst condition is better than the average level of bootstrapping resampling.

(2) Under the condition of data reduction by 70%, comparing with the average accuracy trained by dataset, the average accuracy of SEFR method declines by 4.5%. We use CNNs to learn sample features, so from the perspective of feature, the few-shot extracted by utilizing SEFR method can cover most features of dataset.

(3) When SEFR extraction method is utilized, and remaining dataset is used to test, the optimal condition of typical few-shot can replace 94.5% of the original dataset.

(4) Among the experiments using SEFR method, there are many experiments with their accuracy higher than 90%, and even closed to the average accuracy of the model that trained by dataset.

4.4 Few-shot expansion experiment

According to the optimal divided cluster of each class in the previous subsection, data under the same class can be divided into several clusters. Taking sample labeled with 5 in MNIST as an example, dividing that into 4 clusters, and visualization is utilized to show the centroid of each cluster in Fig. 4. It can be seen from the figure of visualization

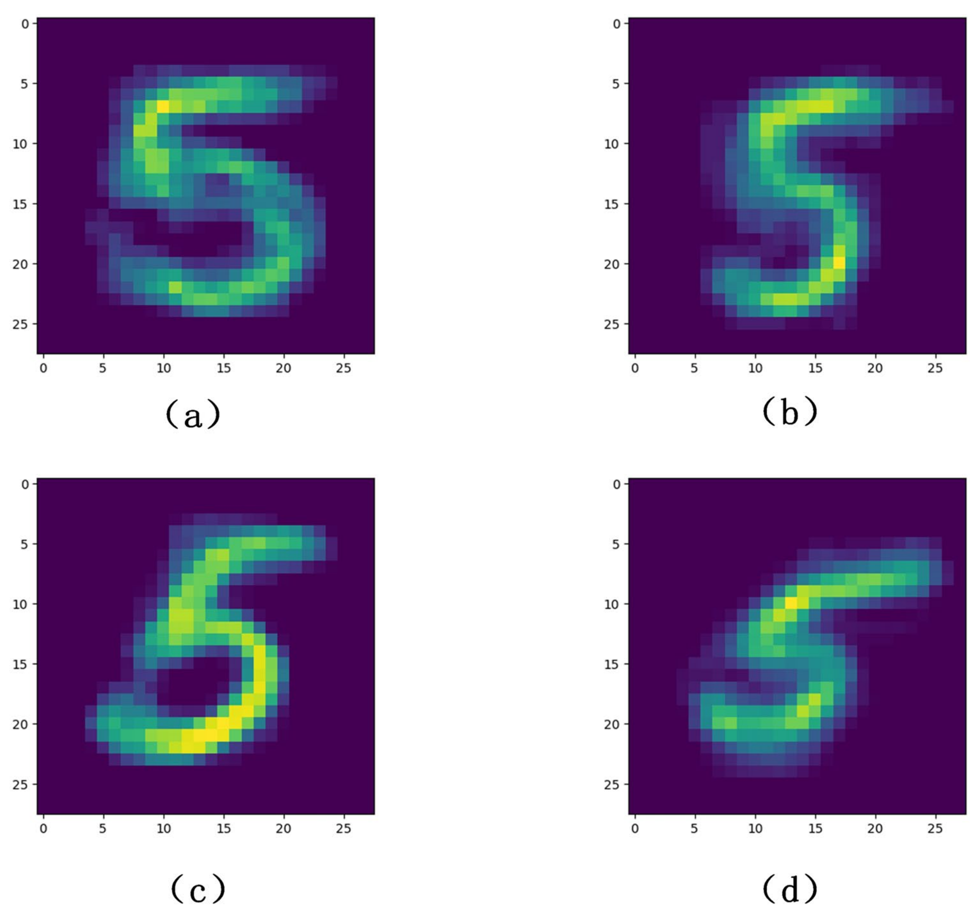


Fig. 4 Taking sample labeled with 5 in MNIST as an example, visualizing the centroid of each cluster among optimal divided clusters

that the centroid shape of each cluster is different, and the deformation information we are looking for is the conversion information between any two clusters (such as Fig. 4(a) and (b)). The difference information is applied to other samples of one cluster (Fig. 4(b)) of these two clusters.

In the previous subsection, the typical few-shot is extracted from MNIST digits datasets and EMNIST characters datasets. Based on the SEIDI method, we conduct several data expansion experiments for the typical few-shot in a way that each time the data volume doubles on the basis of the original data volume. We utilize DAGAN and other expansion methods to compare between the new dataset after expansion and the typical few-shot before expansion, all are shown in Tables 3 and 4.

For the miniImageNet dataset and the CIFAR-100 dataset, the feature vector of the image is extracted by the ResNet network.We use the SEFR method to select the centroid of each category or clusters as few-shot. In this way, a few-shot dataset of 1shot/5shot-5way is constructed. A typical dataset is extracted from remaining samples and formed into sample pairs to learn the deformation information between the same category. Based on the SEIDI method, the original few-shot is expanded. After the expansion, each category contains about 500 feature vectors. Then use the expanded dataset to train the classifier to classify the testing dataset and compare it with some previous state-of-the-art few-shot learning methods, as shown in Table 5.

Analysis of Experiment Results The classification accuracy of SEIDI method on MNIST digits datasets and EMNIST characters datasets is shown in Tables 3 and 4, and the average accuracy of all experiments is reported. Table 3 shows that the average accuracy of typical few-shot in MNIST is increased by more than 2.5% after expansion by utilizing SEIDI method. When there are 100 samples in each class of a typical few-shot, the average accuracy after expansion is closed to that of the original dataset. The experiment results show that model trained by expanded dataset has a better generalization, meaning that new data can be generated by the correct utilization of the deformation information. The newly generated data is different from original data, thus improving the training effect of networks.

Table 3 MNIST SEIDI Classification

Experiment	Samples Per Class	Test Accuracy
MNIST_Strandard	60	0.893
MNIST_SEIDI	60	**0.918**
MNIST_Strandard	100	0.897
MNIST_SEIDI	100	**0.933**
MNIST_Strandard(dataset)	200	0.938

Bold entries font is to reflect the best results in this group in the experiment

Table 4 EMNIST SEIDI Classification

Experiment	Samples Per Class	Test Accuracy
EMNIST_Strandard	50	0.815055
EMNIST_DAGAN_Augmented [18]	50	0.827832
EMNIST_SEIDI	50	**0.845416**
EMNIST_Strandard	100	0.837787
EMNIST_DAGAN_Augmented	100	0.848009
EMNIST_SEIDI	100	**0.865833**

Bold entries font is to reflect the best results in this group in the experiment

Table 4 shows the typical few-shot in EMNIST after expansion by utilizing SEIDI method, which is obviously better than the typical few-shot before expansion by utilizing DAGAN method. Compared with DAGAN method, SEIDI method averagely improves by about 1.8% in two experiments. We make optimal divided clusters in typical few-shot, making networks learn inter-class deformation information more easily and correctly. Moreover, when discriminating the newly generated samples, the centroid of the corresponding cluster is provided, aiming to make the newly generated samples more consistent with the feature and distribution of the cluster.

The classification accuracy of SEIDI on the miniImageNet dataset and the CIFAR-100 dataset is shown in Table 5. In the case of 5shot-5way on the two datasets, the classifier trained with the feature vectors after the expansion of the SEIDI method is better than other comparison methods. An increase of 1.4% on the miniImageNet dataset and an increase of 1.8% on the CIFAR-100 dataset. It shows that the new feature vectors based on the SEIDI method are effective for the same class data. In the process of expansion, as the sample size increases, the

Table 5 miniImageNet and CIFAR100 1shot/5shot 5way SEIDI Classification

Method	miniImageNet	CIFAR100
MatchingNets [24]	46.6/60.0	50.5/60.3
Meta − LearnerLSTM [29]	43.4/60.6	-
MAML [30]	48.7/63.1	49.3/58.3
PrototypicalNetworks [23]	46.1/65.8	48.7/64.7
RelationNet [22]	51.4/67.1	-
TPN [8]	53.8/69.4	-
DEML + Meta−SGD [31]	58.5/71.3	61.6/77.9
TADAM [9]	58.5/76.7	40.1/56.1
DualTriNet [32]	58.1/76.9	**63.4**/78.4
TransductiveFine−tuning [33]	**65.7**/78.4	-
SEIDI	61.7/**79.8**	62.5/**80.2**

Bold entries font is to reflect the best results in this group in the experiment

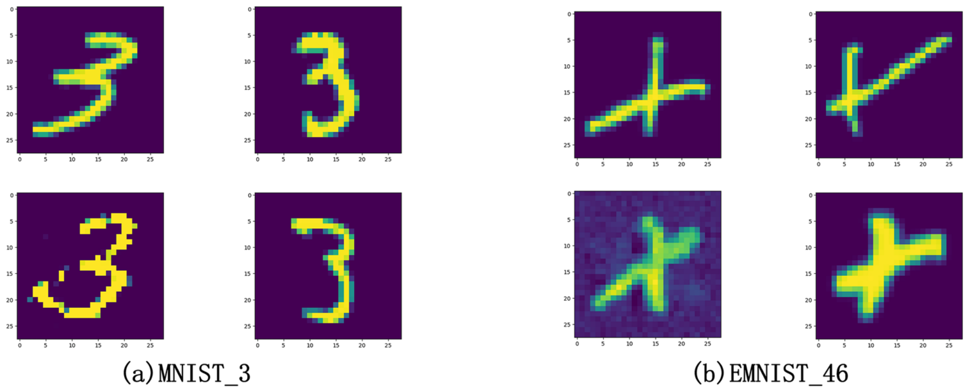


Fig. 5 The newly generated samples in MNIST digits datasets (labeled with 3) and EMNIST characters datasets (labeled with 45) by utilizing network structure of SEIDI method

accuracy of the model is constantly improving. When the sample size reaches about 500 or so, the sample size has little effect on the model.For the case of 1shot-5way, the SEIDI method does not perform well. Each category has only one sample to expand the data, which has certain limitations. It is difficult for one sample to represent the features of the category, and has a certain impact on data expansion.We should consider adding prior knowledge to expand the features of few-shot.

Visualization experiment The bottom left pictures of Fig. 5(a) and (b) are the newly generated samples. Taking Fig. 5 (b) as an example, new samples for that class are generated by utilizing the sample at the bottom right and the deformation information between the upper left sample and the upper right sample. Our goal is to guarantee that the newly generated samples belong to that class, and that the newly generated samples are different from the supplied samples, thus guaranteeing the availability of newly generated samples, and improving the networks generalization of the expanded dataset.

Figure 6 shows the two-dimensional t-SNE visualization results of the feature vectors generated by 5way-5shot in CIFAR-100. The gray points represent the few-shot of 5way-5shot, and the remaining points with different colors represent different categories. It can be seen from the figure that the new feature vectors generated in the same category are near the original few-shot feature vectors. It indicates that the SEIDI method correctly utilizes the inter-class deformation information to generate new data.

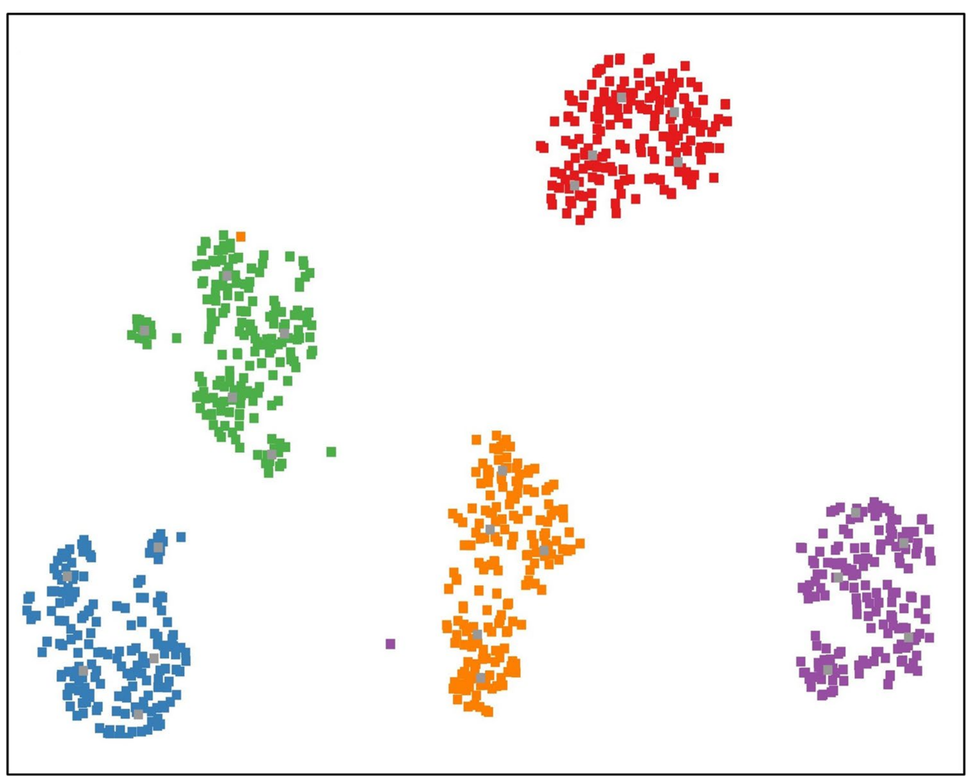


Fig. 6 The t-SNE visualization results of the features in CIFAR100 dataset.It shows generated samples for 5way-5shot. The gray points are the few-shot feature vetors and other points are generated.Different colors represent different categories

Implementation details We implement SEIDI method that bases on keras, and train networks structure by utilizing Adam optimizer. For digits experiments, each class is divided into 25 sample pairs, taking as the input of model. For characters experiments, in accordance with different number of typical few-shot, each class is divided into 25 samples or 40 samples for training. We map samples as a 1568-dimensional vector, the deformation information output by inference model is a 64-dimensional vector, and the size of sample generated by generator is 28*28. Inference model and generative model are constructed by multilayer perceptron (MLP) with a single hidden layer, and each layer is followed with ReLU activation layer. For classifier and discriminator, we utilize CNNs structure to implement them. As for every experiment, we adopt 32 sample pairs in each batch, set learning rate as 10^{-3}, and repeat every experiment for 50 times. Taking the training process of the MNIST dataset as an example, it takes about 20 epochs to reach convergence, and the running time of each epoch is about 30 seconds.

5 Conclusions and future work

Data expansion is a widely used method for performance improvement in the few-shot learning. In this paper, we provide a new idea for few-shot learning, and study new sample extraction method and sample expansion method. Considering two problems in our previous work, this paper puts forward two methods combined with homogeneous cluster loss and centroid loss, the sample extraction method based on feature reconstruction (SEFC) and the sample expansion method based on inter-class deformation information (SEIDI), thus realizing the extraction and expansion of typical few-shot, and trying to improve the two problems. The experiment analysis proves the availability of SEFC and SEIDI, and the visualization experiment can describe our assumptions and results more directly and vividly. We consider to take newly generated samples as the input of reference example, and we consider how to supplement and adjust few-shot when the acquired few-shot lacks features. These will be our future research directions.

Acknowledgements This work is supported by the Basic Research Project (JCKY2019604C004).

References

1. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
2. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
3. Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K (2016) Google's neural machine translation system: Bridging the gap between human and machine translation
4. Wang X, Chen W, Wu J, Wang Y-F, Wang WY (2018) Video captioning via hierarchical reinforcement learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4213–4222
5. Wang X, Chen W, Wu J, Wang YF, Wang WY (2017) Video captioning via hierarchical reinforcement learning. Pattern Recognition
6. Wang Y, Yao Q, Kwok JT, Ni LM (2020) Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys (CSUR) 53(3):1–34
7. Gidaris S, Komodakis N (2018) Dynamic few-shot visual learning without forgetting. In: 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR)
8. Liu Y, Lee J, Park M, Kim S, Yang E, Hwang SJ, Yang Y (2018) Learning to propagate labels: Transductive propagation network for few-shot learning. In: International conference on learning representations, ICLR
9. Oreshkin BN, Lacoste A, Rodriguez P (2018) Tadam: Task dependent adaptive metric for improved few-shot learning. Advances in Neural Information Processing Systems
10. Ramalho T, Garnelo M (2019) Adaptive posterior learning: few-shot learning with a surprise-based memory module. In: International conference on learning representations, ICLR
11. Hariharan B, Girshick R (2017) Low-shot visual recognition by shrinking and hallucinating features. In: Proceedings of the IEEE international conference on computer vision, pp 3018–3027
12. Wang Y-X, Girshick R, Hebert M, Hariharan B (2018) Low-shot learning from imaginary data. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7278–7286
13. Alfassy A, Karlinsky L, Aides A, Shtok J, Harary S, Feris R, Giryes R, Bronstein AM (2019) Laso: Label-set operations networks for multi-label few-shot learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 6548–6557
14. Chu W-H, Li Y-J, Chang J-C, Wang Y-CF (2019) Spot and learn: A maximum-entropy patch sampler for few-shot image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 6251–6260
15. Chen Z, Fu Y, Wang Y-X, Ma L, Liu W, Hebert M (2019) Image deformation meta-networks for one-shot learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8680–8689
16. Benaim S, Wolf L (2018) One-shot unsupervised cross domain translation. Neural Information Processing Systems,
17. Schwartz E, Karlinsky L, Shtok J, Harary S, Marder M, Feris R, Kumar A, Giryes R, Bronstein AM (2018) Delta-encoder: an effective sample synthesis method for few-shot object recognition. Advances in Neural Information Processing Systems, pp 2845–2855
18. Antoniou A, Storkey A, Edwards H (2017) Data augmentation generative adversarial networks. In: Proc Int Conf Learn Represent (ICLR) Workshop
19. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Advances in neural information processing systems, 27

20. Wang X, Yu F, Wang R, Darrell T, Gonzalez JE (2019) Tafe-net: Task-aware feature embeddings for low shot learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1831–1840
21. Karlinsky L, Shtok J, Harary S, Schwartz E, Aides A, Feris R, Giryes R, Bronstein AM (2019) Repmet: Representative-based metric learning for classification and few-shot object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5197–5206
22. Sung F, Yang Y, Zhang L, Xiang T, Torr PHS, Hospedales TM (2018) Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1199–1208
23. Snell J, Swersky K, Zemel RS (2017) Prototypical networks for few-shot learning. Neural Information Processing Systems
24. Vinyals O, Blundell C, Lillicrap T, Wierstra D et al (2016) Matching networks for one shot learning. Advances in Neural Information Processing Systems 29:3630–3638
25. Thelen M, Riloff E (2002) A bootstrapping method for learning semantic lexicons using extraction pattern contexts. In: Proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP 2002), pp 214–221
26. Cohen G, Afshar S, Tapson J, Van Schaik A (2017) Emnist: Extending mnist to handwritten letters. In: 2017 international joint conference on neural networks (IJCNN), pp 2921–2926
27. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25:1097–1105
28. Krizhevsky A, Hinton G, et al (2009) Learning multiple layers of features from tiny images
29. Ravi S, Larochelle H (2017) Optimization as a model for few-shot learning. In: International conference on learning representations (ICLR)
30. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning, pp 1126–1135
31. Zhou F, Wu B, Li Z (2018) Deep meta-learning: Learning to learn in the concept space. Advances in neural information processing systems
32. Chen Z, Fu Y, Zhang Y, Jiang Y-G, Xue X (2019) Sigal L (2019) Multi-level semantic feature augmentation for one-shot learning. IEEE Trans Image Process 28(9):4594–4605
33. Dhillon GS, Chaudhari P, Ravichandran A, Soatto S (2019) A baseline for few-shot image classification. In: International conference on learning representations (ICLR)

Zhengchao Zhang was born in 1996. He received the B.E. degree from Harbin University of Science and Technology. He is currently a Ph.D. candidate with the College of Computer Science and Technology, Harbin Engineering University. His main research interests include few-shot learning.

Hongbin Wang was born in 1979. He received the Ph.D. degree in computer application technology from Harbin Engineering University in 2010. He is member of CCF, and he is currently an associate professor with College of Computer Science and Technology, Harbin Engineering University. His interests include: Artificial Intelligence, Inversion Analysis of Underwater Environment, Deep Learning and Transfer Learning.

Nianbin Wang was born in 1967. He received the Ph.D. degree in computer science and technology from Harbin Institute of Technology. He is currently a professor with the College of Computer Science and Technology, Harbin Engineering University. He is a member of China Computer Federation (CCF). His main research interests include dataspace, deep learning, and data integration.