Application of Meta Learning in Face Recognition

Sisi Peng *, Kaining Zheng
Information system engineering college, PLA Strategic Support Force Information Engineering University, 93 high-tech Zone, Zhengzhou 450000, China

*1335303371@qq.com

Abstract. Now is the information age, artificial intelligence rises under the background of this era, but in the field of artificial intelligence, deep learning stands out in machine learning with its excellent learning ability and is rapidly applied to face recognition engineering. However, the success of mainstream deep learning often depends on a lot of training data and training time. This paper proposes to use meta-learning technology to realize face recognition, meta-learning is to use the prior knowledge and experience to guide the learning of new tasks and has the ability to learn so that it can avoid the phenomenon of overfitting in the case of only a small number of samples. We regard the data set as a task, that is, meta-train task and meta-test task, using the MAML algorithm [1] fine-tuning gradient model based on prior knowledge to obtain optimization parameters to realize face recognition. Experiments show that deep learning neural networks using meta-learning technology can achieve higher accuracy and rate than ordinary face recognition neural networks. In the experiment, the recognition rate of the face can reach more than 92.6%. It has more intelligence and has made a contribution to the development of face recognition technology.

Keywords: Meta-learning, Few-shot learning, Face recognition, Secondary gradient

1. Introduction
At present, the artificial intelligence technology began to rise, the company enterprise, the home life to face recognition demand is also increasing, using the increasingly perfect deep learning technology, gradually meet the needs of face recognition on various occasions, such as bank and privacy-related software login verification, easy to work and learn face recognition card equipment, these gradually spread throughout every corner of life. However, the mainstream deep learning technology needs to rely on a large number of training data and training time to train a good deep model, and cannot use prior knowledge, its generalization ability is weak. This makes the performance of the face recognition algorithm difficult to realize human intelligence.

In this paper, using the recently emerging meta-learning and small sample learning techniques, and using prior knowledge to adjust and optimize network parameters, the problem of less sample learning can be solved. It can avoid overfitting and achieve faster learning. Compared with traditional deep learning, it can obtain higher accuracy and rate. The Model-Agnostic Meta-Learning algorithm is used to learn a good initialization weight. By using hyperparametric gradients (double gradients) [2] descent, the network learns useful representations from the full distribution of tasks, thus has fast adaptation on new tasks, rapid convergence, and fine-tuning on small-scale samples.
This paper analyzes the application of deep learning technology in face recognition and finds that the existing face recognition algorithms still have the disadvantages of time-consuming and laborious. Hence, the study of introducing meta-learning and small sample [3] learning into face recognition VGG net model is established. In the research, the experiment realizes the classification with high accuracy under a few samples, and the learning speed is faster than the mainstream neural network algorithm, which improves the ability of the model and optimizes the performance of the face recognition algorithm.

2. MAML algorithm and small sample processing in face recognition

MAML algorithm, by initializing the network parameters, the neural network can obtain stronger learning ability, MAML focus on improving the overall learning ability of the model, rather than the ability to solve a specific problem. However, the mainstream face recognition algorithm focuses on solving the current task and starts learning again in the face of the new task, which leads to the weak generalization ability in face recognition. In order to solve the above problems, the MAML algorithm is used to make the model switch on different tasks during training, so as to initialize the network parameters. The final model is based on prior knowledge training. You can learn faster when facing new tasks.

This article works with a university's pattern recognition and artificial intelligence lab, the data set in the face recognition database is selected as the test object, set up the parameters according to the 6-way 10-shot experiment[4], and select a total of 400 labeled samples from 10 categories $C_1$ to $C_{10}$ for training meta-models $M_{meta}$, a total of 200 annotated samples were selected from 10 categories $P_1$ to $P_{10}$ to get the model $M_{fine-tune}$. Randomly selected 6 categories in $C_1$ to $C_{10}$, each category randomly selects 30 tagged samples to form a task $\Gamma$. Among them, 10 samples were support set, 20 remaining samples were query set. Using the support set and query set in each task $\Gamma$ to calculate the gradient of each parameter, update the parameters of the model. MAML algorithm performs a secondary gradient update, acting directly on the original model through SGD.

The optimization process [5] of the face recognition model is based on the MAML algorithm. In the MAML algorithm, $\phi$ are initial parameters for the network, and $\tilde{\theta}^n$ are unique network parameters obtained from the task n based on a specific learning rate[6]. The loss function is as follows:

$$L(\phi) = \sum_{n=1}^{N} l^n(\tilde{\theta}^n)$$ (1)

The loss value is calculated by using cross entropy as follows:

$$\frac{\partial C}{\partial w^l} = \frac{1}{n} \sum_{i} \alpha_l^{-1}(a_l^i - y)$$ (2)

$$\frac{\partial C}{\partial f^j} = \frac{1}{n} \sum_{i} (a_l^i - y_i)$$ (3)

Unlike the traditional pre-training model, the MAML algorithm carries out two gradient update[7], does not pursue the optimal loss value on the current model, and focuses on the stronger generalization ability in the face of new tests. The first gradient update is based on the parameters of the random initialization model, and the new parameters are calculated using the samples in the support set in the task. The second gradient update is based on the sum of the query set calculated losses of all the tasks that make up a batch to calculate the parameters SGD stochastic gradient descent, which enhances the generalization ability of the model and avoids overfitting the previous support set.

The specific parameter update process is shown in Fig.1 below:
Application and result analysis of face recognition

To solve the problem of face image classification and recognition, a face image recognition method based on the MAML algorithm and deep learning network VGG is proposed. First, the labeled samples in the training set are divided into meta-training domain and meta-test domain. Second, for model optimization, the quadratic gradient updated model is used to obtain the fine-tuning model $M_{fine-tune}$. Third, the MAML algorithm is used to classify the face images of the new category. In order to verify the effectiveness of the algorithm, the algorithm is tested by the face image database.

In this paper, a few face images are selected as training samples in cooperation with a pattern recognition artificial intelligence laboratory. In view of the more scenes applied to face recognition, the images are all positive face images. There are 6 categories in the total category. According to the requirement of 6-way 10-knot, the experimental setup is divided into face 1~6. The experimental results are shown in Fig.2. Experimental results show that the recognition rate of all images can reach more than 92.6%, and the loss value is about 3~4. This method has good recognition accuracy on a small number of samples. Although the loss value is larger than that of the mainstream network algorithm, it has a better generalization ability for the new samples. Therefore, it has a certain application value.

| Face category | Losses  | (%) accuracy |
|---------------|---------|-------------|
| Face 1        | 3.7866  | 98.5        |
| Face 2        | 3.5743  | 97.4        |
| Face 3        | 4.2343  | 92.6        |
| Face 4        | 2.9876  | 98.3        |
| Face 5        | 3.0234  | 94.7        |
| Face 6        | 3.2875  | 93.2        |

Fig.2 Experimental Results of Face Recognition

3. Face recognition based on MAML algorithm and small sample learning

Unlike the traditional face recognition algorithm, the dataset only needs a small number of tagged samples. Each task is divided into a support set and a query set. In the presence of a few tagged samples, each task can contain some of the same samples[8].

In the supervised task area, few-sample learning is well studied, using previous data of similar tasks in meta-learning, only learning a new function from several input/output pairs of that task. For example, the target can classify images after seeing only one or more examples of the test set, because the model has seen many other types of objects before. In the presence of such prior knowledge, faster classification can be achieved.
Secondary gradient update model. After using the support set data to calculate the gradient to train the model, the model is re-trained using the query set on the basis of the model, which makes the model have better adaptability, while in the training process, it does not pursue the minimum loss value of the current model, but the generalization ability of the model to unknown samples.

To verify the performance based on the MAML face recognition algorithm, the same two training samples and test samples were selected to test the MAML model as well as the VGG network model which began to undergo parameter pre-training and the random initialization parameters. The experimental results are shown in Fig.1. Analysis in Fig.3 shows that the accuracy of the MAML method is higher than that of the pre-trained VGG network face recognition method and higher than that of the parameter randomly initialized VGG network. Moreover, the loss value of the proposed method is higher than that of the VGG network of the unused MAML algorithm, the results are shown in Fig.4.

![Fig. 3 Recognition accuracy of different algorithms](image)

![Fig.4 Loss values under different support vector algorithms](image)
4. Improved Face Recognition Strategy
We have increased depth and small convolution kernels. The basic network adopts VGG net, in order to deepen the network layer number and avoid too many parameters, a small convolution kernel of 3 x 3 or 1 x 1 is used in all layers, and the convolution layer step size is set to 1. 1x1 convolution kernel reduces the dimension and increases the nonlinearity; 3 x 3 convolution kernel makes multiple convolution kernels superimpose, increases spatial sense field and reduces the use of parameters.

We have the basic idea of a secondary gradient. The main difference between the MAML algorithm and the mainstream algorithm is that the number set is divided into multiple tasks, each task is divided into the training domain and test domain. The optimization of partial loss value is sacrificed to improve the adaptability of the model to unknown targets. At first, the classification of the model is based on prior knowledge, using the network parameters fully learned, thus speeding up the convergence rate of the network supervision [9] process in classification, so as to complete face recognition faster.

5. Conclusion
In the research of the face image recognition method, this paper takes meta-learning and small sample technology as the main research line, mainly based on the MAML algorithm, through different data testing, combined with meta-model, to improve the learning ability of the model [10]. Ability to generate knowledge, make the model more intelligent, and gradually expand and apply. At the end of this study, the system performance is tested. In the test, some experiments are carried out, including the influence of sample number and algorithm category on recognition accuracy. Through the analysis of the test data, it can be seen that the number of training samples has a certain impact on the recognition accuracy, and the more samples the better the recognition performance. In the model using the MAML algorithm, the effect of a small sample number can be eliminated, and the model also has good performance. This study has achieved ideal results and provided technical support for the application of the color image recognition method.

Funding
This work was supported by the National Natural Science Foundation of China (No.61673395, and No.61403415).

References
[1] Chelsea F, Pieter A and Sergey L 2017 Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks Proc. Int. Conf. on Machine Learning 70 pp1126-1135
[2] Marcin A, Misha D, Sergio G, Matthew W.Hoffman, David Pfau, Tom Schaul, Brendan Shillingford and Nando de Freitas 2017 Learning to learn by gradient descent by gradient descent, Proc. Int. Conf. on Machine Learning 70 pp748-756
[3] Huaxiu Yao, Chuxu Zhang, Ying Wei, Meng Jiang and Zhenhui Li 2020 Graph Few-shot Learning via Knowledge Transfer, Proc. Int. Conf. on Artificial Intelligence 34(04) pp6656–6663
[4] Yunxiao Qin, Chenxu Zhao, Xiangyu Zhu, Zezheng Wang and Zhen Lei 2020 Learning Meta Model for Zeroand Few-shot FaceAnti-spoofing, Proc. Int. Conf. on Artificial Intelligence 34(07) pp11916–11923
[5] Jianzhu Guo, Xiangyu Zhu, Chenxu Zhao and Dong Cao 2020 Stan Z. Li, Learning Meta Face Recognition in Unseen Domains, IEEE/CVF Conf. on Computer Vision and Pattern Recognition
[6] Antreas Antoniou, Harrison Edwards and Amos Storkey 2019 How to train your MAML Int. Conf. on Learning Representations
[7] Harkirat Singh Behl, Mohammad Najafi, Anurag Arnab and Philip H.S.Torr Meta Learning Deep Visual Words for Fast Video Object Segmentation, arXiv 1812 01397
[8] Soh Jae Woong, Cho Sunwoo and Cho Nam Ik Meta 2020 Transfer Learning for Zero-Shot Super-Resolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition*

[9] Kyle Hsu, Sergey Levine and Chelsea Finn 2019 Unsupervised Learning via Meta-Learning, *Int. Conf. on Learning Representations*

[10] Harkirat Singh Behl, Atılım Gunes Baydin and Philip H.S. Torr 2019 Alpha MAML Adaptive Model-Agnostic Meta-Learning, *Proc Int Conf on Machine Learning*