Semi-Supervised Few-shot Learning via Multi-Factor Clustering

Jie Ling[1]*, Lei Liao[1]*, Meng Yang[1,2]†, Jia Shuai[1]
[1] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China
[2]Key Laboratory of Machine Intelligence and Advanced Computing (SYSU), Ministry of Education
{lingj8, liaolei3, shuaij}@mail2.sysu.edu.cn,yangm6@mail.sysu.edu.cn

Abstract

The scarcity of labeled data and the problem of model overfitting have been the challenges in few-shot learning. Recently, semi-supervised few-shot learning has been developed to obtain pseudo-labels of unlabeled samples for expanding the support set. However, the relationship between unlabeled and labeled data is not well exploited in generating pseudo labels, the noise of which will directly harm the model learning. In this paper, we propose a Clustering-based semi-supervised Few-Shot Learning (cluster-FSL) method to solve the above problems in image classification. By using multi-factor collaborative representation, a novel Multi-Factor Clustering (MFC) is designed to fuse the information of few-shot data distribution, which can generate soft and hard pseudo-labels for unlabeled samples based on labeled data. And we exploit the pseudo labels of unlabeled samples by MFC to expand the support set for obtaining more distribution information. Furthermore, robust data augmentation is used for support set in the fine-tuning phase to increase the labeled samples' diversity. We verified the validity of the cluster-FSL by comparing it with other few-shot learning methods on three popular benchmark datasets, miniImageNet, tieredImageNet, and CUB-200-2011. The ablation experiments further demonstrate that our MFC can effectively fuse distribution information of labeled samples and provide high-quality pseudo-labels. Our code is available at: https://gitlab.com/smartllvlab/cluster-fsl

1. Introduction

Deep learning methods based on neural networks [34] have made great breakthroughs and been widely used in the field of computer vision [15] [32] [23]. Unfortunately, deep learning models require a massive amount of labeled data to learn huge-scale parameters. Moreover, the scenes in real life are complex and diverse. For instance, in some practical problems, there exist objects that are rarely observed, such as the faces of the suspects, rare animal photos, and so on, which makes it difficult to collect labeled data. With the inspiration of the human vision system, much attention has been paid to the development of few-shot learning [29] [38] [27] [7], which can release the strong demand for data used for training deep models to some extent.

The goal of few-shot learning is to learn a new concept or behavior from a very small number of samples based on experience. Due to the particularity of few-shot learning, how to alleviate the overfitting problem that the model hardly fits the distribution of the categories caused by the scarcity of samples, has always been the focus and difficulty of few-shot learning. With the development of few-shot learning, methods have been proposed to solve the above difficulty from three perspectives: model, data, and learning algorithm. Model-based methods [28] [39] aim to learn the interactive information between samples by designing a model that adapts to the few-shot case. As for data, researchers have explored various methods to enrich training data, such as data augmentation [28] [1] [48], which preprocesses data through flipping, cropping, translation, rotation, and zooming. However, data augmentation requires expensive labor costs and relies heavily on domain knowledge, resulting in some data augmentation methods being specific to datasets. Few-shot image classification based on learning methods [45] [14] [38] aim to improve the generalization of the model by using meta-learning or transfer learning strategy, which learns meta-knowledge or transferrable knowledge shared in different subtasks. However, the performance of few-shot learning is still far from satisfactory because the small set of labeled data cannot provide rich and essential information for recent model learning.

When the task has few-shot labeled samples and additional unlabeled data for new categories, a straightforward approach is to utilize the unlabeled samples to alleviate the scarcity of labeled samples. Unlabeled data has been effectively explored by semi-supervised learning methods, including a combination of weak data augmentation and strong augmentation [37], consistent regularization [16],

*Equal contribution, † Corresponding author.

adversarial perturbation regularization [40], achieving very exciting performance without using a large number of annotated samples. If unlabeled data can be explored by combing semi-supervised learning with few-shot learning (e.g., mining pseudo-labels of unlabeled samples), problems such as data scarcity can be solved to some extent. Recent few-shot image classification combined with semi-supervised learning [31] [22] [38] focuses on improving the accuracy of pseudo-label prediction for unlabeled samples and is expected to obtain more correct pseudo-labels to expand the training set. However, existing methods cannot obtain completely correct pseudo-labels. How to avoid the influence of false pseudo-labels on model training and how to use sample distribution information to assist the acquisition of pseudo-labels are still current challenges.

To solve the aforementioned issues, we propose a *Clustering-based semi-supervised Few-Shot Learning (cluster-FSL)* method in image classification. By introducing labeled samples as factors of a cluster and representing unlabeled data on a multi-factor dictionary, we propose *Multi-Factor Clustering (MFC)* to guide the acquisition of pseudo-labels of unlabeled samples, which combines the distribution information of labeled and unlabeled samples for assisting clustering and obtaining more accurate pseudo-labels. In the fine-tuning stage, we design robust data augmentation to augment the support set and adopt the MFC module to predict soft labels of query samples, learning more discriminative feature distribution. In the testing stage, we use MFC instead of label propagation to assign more accurate pseudo-labels to unlabeled samples for expanding the test support set. The experimental results have shown that our cluster-FSL has achieved state-of-the-art performance (e.g., 2.45% improvement in the 5-way 1-shot scene of miniImageNet when the backbone is ResNet-12) on miniImageNet, tiered-ImageNet, and CUB-200-2011.

Our main contributions can be summarized as follows.

1. A novel multiple factor clustering (MFC) algorithm, which includes factors of labeled and unlabeled samples and exploits the distribution information among these factors via a multi-factor dictionary, is proposed, generating more accurate clustering results.

2. The proposed MFC is integrated into the fine-tuning stage and testing stage in a new way, with outputting high-quality pseudo labels for expanding the query and test support set effectively.

2. Related Works

2.1. Learning-based Few-shot Image Classification

Learning-based few-shot image classification algorithm aims to design a training and update mode suitable for few-shot scenarios. Transfer learning and meta-learning methods have been successfully applied to few-shot classification recently.

The transfer-learning-based methods pre-train the model on a large amount of data from the base class and adapt the pre-trained model to the few-shot learning task of identifying new categories. Kozerawski et al. [14] learn a transfer function that maps the embedding features extracted by the pre-trained model to the classification decision boundary. In [45], Yoo et al. identify groups of neurons within each layer of a deep network that shares similar activation patterns, and then use a training set for fine-tuning by group-by-group backward propagation. Transfer-learning-based methods can train a feature extractor with a powerful representation ability. To take advantage of these methods, We fine-tune a pre-trained model with the aid of the proposed MFC to improve the performance of our method.

Meta-learning-based methods, also known as learning to learn, aim to learn a paradigm that can be adapted to recognize new categories using few-shot train examples [46]. Researchers improve the meta-learning-based methods from the following aspects: the embedded module and shared distance measurement method [41] [36] [39], the initial distribution of model parameters [4] [25], and update strategies and rules of model parameters [21] [30]. As a classic and commonly used and effective method for solving few-shot classification tasks, although these methods have achieved rapid development, the problem of sample scarcity is still a remaining challenge.

Recently, Rodríguez et al. [33] propose the use of Embedding Propagation (EP) as an unsupervised non-parametric regularizer for manifold smoothing and apply EP to a transductive classifier. However, this method has a limited expansion of the labeled samples, which prevents further improvement of the performance. To solve the issue of EPNet [24], we use the proposed MFC module to obtain pseudo-labels of unlabeled samples, which can alleviate the scarcity of labeled samples effectively.

2.2. Semi-supervised Few-shot Methods Based on Pseudo-label Acquisition

The few-shot image classification methods combined with semi-supervised learning expand the labeled data by obtaining the pseudo-labels of the unlabeled data and alleviates the problem of sample scarcity, such as introducing class prototypes into the distribution of unlabeled samples to predict pseudo-labels in [31], using label propagation to obtain pseudo-labels in [22], using a pre-trained classifier to predict pseudo-labels in [46], and leveraging the manifold structure of labeled and unlabeled data distribution to predict pseudo-labels in [17]. After obtaining the pseudo-labels, these methods directly use them as labeled samples for model fine-tuning and training. However, they ignore the impact of incorrectly pseudo-labeled data on model training.

To reduce the impact of incorrect pseudo labels, Wu et al. [44] set a priority to select informative unlabeled samples to be used in the subsequent training process, while Sun et al. [20] proposed to limit the number of unlabeled samples selected in each round of optimization and preferentially select pseudo-labeled samples with high confidence. Huang et al. [11] presented a pseudo-loss confidence metric (PLCM), which maps pseudo-labeled data of different tasks to a unified metric space and estimates the confidence of pseudo-labeled according to the distribution component confidence of its pseudo-loss. However, the method of using a trained classifier to predict unlabeled samples one by one individually ignores the information at the data distribution level. The influence of the intra-class and inter-class relationship between unlabeled samples and the distribution information of unlabeled samples and labeled samples on the acquisition of pseudo-labels is not considered.

Recently, Huang et al. proposed Poisson Transfer Network (PTN) [10], which is a transfer-learning-based semi-supervised few-shot method. PTN model improves the capacity of mining the relations between the labeled and unlabeled data for graph-based few-shot learning, and our cluster-FSL is similarly dedicated to enhancing such capacity. However, our proposed MFC utilizes multiple factors to build a feature dictionary, which allows the overall clustering to effectively strengthen the relationship between labeled and unlabeled data, making the clustering being more concise and more interpretable. Wang et al. [42] proposed that the model iteratively selects the pseudo-labeled instances according to its credibility measured by Instance Credibility Inference (ICI) for classifier training. However, if the quality of pseudo-labels is not directly improved, the impact of incorrectly pseudo-labels on the model cannot be reduced. In addition, ICI focuses on solving a linear regression hypothesis by increasing the sparsity of the incidental parameters and ranking the pseudo-labeled instances with their sparsity degree, while our MFC constructs a dictionary by fusing the distribution of the labeled data, and uses the reconstruction error to calculate the distance when clustering.

3. Clustering-based semi-supervised Few-Shot Learning

Obtaining pseudo-labels of unlabeled samples with high confidence is a major challenge that the semi-supervised few-shot learning model needs to solve. To effectively solve the above challenge, we propose a novel model of clustered-based semi-supervised few-shot learning (cluster-FSL), which uses multi-factor clustering to obtain the high-quality pseudo labels of unlabeled data. Furthermore, our proposed cluster-FSL utilizes robust data augmentation and MFC modules to learn more discriminative features for improving performance. Here we first present a multi-factor clustering and then introduce three stages (e.g., pre-training, fine-tuning, and testing) of cluster-FSL in detail.

3.1. Problem Definition

In the few-shot classification task, the dataset is divided into training set, validation set, and test set, namely $\mathcal{D} = \{\mathcal{D}^{train}, \mathcal{D}^{val}, \mathcal{D}^{test}\}$. where $\mathcal{D}^{train} = \{X^{train}, Y^{train}\}$ contains all training data and corresponding labels, and $\mathcal{D}^{test} = \{X^{test}, Y^{test}\}$ contains all test data and corresponding labels. All categories in the training set and test set are denoted by $\mathcal{C}^{train}$ and $\mathcal{C}^{test}$, respectively, such that $\mathcal{C}^{train} \cap \mathcal{C}^{test} = \varnothing$. The categories in the validation set do not overlap with $\mathcal{C}^{train}$ and $\mathcal{C}^{test}$. Referring to episodic learning, we constructs n independent few-shot tasks to form episodic set $\Gamma = \{T_i^{tr}, T_i^{test}\}_{i=1}^{n}$. For each training task T_i^{tr}, we randomly select N categories from the training set, and randomly select K samples from each category to form a training support set $S^{tr} = \{x_i^s, y_i^s\}_{i=1}^{N\times K}$, where $x_i^s \in X^{train}$ and $y_i^s \in Y^{train}$. From the same N categories, we select q non-repeated samples to form the training query set $Q^{tr} = \{x_i^q, y_i^q\}_{i=1}^{N\times q}$, where $x_i^q \in X^{train}$ and $y_i^q \in Y^{train}$. The validation set $\mathcal{D}^{val}$ is used to determine the best model, and the model with the highest accuracy on the validation set will be selected. For each testing task T_i^{test}, the model also randomly selects N categories from the test set, and randomly select K samples from each category to obtain the test support set $S^{test} = \{x_j^s, y_j^s\}_{j=1}^{N\times K}$, and select q non-repeated samples from each category to form the test query set $Q^{test} = \{x_j^q, y_j^q\}_{j=1}^{N\times q}$. In addition, we randomly select u samples from the unselected samples under the N categories included in the test set, and remove the labels to form an unlabeled set $U^{test} = \{x_j^u\}_{j=1}^{N\times u}$.

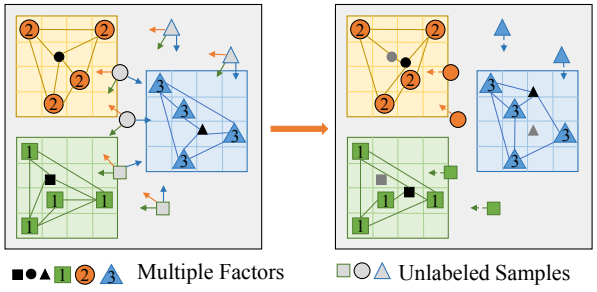

Figure 1. The clustering process of MFC. Among them, black dots represent the cluster centers after the update, and gray-black dots represent the cluster center before the update.

3.2. Multi-Factor Clustering

Given a labeled set L and an unlabeled set U, we propose a clustering method called multi-factor clustering (MFC) to predict the pseudo-labels of samples in U by using the distance between samples and factors (e.g. cluster center and

labeled samples). By utilizing multiple factors to construct a feature dictionary and using reconstruction error as a distance measurement, our MFC is suitable for the few-shot scenario to obtain pseudo-labels of unlabeled samples.

Pseudo-label Acquisition There are a small number of labeled samples in the few-shot learning task, and the number of categories is known. Under the setting of N-way K-shot, the clusters in MFC are expressed as $C = \{C_i\}_{i=1}^{N}$, where N is the number of clusters. The center of each cluster is initialized to the within-class mean of the labeled samples by

$$\boldsymbol{c}_i^* = \frac{1}{K}\sum_{j=1}^{K} \boldsymbol{l}_j^i \tag{1}$$

where $\boldsymbol{l}_j^i \in L$ is the j^{th} embedding representation of the i^{th} category labeled sample and $\boldsymbol{c}_i^*$ represents the center of i^{th} cluster.

Traditional clustering method [31] only considers the distance between an unlabeled sample and each cluster center, while ignoring data distribution information and the feature information of the K labeled samples belonging to the same category. We regard the factor of a cluster as an embedded representation that can represent the cluster (e.g. the cluster center and a labeled sample) and then propose multi-factor clustering (MFC), which sets up multiple factors for each cluster and improves the calculation method of the distance from the unlabeled sample to each cluster.

The factors $\boldsymbol{F}$ of the i^{th} cluster in MFC are defined as $\boldsymbol{F}_i = [l_1^i, \cdots, l_K^i, \boldsymbol{c}_i^*]$, which includes a cluster center and the samples belonging to i^{th} cluster in the labeled set L.

To release the small-sample-size problem, the proposed MFC adopts all class-specific factors as a dictionary to collaboratively represent an unlabeled sample $\boldsymbol{u}_i \in U$:

$$\tilde{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}_i} ||\boldsymbol{u}_i - \boldsymbol{F}\boldsymbol{\beta}_i||_2^2 + \epsilon||\boldsymbol{\beta}_i||_2^2 \tag{2}$$

where ϵ is a constant of 0.01 to regularize the representation, $\tilde{\boldsymbol{\beta}}_i = [\tilde{\boldsymbol{\beta}}_{i,1}, \tilde{\boldsymbol{\beta}}_{i,2}, \cdots, \tilde{\boldsymbol{\beta}}_{i,N}]^T$, $\tilde{\boldsymbol{\beta}}_{i,j} \in \mathbb{R}^{(K+1)\times 1}$ is the sub-coefficient vector associated to the j^{th} cluster. We refer to the solution in collaborative representation [47] to solve the $\tilde{\boldsymbol{\beta}}_i$ in Eq. 2. Then we use the reconstruction errors associated with each cluster as the distance from the unlabeled sample to the cluster as shown in Fig.1. The reconstruction error is defined as:

$$d_{i,j} = ||\boldsymbol{u}_i - \boldsymbol{F}_j\tilde{\boldsymbol{\beta}}_{i,j}||_2^2 \tag{3}$$

Calculating this reconstruction error takes into account the factors of all categories, which make different cluster factors competitively represent the unlabeled data, outputting a more robust clustering.

According to the minimum reconstruction error from the unlabeled data $\boldsymbol{u}_i$ to each cluster, the clustering result is $\alpha_i = \arg\min_{1\le j\le N} d_{i,j}$. And each cluster is updated by,

$$\begin{aligned} \tilde{C}_{\alpha_i} &= C_{\alpha_i} \cup \{\boldsymbol{u}_i\} \\ \boldsymbol{c}_j^* &= \frac{1}{|\tilde{C}_j|}\sum_{\boldsymbol{z}_k \in \tilde{C}_j} \boldsymbol{z}_k \quad \forall j = 1, 2, \cdots, N \end{aligned} \tag{4}$$

where $\boldsymbol{c}_j^*$ is the j^{th} cluster center, which is used to update the cluster factor.

Multiple factors can make the clustering more accurate compared to only one factor because they can better represent the embedding manifold of the cluster. After several iterations, when the clusters and cluster centers do not change, the cluster is denoted as $C = \{C_1, C_2, \cdots, C_N\}$. We take the clustering result of the last iteration to obtain soft labels and hard labels for unlabeled data The soft label $\boldsymbol{p}_i = [p_{i,1}, p_{i,2}, \cdots, p_{i,N}]$ of the unlabeled sample $\boldsymbol{u}_i$ are defined as:

$$p_{i,j} = \frac{e^{-\log(d_{i,j}/\tau)}}{\sum_{k=1}^{N} e^{-\log(d_{i,k}/\tau)}} \tag{5}$$

where τ is a temperature parameter and $p_{i,j}$ denotes the probability that the unlabeled sample $\boldsymbol{u}_i$ is predicted to be the j^{th} class. Thus, the hard label is $\tilde{y}_i = \arg\max_j p_{i,j}$. The procedure details of MFC are shown in Algorithm. 1.

Algorithm 1 Process description of Multi-Factor Clustering

Input: Labeled set $L = \{\boldsymbol{l}_i, y_i\}_{i=1}^{N\times K}$, unlabeled set $U = \{\boldsymbol{u}_i\}_{i=1}^{N\times u}$
Output: Pseudo-labels of samples in unlabeled set $P = \{\boldsymbol{p}_i, \tilde{y}_i\}_{i=1}^{N\times u}$

```
while True do
    for Each unlabeled sample u_i ∈ U do
        Compute the reconstruction error d_{i,j} by Eq.(3)
        Obtain the clustering result α_i = arg min_{1≤j≤N} d_{i,j}
        Update the clusters and the cluster centers by Eq.(4)
    end for
    if The cluster centers remain unchanged then
        Calculate the soft pseudo label p_i and hard pseudo label ỹ_i by Eq.(5).
        Break
    end if
end while
```

3.3. Training procedure

To train a model with good generalization ability, we firstly use $\mathcal{D}^{train}$ to train a pre-trained model by fully-supervised learning and self-supervised learning. Then we use the episode set Γ to fine-tune the pre-trained model by multi-factor clustering and label propagation [12].

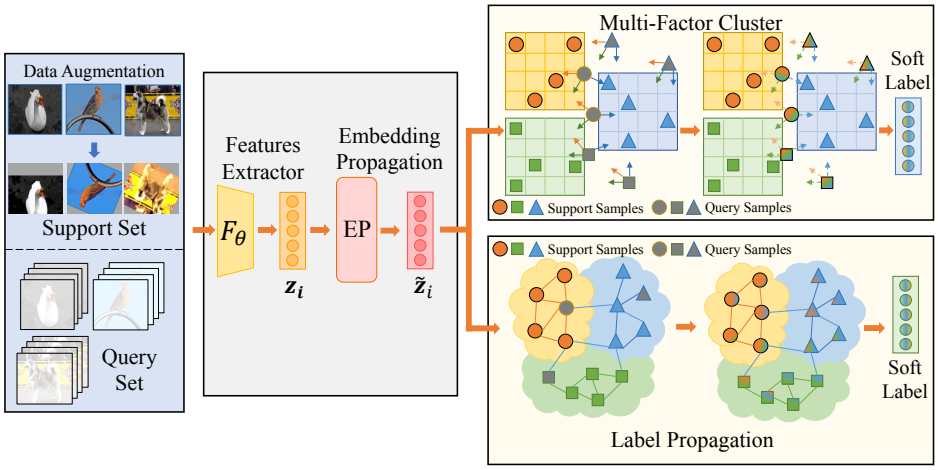

Figure 2. The framework of the fine-tuning phase. At first, we use more robust data augmentation to augment the support set. The samples from the augmented support set and the query set undergo feature extraction and embedding propagation to obtain features $\tilde{z}$. Then based on the support set, the MFC module and the label propagation module obtain the soft labels of the query set samples.

3.3.1 Pre-training phase

We use the cross-entropy loss to fit the model's classification predictions to the ground-truth labels. For self-supervised learning, another classifier that predicts image rotations is added to this model which is same as [33]. In this case, the pre-trained model can extract representative embedding from an image for few-shot tasks.

3.3.2 Fine-tuning with data augmentation and MFC

In this phase as shown in Fig.2, we introduce data augmentation for support set and a cross-entropy loss based on MFC to learn an encoder with good generalization that can extract powerful embeddings. We retain the encoder of the pre-trained model and discard its classifiers to fine-tune the encoder by episodic learning. For a training task T_i, it contains a support set S^{tr} and a query set Q^{tr}. We use Randaugment [3] for support set to avoid over-fitting. Because the size of the support set is too small in few-shot learning, the model can easily overfit a small amount of data. However, Randaugment can dramatically change the image to increase the difficulty of the model to fit data distribution. The augmented support set is $ag(S^{tr}) = \{ag(x_i^s), y_i^s\}_{i=1}^{N\times K}$. After the model extracts embeddings for all samples of the augmented support set and the query set, we use embedding propagation [33] to process these features:

$$\tilde{\boldsymbol{z}}_i = ep(\boldsymbol{z}_i, \boldsymbol{Z}) \tag{6}$$

where $\boldsymbol{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_{N\times(K+q)}\}$ denotes the embedding set of the images from $ag(S^{tr})$ and Q^{tr}. $ep(.,.)$ is the embedding propagation which can increase the smoothness of the embedding manifold by the Euclidean distances between embeddings.

We construct two cross-entropy losses based on multi-factor clustering and label propagation [12], respectively:

$$l_{ft} = \lambda\, l_{MFC} + (1-\lambda)\, l_{LP} \tag{7}$$

where λ is a hyperparameter, l_{MFC} is the cross-entropy loss based on multi-factor clustering, and l_{LP} is the cross-entropy loss based on label propagation.

For multi-factor clustering, $ag(S^{tr})$ can be considered as labeled set and Q^{tr} can be considered as unlabeled set in a training task T_i. We can obtain the soft pseudo-labels of training query samples by MFC, and then use ground-truth labels to calculate the cross-entropy loss:

$$l_{MFC} = -\frac{1}{q}\sum_{i=1}^{q} \log p_{i,y_i^q} \tag{8}$$

where p_{i,y_i^q} denotes the probability that the i^{th} query sample is predicted to be the y_i^q category in the MFC. The probability $p_{i,j}$ is defined in Eq.(5).

Similarly, the cross-entropy loss based on label propagation is defined as:

$$l_{LP} = -\frac{1}{q}\sum_{i=1}^{q} \log \tilde{p}_{i,y_i^q} \tag{9}$$

where $\tilde{p}_{i,y_i^q}$ is the predicted probability obtained by label propagation [12].

3.3.3 Expanding support set by MFC in testing phase

In this phase as shown in Fig.3, we exploit the samples in the test support set S^{test} and the unlabeled set U^{test} to infer

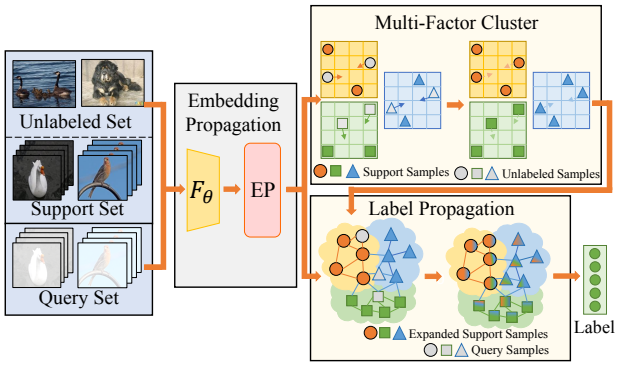

Figure 3. The framework of the testing phase. After features extraction of unlabeled samples and supporting samples, the MFC module is used to obtain pseudo-labels of unlabeled samples to expand the support set. The expanded support set is used for label propagation with the query set, and the labels of the query set samples are obtained.

the categories of samples in the test query set Q^{test}. Different from EPNet [33], we use MFC to assign pseudo-labels to unlabeled data and expand the test support set. And then the categories of samples in test query set Q^{test} are obtained by label propagation using the expanded support set. The advantage of using MFC is that MFC can generate high-quality pseudo-labels for unlabeled data in few-shot tasks.

We extract the embedding representations by the trained encoder from $\{S^{test}, Q^{test}, U^{test}\}$, and apply embedding propagation on these embedding representations. S^{test} is viewed as the labeled set to cluster unlabeled data in U^{test} by MFC. We select the unlabeled data with high-confidence pseudo-labels to form a subset by:

$$\tilde{U}^{test} = \{(x_i^u, \tilde{y}_i) | x_i^u \in U^{test}\} \tag{10}$$

where $\tilde{y}_i$ is the hard pseudo-label obtained by MFC for x_i^u . The test support set is updated by:

$$S^{'} = S^{test} \cup \tilde{U}^{test} \tag{11}$$

At last, we predict the labels of the data in Q^{test} by label propagation based on expanded test support set $S^{'}$.

4. Experiment

4.1. Datasets

The miniImageNet dataset is a subset extracted from the ImageNet [34] dataset by Vinyals et al. [41]. Our experiment uses all 60,000 images in the dataset, and according to the standard of the traditional few-shot learning methods [41] [39] [5] [13], the 100 categories are divided into 64 training categories, 16 verification categories and 20 test categories.

The tieredImageNet is proposed by Ren et al. [31] for few-shot learning, and is a subset extracted from the ImageNet [34] dataset. It contains 34 superclasses which can be divided into 608 categories, with a total of 779,165 images. Our experiments use all the image data in this dataset and divide these superclasses into 20 training superclasses (351 categories), 6 validation superclasses (97 categories) and 8 test superclasses (160 categories), following the traditional few-shot learning methods [6] [35].

The CUB [2] [8] dataset is a fine-grained dataset based on CUB200 [43], which contains 200 classes and 11,788 images split in 100 base, 50 validation and 50 novel classes.

4.2. Model Settings

In the pre-training phase, the cluster-FSL model is trained on 200 epochs with a batch size of 128, using all training categories and data. The dropout rate is 0.1, the weight decay is 0.0005, and the momentum is 0.9. The pre-trained network is updated using the stochastic gradient descent algorithm, with an initial learning rate of 0.1. The learning rate is multiplied by 0.1 when the model reaches a plateau that the validation loss had not improved for 10 epochs. When the structure of the feature extractor is ResNet-12, the output feature dimension n is 512. When the structure is WRN-28-10, the output feature dimension n is 640. In the fine-tuning phase, the cluster-FSL model is trained on 200 epochs, the number of iterations is 600, the weight decay is 0.0005, and the momentum is 0.9. The model is updated using the stochastic gradient descent method. The learning rate is 0.001, and it is multiplied by 0.1 when the model reaches a plateau. In the testing phase, the cluster-FSL model is tested on 1000 epochs, and the average of the accuracy of the model classification results is used as the evaluation result. The number of categories in each few-shot task N is 5. For each category, the number of supported samples K is 1 and 5, the number of query samples q is 15, and the number of unlabeled samples u is 100. The hyperparameter λ is 0.8, the temperature parameter τ is 0.1 and the hyperparameter ϵ is 0.01. The number of iterations of the clustering process is 10.

4.3. Comparative Experiment

The baseline models for comparative experiment are EPNet [33], ICI [42] and some classic or state-of-the-art few-shot learning methods, which are TADAM [26], MTL [38], MetaOpt-SVM [18], CAN [9], LST [39] and LEO [35]; semi-supervised few-shot learning methods, which are TPN [22], TransMatch [46] and PLAIN [19]; and graph-network-based method, wDAE-GNN [6]. In addition, comparative experiments were conducted for PTN [10] and for clustering methods proposed by Ren et al. [31], such as Soft K-Means, Soft K-Means+Cluster, Masked Soft K-Means. We use quantitative analysis and comparison meth-

Table 1. The average classification accuracy of 1000 few-shot tasks in the 5-way 1-shot and 5-way 5-shot scenarios of the cluster-FSL model on the miniImageNet dataset.

Methods	5-way 1-shot	5-way 5-shot
ResNet-12		
TADAM	58.50±0.30%	76.70±0.30%
MTL	61.20±1.80%	75.50±0.80%
MetaOpt-SVM	62.64±0.61%	78.60±0.46%
CAN	67.19±0.55%	80.64±0.35%
LST	70.10±1.90%	78.70±0.80%
EPNet	66.50±0.89%	81.06±0.60%
TPN	59.46%	75.65%
PLAIN	74.38±2.06%	82.02±1.08%
EPNet-SSL	75.36±1.01%	84.07±0.60%
cluster-FSL(our)	**77.81±0.81%**	**85.55±0.41%**
WRN-28-10		
Soft K-Means	50.09±0.45%	64.59±0.28%
Soft K-Means+Cluster	49.03±0.24%	63.08±0.18%
Masked Soft K-Means	50.41±0.31%	64.39±0.24%
LEO	61.76±0.08%	77.59±0.12%
wDAE-GNN	62.96±0.62%	78.85±0.10%
EPNet	70.74±0.85%	84.34±0.53%
TransMatch	62.93±1.11%	82.24±0.59%
ICI	71.41%	81.12%
EPNet-SSL	79.22±0.92%	88.05±0.51%
PTN	81.57±0.94%	87.17±0.58%
cluster-FSL(our)	**82.63±0.79%**	**89.16±0.35%**

ods to test the classification accuracy of the cluster-FSL model and evaluate the performance of the model. And we conduct experiments in two scenarios, 5-way 1-shot and 5-way 5-shot, which are two common scenarios in the field of few-shot learning. Using ResNet-12 and WRN-28-10 as the backbone, we get the comparative experimental results on miniImageNet, tieredImage and CUB-200-2011 datasets as shown in Table 1, Table 2 and Table 3.

In Table 1, compared with the best performance EPNet-SSL [33] on the miniImageNet dataset, when the backbone is ResNet-12, the cluster-FSL model achieves 2.45% and 1.48% improvement in the 1-shot and 5-shot scenes respectively. When the backbone is WRN-28-10, our cluster-FSL achieves 1.06% and 0.73% improvement at 1-shot and 5-shot scenes, respectively. Table 2 shows the comparative experimental results on the tieredImageNet dataset. Our cluster-FSL achieves 2.10% and 1.46% imporvement at the 1-shot and 5-shot when the backbone is ResNet-12, which is the state-of-the-art performances. When the backbone is WRN-28-10, our cluster-FSL also shows a 1.04% and 1.04% improvement under 1-shot and 5-shot,respectively. On CUB-200-2011 dataset, Table 3 shows the improvement of our cluster-FSL compared to EPNet [33] and ICI [42].

The results of comparative experiments illustrate that our cluster-FSL has excellent performance under both 1-shot and 5-shot. Moreover, our MFC has stronger ability to strengthen the relationship between multiple factors.

Table 2. The average classification accuracy of 1000 few-shot tasks in the 5-way 1-shot and 5-way 5-shot scenarios of the cluster-FSL model on the tieredImageNet dataset.

Methods	5-way 1-shot	5-way 5-shot
ResNet-12		
MetaOpt-SVM	65.99±0.72%	81.56±0.53%
CAN	73.21±0.58%	84.93±0.38%
LST	77.70±1.60%	85.20±0.80%
EPNet	76.53±0.87%	87.32±0.64%
PLAIN	82.91±2.09%	88.29±1.25%
EPNet-SSL	81.79±0.97%	88.45±0.61%
cluster-FSL(our)	**83.89±0.81%**	**89.94±0.46%**
WRN-28-10		
Soft K-Means	51.52±0.36%	70.25±0.31%
Soft K-Means+Cluster	51.85±0.25%	69.42±0.17%
Masked Soft K-Means	52.39±0.44%	69.88±0.20%
LEO	66.33±0.05%	81.44±0.09%
wDAE-GNN	68.16±0.16%	83.09±0.12%
EPNet	78.50±0.91%	88.36±0.57%
ICI	85.44%	89.12%
EPNet-SSL	83.68±0.99%	89.34±0.59%
PTN	84.70±1.14%	89.14±0.71%
cluster-FSL(our)	**85.74±0.76%**	**90.18±0.43%**

Table 3. The average classification accuracy of 1000 few-shot tasks in the 5-way 1-shot and 5-way 5-shot scenarios of the cluster-FSL model on the CUB-200-2011 dataset. $(\cdot)\dagger$denotes this method uses ResNet-12 as the backbone, while $(\cdot)\ddagger$denotes this method uses WRN-28-10 as the backbone.

Methods	5-way 1-shot	5-way 5-shot
EPNet†	82.85±0.81%	91.32±0.41%
cluster-FSL(our)†	**87.36±0.71%**	**92.17±0.31%**
ICI‡	91.11%	92.98%
EPNet‡	87.75±0.70%	94.03±0.33%
cluster-FSL(our)‡	**91.80±0.58%**	**95.07±0.23%**

Table 4. In 5-way 5-shot setting, the impact of different methods of obtaining pseudo-labels based on cluseter-FSL.

Methods	miniImageNet	tieredImageNet
Label propagation	87.99±0.37%	89.45±0.45%
Kmeans	88.05±0.40%	88.48±0.53%
MFC(our)	**89.16±0.35%**	**90.18±0.43%**

4.4. Ablation Studies

In this section, we have done a series of complete ablation experiments for the role of each module at each stage,

Table 5. In 5-way 1-shot setting, the impact of MFC and label propagation (LP) on fine-tuning and testing phase of cluseter-FSL.

Settings		Fine-tune		
		LP	MFC	MFC + LP
Test	LP	78.71%	79.32%	79.81%
	MFC + LP	**80.88%**	**82.56%**	**82.70%**

Table 6. In 5-way 1-shot setting, the impact of MFC and data augmentation on fine-tuning phase of cluseter-FSL.

Data Augmentation	MFC	ACC
×	×	79.87±1.10%
✓	×	80.88±1.07%
×	✓	81.73±1.09%
✓	✓	**82.70±1.03%**

which are unified with WRN-28-10 as the backbone and evaluate the accuracy under 5-way 1-shot or 5-way 5-shot. To verify that MFC is effective in obtaining pseudo-labels with higher correctness, we compared different methods of obtaining pseudo-labels, such as label-propagation [33] and Kmeans, on the miniImageNet and tieredImageNet under the cluster-FSL, as shown in Table 4. The experimental results show that compared with label propagation and Kmeans clustering, the MFC module has improvements on the miniImageNet and tieredImageNet datasets, which shows that multiple factors contain more sample distribution information and improve classification accuracy.

In order to verify that the MFC and label propagation modules included in the cluster-FSL play a crucial role in the experimental results, a full-scale ablation experiment is conducted for each component of the fine-tuning and testing phases, as shown in Table 5. For the fine-tuning stage, we compared the model with only label propagation, only MFC module, and both. For the three situations of the fine-tuning stage, we explored the impact of different settings in the testing stage on performance. For testing stage, the "MFC+LP" in Table 5 means that the testing model uses the MFC to obtain the expanded support set and then uses the label propagation to predict the labels of the query set. And the "LP" means that we only use label propagation to infer the labels of the query set. The results show that the MFC plays a important role in the model, and it can improve the performance of the model.

During the fine-tuning phase of the model, the support set was extended with data augmentation and the query samples were predicted by MFC module. Therefore, we carry out ablation experiments for both parts and the results are shown in Table 6, which shows the effect of both data augmentation and MFC.

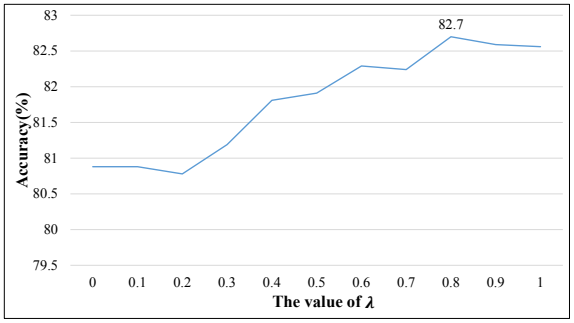

Figure 4. The impact of different λ on the miniImageNet dataset in 5-way 1-shot setting.

4.5. Parameters Analysis

Our cluster-FSL model has one hyperparameter, λ, where λ in Eq.(7) is used to control the proportion of MFC and label propagation during the fine-tuning phase. The parameter analysis experiment is done on the miniImageNet dataset with WRN-28-10 as the backbone, and the 5-way 1-shot setting is adopted. As shown in Fig.4, the best accuracy is obtained at 0.8, which can achieve 82.70%. When the value of λ is 0 or 1, it corresponds to not considering the effect of MFC or label propagation in Eq.(7) during the fine-tuning phase of the model.

5. Conclusion

We proposed a novel clustering-based semi-supervised few-shot learning (cluster-FSL) image classification method, which has effectively alleviated the problem of sample scarcity. In cluster-FSL, we presented a multi-factor clustering (MFC) algorithm by integrating factors of labeled and unlabeled data, which effectively improve the quality of pseudo-labels. Furthermore, in the model fine-tuning stage, we used more robust data augmentation to further augment the dataset and learned the model with the joint supervision of multi-factor clustering and label propagation. On the three benchmark datasets, our cluster-FSL has state-of-the-art performances than other few-shot learning methods.

Limitations: our proposed cluster-FSL needs to introduce extra unlabeled data and make an assumption that unlabeled data and labeled data are implicitly embedded in a manifold to ensure the generation of pseudo-labels.

Acknowledgement: This work was partially supported by the National Natural Science Foundation of China (Grants no. 62176271 and 61772568) and Guangdong Basic and Applied Basic Research Foundation (Grant no. 2019A1515012029).

References

[1] Sagie Benaim and Lior Wolf. One-shot unsupervised cross domain translation. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 2108–2118, 2018. 1

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019. 6

[3] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020. 5

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017. 2

[5] Victor Garcia and Joan Bruna. Few-shot learning with graph neural networks. In *6th International Conference on Learning Representations, ICLR 2018*, 2018. 6

[6] Spyros Gidaris and Nikos Komodakis. Generating classification weights with gnn denoising autoencoders for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21–30, 2019. 6

[7] Jianzhu Guo, Xiangyu Zhu, Chenxu Zhao, Dong Cao, Zhen Lei, and Stan Z Li. Learning meta face recognition in unseen domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6163–6172, 2020. 1

[8] Nathan Hilliard, Lawrence Phillips, Scott Howland, Artëm Yankov, Courtney D Corley, and Nathan O Hodas. Few-shot learning with metric-agnostic conditional embeddings. *arXiv preprint arXiv:1802.04376*, 2018. 6

[9] Ruibing Hou, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Cross attention network for few-shot classification. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 4003–4014, 2019. 6

[10] Huaxi Huang, Junjie Zhang, Jian Zhang, Qiang Wu, and Chang Xu. Ptn: A poisson transfer network for semi-supervised few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1602–1609, 2021. 3, 6

[11] Kai Huang, Jie Geng, Wen Jiang, Xinyang Deng, and Zhe Xu. Pseudo-loss confidence metric for semi-supervised few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8671–8680, 2021. 3

[12] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5070–5079, 2019. 4, 5

[13] Jongmin Kim, Taesup Kim, Sungwoong Kim, and Chang D Yoo. Edge-labeling graph neural network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11–20, 2019. 6

[14] Jedrzej Kozerawski and Matthew Turk. Clear: Cumulative learning for one-shot one-class image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3446–3455, 2018. 1, 2

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. 1

[16] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016. 1

[17] Michalis Lazarou, Tania Stathaki, and Yannis Avrithis. Iterative label cleaning for transductive and semi-supervised few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8751–8760, 2021. 2

[18] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10657–10665, 2019. 6

[19] Pan Li, Guile Wu, Shaogang Gong, and Xu Lan. Semi-supervised few-shot learning with pseudo label refinement. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021. 6

[20] Xinzhe Li, Qianru Sun, Yaoyao Liu, Qin Zhou, Shibao Zheng, Tat-Seng Chua, and Bernt Schiele. Learning to self-train for semi-supervised few-shot classification. *Advances in Neural Information Processing Systems*, 32:10276–10286, 2019. 3

[21] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017. 2

[22] Y Liu, J Lee, M Park, S Kim, E Yang, SJ Hwang, and Y Yang. Learning to propagate labels: Transductive propagation network for few-shot learning. In *7th International Conference on Learning Representations, ICLR 2019*, 2019. 2, 6

[23] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1

[24] Tsendsuren Munkhdalai and Hong Yu. Meta networks. In *International Conference on Machine Learning*, pages 2554–2563. PMLR, 2017. 2

[25] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 2

[26] Boris N Oreshkin, Pau Rodriguez, and Alexandre Lacoste. Tadam: task dependent adaptive metric for improved few-shot learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 719–729, 2018. 6

[27] Juan-Manuel Perez-Rua, Xiatian Zhu, Timothy M Hospedales, and Tao Xiang. Incremental few-shot object detection. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, pages 13846–13855, 2020. 1

[28] Hang Qi, Matthew Brown, and David G Lowe. Low-shot learning with imprinted weights. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5822–5830. IEEE, 2018. 1

[29] Limeng Qiao, Yemin Shi, Jia Li, Yaowei Wang, Tiejun Huang, and Yonghong Tian. Transductive episodic-wise adaptive metric for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3603–3612, 2019. 1

[30] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. 2016. 2

[31] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B Tenenbaum, Hugo Larochelle, and Richard S Zemel. Meta-learning for semi-supervised few-shot classification. In *International Conference on Learning Representations*, 2018. 2, 4, 6

[32] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015. 1

[33] Pau Rodríguez, Issam Laradji, Alexandre Drouin, and Alexandre Lacoste. Embedding propagation: Smoother manifold for few-shot classification. In *European Conference on Computer Vision*, pages 121–138. Springer, 2020. 2, 5, 6, 7, 8

[34] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015. 1, 6

[35] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In *International Conference on Learning Representations*, 2018. 6

[36] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 30:4077–4087, 2017. 2

[37] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33, 2020. 1

[38] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2019. 1, 2, 6

[39] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018. 1, 2, 6

[40] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 1195–1204, 2017. 2

[41] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29:3630–3638, 2016. 2, 6

[42] Yikai Wang, Chengming Xu, Chen Liu, Li Zhang, and Yanwei Fu. Instance credibility inference for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12836–12845, 2020. 3, 6, 7

[43] Peter Welinder, Steve Branson, Takeshi Mita, Catherine Wah, Florian Schroff, Serge Belongie, and Pietro Perona. Caltech-ucsd birds 200. 2010. 6

[44] Yu Wu, Yutian Lin, Xuanyi Dong, Yan Yan, Wanli Ouyang, and Yi Yang. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5177–5186, 2018. 3

[45] Donghyun Yoo, Haoqi Fan, Vishnu Naresh Boddeti, and Kris M Kitani. Efficient k-shot learning with regularized deep networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 1, 2

[46] Zhongjie Yu, Lin Chen, Zhongwei Cheng, and Jiebo Luo. Transmatch: A transfer-learning scheme for semi-supervised few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12856–12864, 2020. 2, 6

[47] Lei Zhang, Meng Yang, and Xiangchu Feng. Sparse representation or collaborative representation: Which helps face recognition? In *2011 International conference on computer vision*, pages 471–478. IEEE, 2011. 4

[48] Yabin Zhang, Hui Tang, and Kui Jia. Fine-grained visual categorization using meta-learning optimization with sample selection of auxiliary data. In *European Conference on Computer Vision*, pages 241–256. Springer, 2018. 1