https://doi.org/10.2352/EI.2025.37.8.IMAGE-265

Convolutional Shared Dictionary Module for Few-Shot Learning

Tong Zhou [a], Changyin Dong [a,*], Zhen Wang [b], Bo Chang [b], Junshu Song [b], Bin Shen [c], Baodi Liu [d], Pengfei He [b]
[a] School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, China
[b] Sinopec Group Shengli Oil Administration Bureau co., Dongying, China
[c] Celonis, New York, United States
[d] School of Control Science and Engineering, China University of Petroleum (East China), Qingdao, China
* Corresponding author

Abstract

Few-shot learning is the most prevalent problem which has attracted lots of attention in recent years. It is a powerful research method in the case of limited training data. Simultaneously, few-shot learning methods based on metric learning mainly measure the similarity of feature embeddings between the query set sample and each class of support set samples. Therefore, how to design a CNN-based feature extractor is the most crucial problem. Nowadays, the existed feature extractors are obtained via training the standard convolutional networks (e.g., ResNet), which merely focuses on the information inside each image. However, the relations among samples may also be beneficial to promote the performance of the few-shot learning task. This paper proposes a Convolutional Shared Dictionary Module (CSDM) to find the hidden structural information among samples for few-shot learning and reduce the dimension of sample features to remove redundant information. Therefore, the learned dictionary is more easily adapt to the novel class, and the reconstructed features are more discriminative. Moreover, the CSDM is a plug-and-play module and integrates the dictionary learning algorithm into the feature embedding. Experimental results on several benchmark datasets have demonstrated the effectiveness of the proposed CSDM.

INTRODUCTION

In recent years, deep learning network has made significant progress in the object detection [1], image classification [2–5], image segmentation [6], etc. In contrast, there are limited training data in numerous scenes in reality. Numerous effective few-shot learning methods have recently been proposed to solve low performance caused by insufficient data samples. Unlike traditional methods, the train set and test set in few-shot learning are independent of each other, and each contains the support set and query set. Generally, there are between 1 and 20 examples per class in the support set, which puts forward higher requirements on the model's generalization performance and how to make fair use of the finite data. It is entirely different from the traditional classification problem. Simultaneously, few-shot image classification tasks mainly include data-based methods [7], optimization-based methods [8], and metric-learning-based methods [9]. Among which the few-shot image classification based on metric learning is the most widely studied.

The few-shot image classification method based on metric learning effectively solved the problem with limited image data, which usually includes feature extraction network and metric learning module, such as Prototypical Network [10]. First, the feature extraction network extracts each class of samples' features in the support set and query set to represent the sample data. The metric learning module is then adopted to implement the prediction of the query set samples' labels and update feature embedding network parameters by minimizing the similarity between the support set and query set. This kind of method is simple and effective. However, the feature extraction network merely extracts the feature independently, without considering the distribution among samples.

This paper introduces the Convolutional Shared Dictionary Module (CSDM) to consider the correlation among samples fully. After the feature extraction network, CSDM is used to reconstruct all samples' embedding features. The CSDM algorithm can obtain the samples' distribution according to learning the pattern of embedding features and thus transform the embedding features to a more discrimination subspace. Specifically, CSDM uses the feature embeddings extracted from the feature extraction network to learn and build a shared dictionary and then obtains the representation of samples through the linear combination of dictionary atoms in the shared dictionary. CSDM can be considered a feature reconstruction layer inserted into the feature extraction network. It provides more efficient features by considering the global features' distribution. Figure 1 shows how to apply the CSDM to the metric-learning-based few-shot learning task.

The main contributions in this paper are as follow,

- We propose an effective Convolutional Shared Dictionary Module and integrate it into the deep feature extraction network. This framework is called the convolutional shared dictionary feature extraction framework. CSDM can effectively obtain the relationship among samples to reconstruct the feature embeddings. The reconstructed feature embeddings are more discriminative and thus more suitable for Domain Shift tasks.
- To improve the learning ability of the Convolutional Shared Dictionary Module, CSDM uses the ℓ_2 norm on the dictionary to prevent overfitting and enhance the robustness to data noise. To facilitate optimization, we also propose updating the dictionary during backpropagation.
- The proposed module improves performance $0.2\% \sim 2.4\%$ over baseline systems (Prototypical Network [10]) on 5-way N-shot ($N = 5, 10, 15$) on four benchmark datasets for few-shot learning.
- The proposed CSDM is a plug-and-play module and can be integrated into any feature embedding networks.

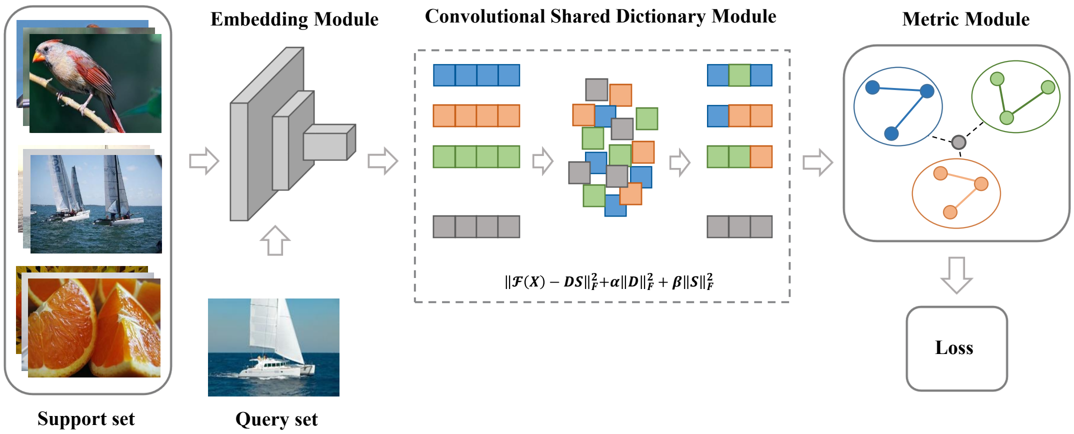


Figure 1. *The figure shows a schematic diagram of CSDM in the 3-way 3-shot classification task. It consists of a feature extraction network, convolution dictionary module, and measurement module. (1) The feature extraction network is to extract the feature embeddings of samples. (2) CSDM uses feature embeddings to build a shared dictionary and transform the feature embeddings into a more efficient representation. (3) The metric learning module predicts the category of the query set and updates the parameters of the feature extraction network and convolution dictionary module via the cross-entropy loss.*

Related Work

Metric Learning

The few-shot learning method based on metric learning models the distance distribution among samples, making the same class samples closer and the different classes farther. Finally, metric the distance of the feature vectors of the samples to complete the classification task. Koch G *et al.* [11] proposed the Siamese Network by permuting and combining samples into pairs and judging whether they belong to the same class by the distance of the sample pairs. Vinyals O *et al.* [12] proposed the Matching Network to construct different feature extraction models for support set and query set, and weighted sum the prediction results of support set and query set to obtain the label of unknown samples. The attention mechanism is introduced into the embedded feature extraction model to improve the few-shot classification method's fast learning ability. Snell J *et al.* [10] proposed the Prototypical Network, which calculates a prototype (the average value of sample feature vectors of each class) for each class sample in the support set. It then predicts the categories according to the distance between the samples and prototypes of each class. Sung F *et al.* [13] proposed Relation Networks. Unlike using pre-defined metrics, relation networks learn a non-linear expression and evaluate the relationship more accurately. Hui B *et al.* [14] proposed adding a self-attention module based on the Relation Network. The self-attention module finds out the correlation between each pixel and all other pixels so that the network can extract non-local long-distance dependency information. Simon C *et al.* [15] proposed finding a suitable subspace for each class. And then, metric the distance between samples in the subspace and predict the class.

Dictionary Learning

Dictionary learning is a representation of learning technology with a long history. It obtains the sparse representation of samples by the linear combination of dictionary atoms. Mallat S G *et al.* [16] first proposed that dictionary learning is a generic sparse representation model. Later, dictionary learning was applied to more fields, such as image classification [17] , image super-resolution reconstruction [18], image denoising [19], and image compression [19].

Mairal J *et al.* [20] proposed a supervised multi-modal dictionary learning method. It uses sparse joint constraints to enhance data of the same category to obtain a multi-modal dictionary with identification ability. Yang M *et al.* [21] introduced the Fischer criterion into sparse representation and proposed a discriminative dictionary method based on the Fischer criterion, which further improved the recognition ability of dictionary learning. Gu S *et al.* [22] introduced an analysis dictionary and proposed a Projective dictionary pair learning method. The analysis dictionary is trained through linear projection, which solves high complexity and low efficiency of sparse constraints of ℓ_0 norm and ℓ_1 norm. Aharon M*et al.* [23] proposed a dictionary learning method based on sparse representation, using iterative alternate learning to optimize the dictionary and better fit the sample data. Jiang Z *et al.* [24] introduced a sparse label matrix. Minimizing the discriminant sparse coding error to obtain a sample similar coefficient matrix improves dictionary learning's discrimination ability. Shao S *et al.* [25] proposed a dictionary learning method based on label embedding, which embeds tag information into the regularized dictionary learning method and transforms the sparse constraint problem into a convex optimization problem.

Methodology

In this section, we mainly introduce the Convolutional Shared Dictionary Module. First, we analyze the shortcomings of the feature extraction network used for few-shot learning and then explain the proposed CSDM in detail.

Convolutional Neural Network

Few-shot learning based on metric learning (such as Prototypical Network) consists of feature extraction network and base learner. The feature extraction network is usually a deep convolutional network (e.g., ResNet-12). We assume that $X = \{x_i|i = 1,2,\cdots,K\times N\}$ is the input samples and $M = K\times N$ is the total number of samples. The feature extraction network is considered as a function $F(\cdot)$, with which the feature embeddings are represented as $F(X) \in \mathbb{R}^{d\times N}$. Here, K represents the number of categories, N is the number of samples for each class, d is the dimension of feature embeddings, x_i is the i_{th} sample.

Each sample corresponds to an output (embedding feature) independently, and the embedding features are obtained separately. This lead to the feature extractor network cannot effectively use relationship among the samples. To find the structural information hidden in the feature embeddings and remove the redundant information, we propose a Convolutional Shared Dictionary Module (CSDM). The reconstructed feature embeddings via the CSDM would have more reliable discriminability.

Convolutional Shared Dictionary Module

As mentioned above, the relations among samples play crucial roles in extracting feature embeddings, and the convolutional neural networks usually ignore the relations. In this paper, we propose the Convolutional Shared Dictionary Module (CSDM) to address this problem. To be more specific, we try to employ a to-be-learned dictionary D to re-represent the feature embedding as X^*, where $D \in \mathbb{R}^{d\times n}$. Here, n is the number of dictionary atoms. This way builds a strong connection among classes, which is helpful for the downstream classification task. We formulate the CSDM as:

$$\{D,S\} = \arg\min_{D,S}\left\{\|F(X)-DS\|_F^2 + \alpha\|D\|_F^2 + \beta\|S\|_F^2\right\} \quad (1)$$

where $S \in \mathbb{R}^{n\times N}$ is the code for $F(X)$, α, and β are constant parameters. Then we optimize (1) to obtain the optimal of shared dictionary D and matrix S. We expand (1) as follows:

$$\begin{aligned}\{D,S\} = trace\{&(F(X)-DS)^T(F(X)-DS)\\ &+\alpha D^T D + \beta S^T S\}\end{aligned} \quad (2)$$

where $trace$ is the trace of matrix. The optimal shared dictionary D is as follows:

$$D = XS^T\left(SS^T+\alpha I\right)^{-1} \quad (3)$$

We propose to embed the shared dictionary D and the coding matrix S into the feature extraction network and update them during backpropagation and forward propagation, respectively. The optimal coding matrix S is as follows:

$$S = \left(D^T D + \beta I\right)^{-1} D^T F(X) \quad (4)$$

The coding matrix S contains inter-class information, we define the embedding features after reconstruction as follows:

$$X^* = \frac{\sum\left(S^T - mean(S^T)\right)}{n-1} \quad (5)$$

We summarize the steps in Algorithm 1.

Algorithm 1 Convolutional Shared Dictionary Module

Input: Samples $X = [x_1, x_1, \cdots, x_N]$ for k classes.
Output: Reconstructed embedding features X^*. Initialize CNN parameters, coding matrix S, shared dictionary matrix D.
Obtain embedding features $F(X)$.
Optimize shared dictionary matrix D and coding matrix S as described in Section .
Update dictionary D by (3).
Update coding S by (4).
Obtain reconstructed embedding features X^* by (5).

Baseline with CSDM

Figure 1 shows the architecture of the baseline with CSDM. We use the base learner to follow the Prototypical Network model and consider reconstructed embedding features of each class center in the support set as the prototype. Then use the query set to calculate the probability of each sample corresponding to each class.

EXPERIMENT

This section mainly shows and analyzes the experimental results on four benchmark few-shot image classification datasets and compares them with the baseline method (Prototypical Network). Besides that, we evaluate the performance of the proposed method CSDM under domain shift.

Datasets

We evaluate the proposed method on four benchmark few-shot image classification datasets: miniImageNet [12], tieredImageNet [9], CIFAR FS [9], and FC100 [9].

The miniImageNet dataset is a sub-dataset of the ILSVRC-2012 dataset [26], which contains 100 classes of the ILSVRC-2012 dataset, and each class includes 600 images. Among them, we randomly select 64, 16, and 20 classes for meta-training, meta-validation, and meta-testing, respectively.

The tieredImageNet dataset contains 608 classes of data selected from the ILSVRC-2012 dataset [26], and each category also includes 600 images. These classes are divided into 34 high-level classes, of which 20 (351), 6 (97), and 8 (160) high-level classes (classes) for meta-training, meta-validation, and meta-testing, respectively.

The CIFAR-FS dataset contains all 100 classes of data in the CIFAR-100 [27], of which 64, 16, and 20 classes for meta-training, meta-validation, and meta-testing, respectively. And each class contains 600 images.

The FC100 dataset also contains 100 classes in the CIFAR-100 [27], divided into 20 high-level classes, among which 12 (60), 4(20), and 4(20) high-level classes for meta-training, meta-validation, and meta-testing, respectively. The number of images is the same as the CIFAR-FS datasets. The number of images is 600. Both the tieredImageNet dataset and the FC100 dataset are more challenging to classify samples belonging to the same high-level class.

Experimental Setups

In the training stage, we use the meta-learning method to train our model. Each meta-task selects 5 classes of samples, and

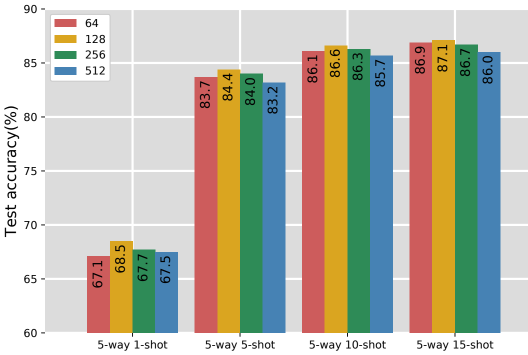


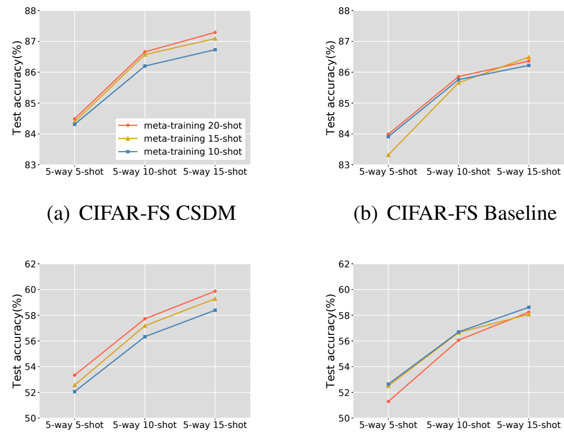
Figure 2. *The effect of the number of dictionary atoms on the model's performance. We conduct various experiments on the CIFAR-FS dataset, setting the number of dictionary atoms to* 64*,* 128*,* 256*, and* 512*, respectively. We train and test the performance of the model under the same environment. The figure shows experiments result on* 5*-way* N*-shot (*$N=1,5,10,15$*). From the figure, with the number of dictionary atoms increasing, the performance of the model gradually decreases after enhancement. When the dictionary atom is* 128*, the performance is the strongest, and we use this parameter in all experiments.*

the number of samples in each category is 15 (5-way 15-shot) on the four datasets. The meta-validation set is used only to adjust the parameters of the model and not for training.

We choose ResNet-12 as our feature extraction network, which is the same as [9], and use the SGD optimizer to train the model. The initial learning rate is set to 0.1 and gradually decreases as the number of epochs increases, and we also use label smooth. Selecting the number of atoms in the parameter dictionary of CSDM is shown in Figure 2. For a fair comparison, we train the baseline method and CSDM model under the same parameters. We randomly evaluate $1,000$ episodes with 95% confidence intervals on the meta-testing set as the final result by public code. We compare the proposed CSDM and baseline methods on four benchmark datasets, such as miniImageNet, tieredImageNet, CIFAR-FS, and FC100 datasets. We present the experimental results in Table 1.

Table 1. Classification results

Datasets	CSDM	1-shot	5-shot	10-shot	15-shot
miniImageNet		**59.25%**	75.60%	79.36%	80.99%
	✓	57.39%	**77.22%**	**80.90%**	**82.31%**
		-1.86%	*+1.62%*	*+1.54%*	*+1.32%*
tieredImageNet		61.74%	80.00%	83.57%	84.34%
	✓	**62.61%**	**82.49%**	**85.65%**	**86.78%**
		+0.87%	*+2.49%*	*+2.08%*	*+2.44%*
CIFAR-FS		**72.20%**	83.50%	85.66%	86.49%
	✓	70.33%	**84.39%**	**86.57%**	**87.09%**
		-1.87%	*+0.89%*	*+0.91%*	*+0.6%*
FC100		37.50%	52.50%	56.65%	58.07%
	✓	**38.43%**	**52.73%**	**57.18%**	**59.28%**
		+0.93%	*+0.23%*	*+0.53%*	*+0.71%*



(a) CIFAR-FS CSDM (b) CIFAR-FS Baseline

(c) FC100 CSDM (d) FC100 Baseline

Figure 3. *Evaluate accuracies on CIFAR-FS and FC100 datasets with the various meta-training shot. The CSDM performance on* N*-shot (N=*$5,10,15$*) gradually increases with increasing meta-training shot.*

Experiment Results

Table 1 shows the accuracy of the baseline method and the baseline method with CSDM and the accuracy improvement of these two methods. We can see that on the 5-way N-shot (N=$5,10,15$) meta-testing, the performance of CSDM has improved to varying degrees over the baseline method. On the 5-way 1-shot meta-testing, the performance on two fine-grained classification datasets, such as tieredImageNet and FC100, increases by 0.87 and 0.93, respectively. However, the experiments on the miniImageNet and CIFAR-FS datasets are not ideal. It may be because it is not stable to find the relationship among extremely few samples through dictionary learning. We speculate that the proposed CSDM is more suitable for situations with more meta-testing data, and we also conduct experiments to verify our speculation. We adopt the same parameters to train the baseline method, and the proposed CSDM obtains better experimental results on 5-way 10-shot and 15-shot. As shown in Table 1, the proposed CSDM has improved at least 0.5% in performance than the baseline method.

We compared our results with previous work on four datasets, as shown in Table 2 and Table 3, and our experimental results were competitive. We conduct more detailed experiments on CIFAR-FS and FC100 datasets to analyze the advantages of CSDM. As shown in Figure 3, the CSDM performance on 5-way N-shot (N=$5,10,15$) gradually increases with increasing meta-training shot, which is more practical. However, the baseline method does not have this trend. We may use a higher shot to train the model and obtain better performance. We analyze that with the increase of meta-training shot, the proposed method can mine more relationships among sample classes to reconstruct sample embeddings, which is more suitable for metric learning.

Feature Visualization with t-SNE

We use t-SNE [34] to visualize features on the CIFAR-FS dataset. As shown in Figure 4, the feature extractor used in the baseline method does not consider the relationship among samples, resulting in partial overlap of adjacent embedded feature

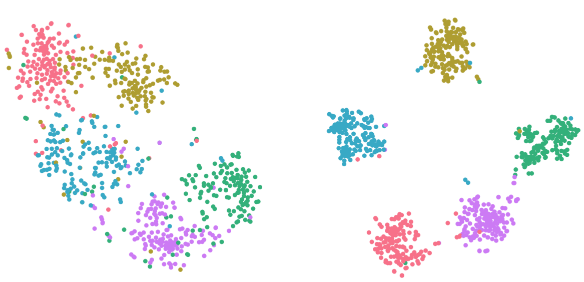
Table 2. **Comparison with previous to prior work on miniImageNet and tieredImagent dataset on 5-way 5-shot case.**

Model	Backbone	miniImageNet	tieredImageNet
MAML [8]	ConV-4	63.11 ± 0.92	70.30 ± 1.75
Relation Nets [13]	ConV-4	65.32 ± 0.70	71.32 ± 0.78
SNAIL [28]	ResNet-12	68.88 ± 0.92	-
AdaResNet [29]	ResNet-12	71.94 ± 0.57	-
TADAM [30]	ResNet-12	76.70 ± 0.30	-
Prototypical [10]	ResNet-12	75.60 ± 0.48	80.00 ± 0.55
TEAM [31]	ResNet-12	72.04	-
Meta-Transfer [32]	ResNet-12	75.53 ± 0.80	-
TapNet [33]	ResNet-12	76.36 ± 0.10	80.26 ± 0.12
SDL-FSL	ResNet-12	77.22 ± 0.71	82.49 ± 0.56

Table 3. Comparison with previous to prior work on CIFAR-FS and FC100 dataset on 5-way 5-shot case.

Model	Backbone	CIFAR-FS	FC100
MAML [8]	ConV-4	71.50 ± 1.00	-
Relation Nets [13]	ConV-4	69.30 ± 0.80	-
TADAM [30]	ResNet-12	-	56.10 ± 0.40
Prototypical Nets [10]	ResNet-12	83.50 ± 0.50	52.50 ± 0.60
SDL-FSL	ResNet-12	84.39 ± 0.5	52.73 ± 0.6

distribution and incorrect classification. The proposed CSDM method fully considered the relationship among samples and found concise patterns. The distribution of the same class is more concentrated, and the distance of different categories is more distant, making it easier to judge the category.



(a) Baseline (b) CSDM

Figure 4. The t-SNE visualization. We test the effect of the CSDM on the embedding features in a 5-way problem.

Performance Under Domain Shift

We train the model on the miniImageNet [12] dataset and evaluate performance on the CUB [35], plantae [36], and places [37] dataset under the same parameters. We follow the setting in [38]. Table 4 shows the accuracy of the baseline method and the CSDM and the accuracy improvement of these two methods. The proposed module improves performance 0.28% $\sim$ 4.63 % over baseline methods on 5-way N-shot ($N = 1, 5, 10$).

The CSDM we proposed can find the relationship among samples, reconstruct all the sample features, and reduce the dimension of the sample features that can remove some redundant information. We found that the performance of CSDM in cross-domain was improved higher than that in the few-shot learning dataset. It would be because the inter-class differences between the meta-training data and the meta-testing data in the few-shot learning dataset were not significant. Consequently, the improvement effect was limited. In contrast, the inter-class differences between the meta-training data in the miniImageNet and test data in the cross-domain experiment are more potent. Thus, the CSDM improves the discrimination of sample features more effectively by utilizing the relationship among samples.

Table 5. Classification results under domain shift

Datasets	CSDM	1-shot	5-shot	10-shot
		36.56%	50.65%	55.48%
mini. $\rightarrow$ CUB	✓	**36.96%**	**53.76%**	**59.92%**
		+0.40%	*+3.11%*	*+4.44%*
		29.57%	39.73%	43.81%
mini. $\rightarrow$ plantae	✓	**30.98%**	**43.19%**	**48.44%**
		+1.41%	*+3.46%*	*+4,63%*
		47.79%	65.34%	69.68%
mini. $\rightarrow$ places	✓	**48.07%**	**67.07%**	**71.72%**
		+0.28%	*+1.73%*	*+2.04%*

Conclusions

This paper proposes a plug-and-play module, Convolutional Shared Dictionary, inserted into the metric-learning-based few-shot learning method and reconstructed embedding features according to the relationship among samples. It helps the feature embeddings obtain the statistical relationship information and thus improves metric-learning-based few-shot image classification tasks.

References

[1] Qi Qi, Sanyuan Zhao, Jianbing Shen, and Kin-Man Lam, “Multi-scale capsule attention-based salient object detection with multi-crossed layer connections,” in *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2019, pp. 1762–1767.

[2] Shuai Shao, Lei Xing, Rui Xu, Weifeng Liu, Yan-Jiang Wang, and Bao-Di Liu, “Mdfm: Multi-decision fusing model for few-shot learning,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[3] Lei Xing, Shuai Shao, Yuteng Ma, Yanjiang Wang, Weifeng Liu, and Baodi Liu, “Learning to cooperate: Decision fusion method for few-shot remote sensing scene classification,” *IEEE Geoscience and Remote Sensing Letters*, 2022.

[4] Lei Xing, Shuai Shao, Weifeng Liu, Anxun Han, Xiangshuai Pan, and Bao-Di Liu, “Learning task-specific discriminative embeddings for few-shot image classification,” *Neurocomputing*, 2022.

[5] Shuai Shao, Lei Xing, Yan Wang, Rui Xu, Chunyan Zhao, Yanjiang Wang, and Baodi Liu, “Mhfc: Multi-head feature

collaboration for few-shot learning," in *ACM MM*, 2021, pp. 4193–4201.

[6] Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua, Alan L Yuille, and Li Fei-Fei, "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 82–92.

[7] Eli Schwartz, Leonid Karlinsky, J. Shtok, Sivan Harary, M. Marder, Abhishek Kumar, R. Feris, R. Giryes, and A. Bronstein, "Delta-encoder: an effective sample synthesis method for few-shot object recognition," in *Advances in Neural Information Processing Systems*, 2018.

[8] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *International Conference on Machine Learning*, 2017.

[9] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto, "Meta-learning with differentiable convex optimization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10657–10665.

[10] Jake Snell, Kevin Swersky, and Richard Zemel, "Prototypical networks for few-shot learning," in *Proceedings of the Advances in Neural Information Processing Systems*, 2017, pp. 4077–4087.

[11] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *International Conference on Machine Learning deep learning workshop*, 2015.

[12] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al., "Matching networks for one shot learning," in *Proceedings of the Advances in Neural Information Processing Systems*, 2016, pp. 3630–3638.

[13] Flood Sung, Yongxin Yang, Li Zhang, and Tao Xiang, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199–1208.

[14] Binyuan Hui, Pengfei Zhu, Qinghua Hu, and Qilong Wang, "Self-attention relation network for few-shot learning," in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops*, 2019, pp. 198–203.

[15] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi, "Adaptive subspaces for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4136–4145.

[16] Stéphane G Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, pp. 3397–3415, 1993.

[17] Shuai Shao, Lei Xing, Wei Yu, Rui Xu, Yan-Jiang Wang, and Bao-Di Liu, "Ssdl: Self-supervised dictionary learning," in *ICME*, 2021, pp. 1–6.

[18] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, pp. 2861–2873, 2010.

[19] S Grace Chang, Bin Yu, and Martin Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, pp. 1532–1546, 2000.

[20] Julien Mairal, Francis Bach, and Jean Ponce, "Task-driven dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 791–804, 2011.

[21] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, pp. 209–232, 2014.

[22] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, "Projective dictionary pair learning for pattern classification," in *Advances in Neural Information Processing Systems*, 2014, pp. 793–801.

[23] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, pp. 4311–4322, 2006.

[24] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2651–2664, 2013.

[25] Shuai Shao, Rui Xu, Weifeng Liu, Bao-Di Liu, and Yan-Jiang Wang, "Label embedded dictionary learning for image classification," *Neurocomputing*, pp. 122–131, 2020.

[26] 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*, pp. 211–252, 2015.

[27] Luca Bertinetto, Joao F Henriques, Philip HS Torr, and Andrea Vedaldi, "Meta-learning with differentiable closed-form solvers," *Proceedings of the International Conference on Learning Representations*, 2019.

[28] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel, "A simple neural attentive meta-learner," in *Proceedings of the International Conference on Learning Representations*, 2018.

[29] Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, and Adam Trischler, "Rapid adaptation with conditionally shifted neurons," in *International Conference on Machine Learning*, 2018, pp. 3664–3673.

[30] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste, "Tadam: Task dependent adaptive metric for improved few-shot learning," in *Proceedings of the Advances in Neural Information Processing Systems*, 2018, pp. 721–731.

[31] 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 International Conference on Computer Vision*, 2019, pp. 3603–3612.

[32] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele, "Meta-transfer learning for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 403–412.

[33] Sung Whan Yoon, Jun Seo, and Jaekyun Moon, "Tapnet: Neural network augmented with task-adaptive projection for few-shot learning," in *International Conference on Machine Learning*, 2019, pp. 7115–7123.

[34] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*,

pp. 2579–2605, 2008.

[35] Peter Welinder, Steve Branson, Takeshi Mita, Catherine Wah, Florian Schroff, Serge Belongie, and Pietro Perona, "Caltech-ucsd birds 200," *Computation & Neural Systems Technical Report*, 2010.

[36] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie, "The inaturalist species classification and detection dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8769–8778.

[37] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1452–1464, 2017.

[38] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang, "Cross-domain few-shot classification via learned feature-wise transformation," in *Proceedings of the International Conference on Learning Representations*, 2020.

Author Biography

Tong Zhou *obtained a Bachelor's degree in Petroleum Engineering from China University of Petroleum (East China) in 2001 and a Master's degree in Oil and Gas Field Development Engineering from China University of Petroleum (East China) in 2007. He is currently pursuing a Ph.D. in Petroleum and Natural Gas Engineering at China University of Petroleum (East China). His research focuses on efficient oil and gas production theory and technology, as well as big data analysis and applications in oil and gas production.*

Changyin Dong *is a professor and Ph.D. supervisor. He graduated with a Bachelor's degree in Petroleum Engineering from China University of Petroleum (East China) in 1998 and a Ph.D. in Oil and Gas Field Development Engineering from China University of Petroleum (East China) in 2003. His research and teaching work focuses on oil and gas extraction engineering, rock mechanics in petroleum engineering, and solid production (sand production) in oil and gas reservoirs and its control.*

Zhen Wang *graduated with a Master's degree in Oil and Gas Field Development Engineering from China University of Petroleum (East China) in 2014. He has been engaged in intelligent oilfield technology research, pilot construction, and technical management work at Shengli Oilfield for many years. His main research areas include machine learning model development and optimization for oil and gas exploration and development, as well as artificial intelligence platform and application architecture design.*

Bo Chang *graduated with a Bachelor's degree in Applied Chemistry from Qingdao University in 1996. He has been working in the production IoT technology department at Sinopec Shengli Oilfield Branch, focusing on research related to oil and gas production IoT data acquisition, transmission, and automatic control technologies.*

Junshu Song *obtained a Bachelor's degree in Engineering from Shandong University of Technology in 2010. He has been working in the information department of Sinopec Shengli Oilfield Branch, mainly engaged in the research and development of oil and gas production information software and hardware products, as well as the development and promotion of intelligent technologies.*

Shen Bin *is Engineering Director (AI/ML) at Celonis. He was with the Advanced Technology Group (ATG) of Pinterest from 2019 to 2022. Before that, he was with Google New York from 2015 to 2019. Bin earned B.S. and M.S. degrees from Department of Electronic Engineering, Tsinghua University, Beijing, in 2007 and 2009, respectively, and M.S. and Ph.D. degrees from Department of Computer Science, Purdue University, West Lafayette, Indiana, in 2011 and 2014.*

Baodi Liu *received the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2013. He is currently an Associate Professor with the College of Control Science and Engineering, China University of Petroleum, Qingdao, China. His research interests include computer vision and machine learning.*

Pengfei He *graduated with a Master's degree in Information and Computing Technology from Saint Petersburg State Electrotechnical University, Russia, in September 2011. He is currently the Deputy Manager of the Petroleum Development Information Services Center at Shengli Oilfield. Since joining the workforce, he has won several awards, including First Prize at the Fourth and Fifth China Petroleum and Petrochemical Intelligent Technology Exchange Conference, First Prize for Outstanding Achievements from the Shandong Petroleum Association, Second Prize in Quality Management Achievements in Shandong Province, and Second Prize in the Shengli Oilfield Informationization Intelligent Innovation Competition.*