A metric learning network based on attention mechanism for Power grid defect identification

Xiaodie Yu¹, Xiaoming Ju¹*, Yong Wang², Haojin Qi²

¹School Computer Science and Software Engineering, East China Normal University, Shanghai, 200062, China
²Ningbo Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Ningbo, 315000, China

*Corresponding author’s e-mail: xmjuxi@sei.ecnu.edu.cn

Abstract: Aiming at the current situation that deep learning has a strong dependence on sufficient samples and it is difficult to obtain sufficient labeled samples in practical applications, a few-shot image classification method combining attention mechanism is proposed. Firstly, the feature extraction module adds an attention mechanism to increase the weight of important parts, pay attention to the more important parts of the image, and obtain the corresponding feature vector. Secondly, the relevance measurement module is used to calculate the similarity between features to determine the category of the query image. Experiments have proved that our proposed model improves the accuracy of about 1.2% compared with the existing model. At the same time, the model is applied to the classification of power grid defects.

1. Introduction

At present, deep learning based on deep convolutional networks has made great achievements in the fields of image recognition, target segmentation, target detection, and speech recognition. And it has been gradually applied in various fields, including smart cities, smart homes, smart grids and so on. However, deep convolutional networks have three huge challenges: Firstly, deep learning relies on a large number of labeled samples, however, due to privacy, security, or the contingency of events, there are often scenarios with insufficient samples in actual application scenarios. In addition, collecting sufficient samples requires a lot of manpower and material resources. Finally, the generalization ability of new tasks is insufficient, and the training model needs to be rebuilt when receiving new tasks.

The limitations of deep learning samples derive the direction of few-shot learning. Its core is to learn the potential associations or internal laws of things from limited samples, and to imitate and learn the ability of humans to obtain knowledge from a few data instances through analogy and promotion. The main challenge of few-shot learning lies in the insufficient number of samples. Based on this, the method of expanding sample data is the first consideration, using data enhancement methods: based on scaling and transformation [1]. GAN generated data [2] and other expansion of the number of samples, So as to solve the problem of insufficient samples; secondly, metric learning and meta-learning are also two mainstream methods to solve few-shot learning.

The security of the power grid is related to the stability of the country. Power grid need regular inspection and maintenance because there are many influencing factors: the influence of natural factors such as wind and rain; secondly, the influence of external factors such as man-made damage; thirdly, the long time period leads to quality problem. However, as far as the current situation is concerned, line
maintenance generally requires manpower to inspect and find, or to troubleshoot after problems, which leads to maintenance that requires a lot of manpower, material resources, and time costs. In addition, there is passivity and lag in repairing problems after they appear.

This requires proactive discovery of defects and timely maintenance. The current common method is to use monitoring images to make human judgments. This method also has two drawbacks: one is that everyone has different judgment standards, and the results are often subjective. The second is the high cost of manpower identification. How to integrate artificial intelligence to solve problems is the current development trend. Power grid defects are the category of few-shot learning. The number of samples is insufficient, and there are only a few samples for each type of defect.

Our paper is based on the metric learning method. The current feature mapping network for metric learning is generally a simple CNN network with a small number of layers. Compared with complex images, the feature learning ability has limitations. Based on this problem, our paper combines the attention mechanism to improve the feature mapping network, and then uses the matching degree calculation module to achieve few-shot classification. At the same time, the trained model is applied to the problem of power grid defect identification and classification to solve the problem of defect classification.

2. Related work

2.1. Metric learning
The basic idea of metric learning is to model the distance distribution between samples, so that similar targets are closer and different objects are farther away. In 2015, [3] proposed Siamese Network that uses the principle of metric learning to solve few-shot learning. The network mines the similarity measurement relationship from known data, and then uses the learned relationship to determine the category of unclassified samples. Matching Network [4] combines the attention mechanism to read a small number of samples each time during training, so that the test set and the training set environment are as consistent as possible. In 2017, Prototype Network [5] mapped the samples in each category to the metric space, and the classification problem was transformed into the problem of finding the nearest neighbors in the metric space. In order to solve the above problem, in 2018 [6] Sung et al. proposed Relation Network to learn the distance measurement between samples. Our method is based on metric learning, and the fusion attention mechanism improves the feature mapping network to realize the problem of few-shot learning.

2.2. Attention mechanism
The attention mechanism was first proposed in machine translation, and has been gradually applied in various fields, including: speech recognition, image classification and so on. In a sense, the attention mechanism belongs to human-like thinking. The visual system or memory system tends to pay attention to the key parts of the object it sees, thereby ignoring the unimportant information points. This is the principle of the attention mechanism, which makes the machine highlight the decisive part in the learning process and weaken the irrelevant important features [7].

SENet [8] proposed an attention mechanism for the channel domain to learn the correlation between channels. By controlling the weight, the important part is strengthened and the unimportant part is weakened. In the same year, the CBAM model [9] focused on the characteristics of the channel dimension and the spatial dimension; the basic idea of SKNet [10] proposed in 2019 was to obtain convolution kernels with different weights for different images.

2.3. Meta-learning
Meta-learning [11] is also called "learning to learn". Its purpose is not to solve a specific scenario or problem through training samples. The basic principle trains a meta-learner on a large number of tasks firstly, so that it learns how to acquire knowledge from a limited sample, and has strong generalization ability. In the face of new tasks, it can achieve rapid adaptation. In 2017, Model-Agnostic Meta-Learning
(MAML) [12] was proposed. The model obtained the initialization parameters of the model through training. When adapting across tasks, it can quickly adapt to new tasks with fewer iterations. In the same year, [13] proposed the Meta-SGD model, which learned the learning rate at the same time of learning initialization.

3. The proposed Method

The current feature mapping network for metric learning is generally a simple 4-layer convolutional network. For complex background images, the ability to extract features is limited. Based on this, this paper introduces an attention mechanism to improve the feature mapping module in metric learning to achieve few-shot classification.

The overall architecture of the model is shown in Figure 1.

![Figure 1. Structure diagram of the Attention-Relation Network](image)

The model contains three modules: Feature extract module based on the attention mechanism, Feature connecting module and Matching degree calculation module.

**Feature extract module based on the attention mechanism.** The attention-based feature mapping module mainly implements feature extraction. The network structure of the model is shown in Figure 2. It includes four convolution modules (convolution layer, BN layer and Relu activation function). The MaxPooling layer is added after the first two convolution modules; the Attention module is followed to learn the relationship and correlation between Channels to obtain different Channels Weights, and output feature vectors with different weights.

N categories are randomly selected in the data set, and K samples of each category are randomly selected to form a support set:

$$I_{support} = \{(x_i, y_i)\}_{i=1}^{K \times N}$$

From the selected N categories, choose C samples from each category to form a query set:

$$I_{query} = \{(x_j, y_j)\}_{j=1}^{C \times N}$$

The support set and one sample of the query set are input to the Attention-based feature extraction network to obtain $F(x_i)|_{i=1}^{K \times N}$ and $F(x_j)$. 
**Feature connecting module.** The purpose of the feature connecting module is to connect the features of the support set $F(x_j)_{i=1}^{K*N}$ and the features of the query set $F(x_j)$.

$$F_i,j(x)_{i=1}^{K*N} = C(F(x_i)_{i=1}^{K*N}, F(x_j))$$

**Matching calculation module.** The matching calculation module learns the matching measurement method of Relation Network [6]. The specific network model structure is shown in Figure 3. It uses two convolution modules (convolution layer, BN layer, Relu layer and MaxPooling layer). It is followed by two fully connected layers to calculate the similarity between the query set features and the support set samples to determine the category information of the query image.

$$r_{i,j} = M(F_i,j(x)_{i=1}^{K*N})$$

---

### 4. Experiment

#### 4.1 Dataset

**MiniImagenet.** We select miniImagenet [14] as the dataset, including 100 categories, each category contains 600 samples. Select 64 categories as the training set, 16 categories as the verification set, and the remaining 20 categories as the test set.

**Power grid defect dataset.** We screened 5 types of sample data from power grid monitoring images, including: self-explosion of insulators, skew of phase plates, foreign objects in tower poles (bird's nest), shock-proof hammer, wire stripping, and identification of power grid defects as a new task.
4.2 Experimental configuration

This experiment is based on the Linux system and is implemented using the PyTorch deep learning framework. The specific experimental configuration is shown in Table 1.

| Configuration          | version  |
|------------------------|----------|
| Cuda                   | 9.0      |
| operating system       | Linux    |
| Graphics card          | Tesla K80|
| Pytorch                | 0.3.1    |

4.3 Experimental steps

**Training phase.** Training the meta-learner in an n-way k-shot task mode by using training set. Firstly, construct a task: N categories are randomly selected in the MiniImagenet training set, and K samples of each category are randomly selected to form a support set. From the selected N categories, each category selects C samples to form the query set. Select a sample in the query set, and input the built model together with the samples in the support set to perform feature extraction, connecting, and matching calculation to determine the category of the query image. Select another picture in the query set and repeat the above operation.

**Validation & testing phase.** The validation set and test set are not intersected with the categories of the training set, that is, the category labels are different. In the same way as the training set construction task, follow the N-way K-shot method to verify & test the trained model.

**Power grid defect dataset.** On the grid defect dataset, the task is still constructed using the 5-way-1-shot method. One picture is selected from each of the 5 categories of defects to form a support set, and all the remaining samples form a query set. Only a few samples are needed to realize the identification of 5 categories of defects in the power grid.

4.4 Experimental results

**MiniImagenet dataset.** The experiment in this article is based on the miniImagenet dataset. The model is trained on this data set and compared with the currently widely used method model.

| Model                   | Fune-tune | 1-shot     | 5-way Accuracy          |
|-------------------------|-----------|------------|-------------------------|
|                         |           | 1-shot     | 5-shot                  |
| Matching Network        | N         | 43.56±0.84%| 55.31±0.73%             |
| Meta-learning LSTM      | N         | 43.44±0.77%| 60.60±0.71%             |
| Prototypical network    | N         | 49.42±0.78%| 68.20±0.66%             |
| MAML                    | Y         | 48.70±1.84%| 63.11±0.92%             |
| Relation Network        | N         | 49.40±0.50%| 64.55±0.40%             |
| **Our**                 | N         | **50.59±0.70%** | **64.70±0.5%**          |

Note: the best results are shown in bold in the table.

As shown in Table 2, compared with other models, our method achieves the best accuracy results in 5-way 1-shot. Compared with Relation Network, the accuracy is improved by 1.2% approximately.

**Power grid defect dataset.** We conduct experiments on the constructed grid defect dataset to realize grid defect detection. The example in the figure below is (a) picture of an insulator defect, where a is the original image and (b-h) is the output of the middle layer of the model.
There are 5 categories of defects in the grid defect dataset, and each category of defect has 20 samples. In the experiment, the sample classification accuracy rate is 60%.

| Model | Dataset                          | 5-way 1-shot Accuracy |
|-------|----------------------------------|-----------------------|
| Our   | Power grid defect dataset        | 60%                   |

5. Conclusion

The paper proposes a metric learning model based on the attention mechanism, which includes three modules: a feature extraction module based on the attention mechanism, a feature connection module and a matching degree calculation module. Experiments have proved that on the 5-way 1-shot few-shot learning problem, our proposed model improves the accuracy of about 1.2% compared with the existing model.

Secondly, the proposed model is applied to the actual application of power grid defect detection, using several pictures to realize the detection of five categories of grid defects.

At present, due to the limitations of the sample, the accuracy of most methods has not reached a very high level. There will be a lot of research for few-shot learning in the future: how to use a small number of samples to achieve accurate detection.

References

[1] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features[C]//Proceedings of 2017 IEEE Conference on International Computer Vision. Venice, Italy: IEEE, 2017: 3037-3046. [DOI: 10.1109/ICCV.2017.328]
[2] Charalambous, C.C., Bharath, A.A.: A data augmentation methodology for training machine/deep learning gait recognition algorithms. In: BMVC. (2016)
[3] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition[C]. ICML Deep Learning Workshop. 2015, 2
[4] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]. Advances in Neural Information Processing Systems. 2016:3630-3638.
[5] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[C]. Advances in Neural Information Processing Systems. 2017:4080-4090.
[6] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: CVPR. (2018)
[7] K. Xu et al, "Show, attend and tell: Neural image caption generation with visual attention," in 2015.
[8] J. Hu et al, "Squeeze-and-Excitation Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, (8), pp. 2011-2023, 2020; 2019;.
[9] S. Woo et al, "CBAM: Convolutional block attention module," in 2018. DOI: 10.1007/978-3-030-01234-2_1.
[10] L. Xiang et al,” Selective Kernel Networks, cvpr2019
[11] Munkhdalai T,Yu H. Meta networks[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70,2017:2554-2563
[12] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[J].arXiv preprint arXiv:1703.03400, 2017
[13] Li, Z., Zhou, F., Chen, F., Li, H.: Meta-sgd: Learning to learn quickly for few shot learning. In: arxiv:1707.09835. (2017)
[14] Ravi S, Larochelle H. Optimization as a model for few-shot learning[C]//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR, 2017