Journal of Informatics and Web Engineering
Vol. 2 No. 2 (September 2023) eISSN: 2821-370X

A Multi-Scale Feature Attention Image Recognition Algorithm

Xin MingYuan[1*], Ang Ling Weay[2] , Sellappan Palaniappan[3]
[1,2,3] School of Information Technology, Malaysia University of Science & Technology, Petaling Jaya, Malaysia
**corresponding author: (xin.mingyuan@phd.must.edu.my; ORCiD: 0000-0001-9336-5612)*

ABSTRACT- The success of image classification using small samples is contingent on neural network models' capability to derive image representations from the data. A proposed solution is a small-sample image classification system that leverages attention mechanisms and meta-learning to capture more comprehensive image information. Due to its ability to efficiently suppress irrelevant characteristics and accentuate pertinent ones, this technique may extract more robust multiscale features and enhance classification performance through meta-learning.In this paper, the effectiveness of the multi-scale attention network is verified on two datasets, namely, Mini-ImageNet and Tiered-ImageNet, and the accuracy of the method is 58.54% for 5-way 1shot and 74.76% for 5-way 5shot on the Mini-ImageNet dataset. In the dataset of the Tiered-ImageNet,the accuracy of 5-way 1-shot and 5-way 5-shot increased to 59.74% and 78.65%, respectively. The experimental results show that the multi-scale sub-attention can pay more attention to the global information of the image than the single-scale attention network, and significantly improve the accuracy of small-sample image classification.

Keywords— Small-Sample Classification, Meta-learning, Attention mechanism, Multi-scale, Feature extraction

Received: 17 December 2022; Accepted: 29 March 2023; Published: 16 September 2023

I. INTRODUCTION

Data-driven deep learning models have had considerable success with picture categorization problems since the advent of big data [1]. In general, three key components—adequate processing power, complex neural networks, and huge datasets—are what make deep learning successful. There are not enough training samples available in many real-world application scenarios, including those in the sectors of medicine, the military, and finance, due to worries about privacy and security as well as high human expenses. Training models are susceptible to overfitting in the absence of sufficient supervised instances, which results in the model performing well on the training sample but failing to generalise effectively on the test set [2]. Small-sample learning [3] uses a much smaller sample size of data than that required for deep learning to achieve results that approach or even surpass those of deep learning with large data. To improve the generalisation performance of models with limited sample learning, several experiments have been conducted in China and worldwide.

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2023.2.2.1

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

Although deep learning has proved quite effective at categorising photos, it requires a lot of annotated data to work effectively. Data collecting may be challenging in industries like security and healthcare, and data labelling and purification are labor-intensive processes. Deep neural networks may also overfit as a result of the lack of labelled data, which would reduce their capacity to classify data. Small sample learning has grown to be a common problem in machine learning due to the fact that it can be difficult or expensive to acquire samples in the real world [4] , although data augmentation techniques based on GAN networks can generate training data and increase training data [5].

By enabling network models to quickly abstract representative features from a small number of samples and to quickly compare key information about an image when they encounter similar tasks, small-sample learning solves the aforementioned issue and allows for the classification of a new class without having to retrain the model. FSL aims to create a model that can recognise novel positions from a limited sample size [6]. This necessitates setting up the network parameters for each data collection, which slows down network optimization. Small-sample learning, in particular, is a subset of the image classification problem where the deep network is first pre-trained on a large number of samples with comparable tasks to enable the model to continuously learn the common knowledge of the images, and then the trained model is applied to optimise the current small-sample task. In contrast to conventional deep learning, small-sample learning concentrates on common features among images, making it more adaptable to new classification tasks. It also produces better classification results for sample data with a smaller number of categories without the need to train on large labelled samples. For Few-shot sample [7].To overcome these issues, transfer learning [8] has become a more popular approach. The goal of transfer learning is to learn new information fast and apply it to a different domain.

Using the aim of making deep networks learn similarly to people, the idea of few-shot learning with few samples was suggested to address this problem. However, due to the constrained amount of parameter changes during training, gradient-based optimization methods struggle with few-shot learning [9]. An strategy that is frequently employed to solve this issue is transfer learning. It entails quickly transferring knowledge to a new domain while building on prior knowledge. Knowledge transfer between many disciplines is now feasible thanks to transfer learning. Data augmentation [10], meta-learning [11], and metric learning [12] are popular solutions in small sample domains.

A generator network and a discriminator network make up GANs [13]. Since the basic GAN model is an adversarial game problem between the generator and the discriminator, with the generator's purpose being to produce data with a distribution that is as near to the real data as possible, the objective function has a direct impact on the basic GAN model. The model learns commonalities from a few marked-up samples before the task in order to solve few-shot training challenges better [14]. Every job is made up of a support set, which typically only contains a limited amount of data for model tuning and is the source of the tiny sample, and a query set, which is used to test the model's generalizability on the task. The model will discover a set of sensitive parameters through training on several distinct tasks that may be rapidly modified to learn a new task [15].

In order to solve the problem of a finite amount of training samples, we suggest the use of multi-scale Attention Feature (MAF) for the expansion convolution-based generation of feature maps of different sizes. In this study, the relational network embedding module's channel attention method is implemented to help the model better identify tiny samples by learning larger multi-scale features and rescaling them.

The following are the paper's contributions:

- To overcome the limitations posed by limited training data in image recognition and to prevent overfitting, we present the Multi-scale Feature Framework to enhance sample diversity and combine multiscale features for improved recognition.
- Our proposed MAF Framework features an attention mechanism for both multi-scale space and channels, which enables the relational network's embedding module to learn more comprehensive multi-scale features and adjust the channel attention weights in multiple dimensions, leading to enhanced classification performance with small samples.
- To illustrate the intricacy of the multi-scale attention technique proposed in this study, experiments are planned on the open-source Mini-ImageNet and Tiered-ImageNet datasets.The research is conducted using two distinct small grouping tasks, 5-way 1-shot and 5-way 5-shot, respectively. To show the usefulness and development of

the algorithm suggested in this work, more sophisticated algorithms like MAML,ADM and L2F are also chosen as baseline.

II. METHODOLOGY

A. Multi-scale feature learning

The Multi-scale Feature Learning method is used to improve the picture recognition classification algorithm's accuracy. Using various preprocessing methods or convolution kernels of varying sizes, this method generates images with varying scales, merges the resulting image features for analysis. In the measurement based meta-learning approach, the classifiers of Siamese, coordinating, and model organizations are planned physically utilizing Euclidean or cosine removes, the social organization further develops the distance metric, and the brain network is utilized as the classifier to gain proficiency with the between highlight metric. The first goal of the metric-based meta-learning is to learn an embeddable module with predefined fixed metrics. Then, the relational network is used to learn a transferable depth metric module to improve the accuracy with which images can be classified. The object's semantic and contextual information in the depth feature map is enhanced by the multi-scale receptive field.Studies have been conducted on the impact of multi-scale features on image recognition, with positive results reported in the literature [16].

B. Attention Mechanism

The Attention Mechanism was proposed to enhance the feature extraction ability of neural networks by enabling them to focus on the most important parts of an image [17]. This idea was inspired by how humans naturally focus on specific areas in visual scenes. The channel attention mechanism, introduced in the SENet [18] network, focuses on the internal dependencies between different channels in an image. By assigning different attention weights to each channel, it creates a channel attention weight vector to enrich the global information of the image. On the other hand, the position attention mechanism creates an attention weight vector by weighting and aggregating features at each position in the image. This mechanism emphasizes the distribution of image features by focusing on the position information in the image

C.Multi-scale Attention Feature (MAF)

The Multi-scale Feature (MAF) architecture is designed to address the overfitting issue in image recognition caused by small training samples. Since the current bottleneck in the development of data networks is still the computing power of computers, when the input information keeps increasing, the model will be extended to be more complex, and the amount of information to be processed can be reduced by introducing the attention mechanism. Machine learning models typically include the attention mechanism, a particular structure used to automatically determine and compute the contribution of the input data to the output data.The influence of the features on the lower-level model or network can also make use of more correlation information from the original data.The multi-scale classification network is convolved and the multi-scale feature map features are extracted at the beginning of each network using an attention technique. This results in the output, which is represented by r={r1,r2,r3,...,rK}. The network can swiftly focus on the crucial areas of an image while still taking the complete image into account by using different weights that are determined by the attention module. The network can more accurately and efficiently extract pertinent data from the image by mixing multi-scale characteristics for recognition, which enhances the performance of image classification.

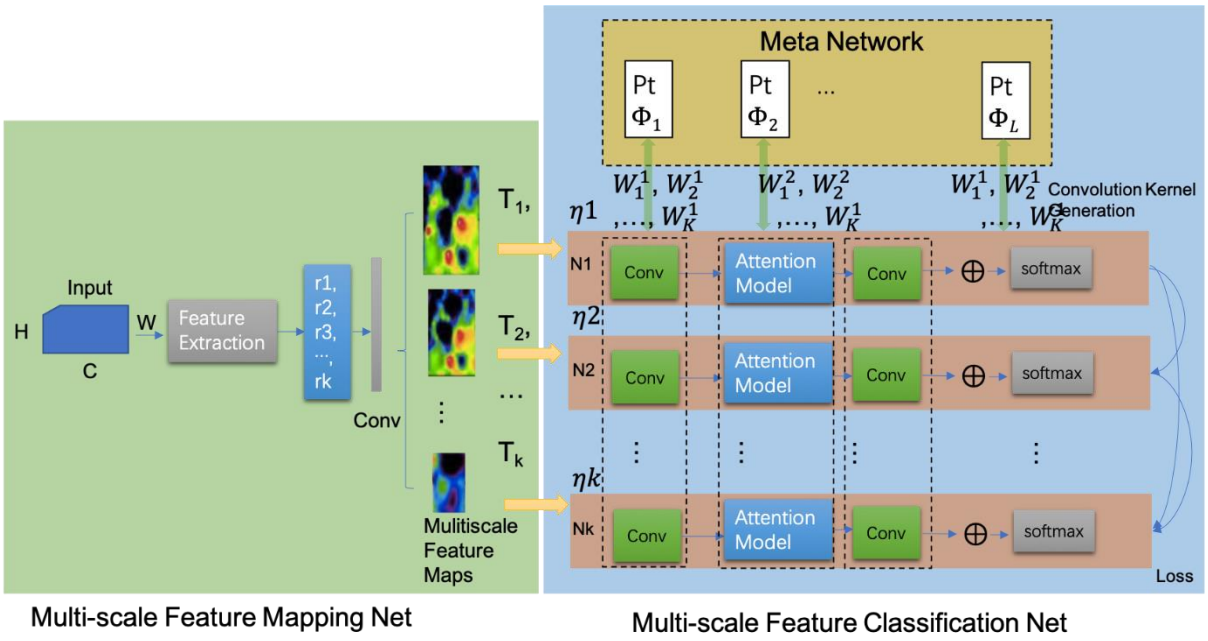

Figure 1. Multi-scale Feature Flowchart

The multiscale feature maps obtained using different expansion coefficients are converted into multiscale feature maps of the input in the channel direction. Considering that the location detail information of high-dimensional image features is gradually lost, which is not conducive to image classification, the local features used do not allow the network to focus on the image location information. In order to build a rich contextual relationship model on the local features of the image, and considering that the attention may be focused on the unimportant object featurees, we add a location attention module, which can integrate the similar information of the image from a global perspective adaptively and focus on the location information of the image as a reference basis for classification.

First, create a multiscale feature map in the channel direction from the input.In order to process these components independently and concurrently with various scales, the group convolution method is introduced.Dealing with multi-scale kernel and group size relationships. where K represents the kernel size and G the group size.The obtained multi-scale feature maps are stitched together.

$$F_i = conv(k_i \times k_i, G_i)(X_i), i = 1,2, \dots, S-1$$

Second, for each transformation, t Translation of the input X to the feature map U, where U∈RH ×W ×C. The Squeeze procedure first reduces the spatial dimension of the feature U to an 1×1×C vector, also known as global average pooling. The following equation illustrates the squeezing operation.

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j)$$

The high-dimensional features of an image can be regarded as the response of a class.Each new channel feature vector is an expression of a local area feature, and we construct a channel global subnotice module to model the image channel information. From the perspective of channel information, we focus on the channel dependencies between images as another discriminative method to solve the problem.

Thirdly, Softmax is used to recalibrate the channel direction of the attention vector and obtain the attention weight of the multi-scale channel recalibration.To supply the attention weights required for the multiscale channel calibration, the attention vectors for the channel directions are recalibrated using Softmax.

$$att_i = Softmax(Z_i) = \frac{\exp(Z_i)}{\sum_{i=1}^{S-1} \exp(Z_i)}$$

Fourth, in particular, Softmax is used to calibrate the multi-scale channels,in where att stands for the multiscale channel weights following an attention cascade.The related feature maps, F_i are weighted using the derived multiscale channels' weights.

$$Y_i = F_i \odot att_i, i = 1,2, \dots, S-1$$

where $\odot$ denotes channel multiplication.

Finally, a fine-grained feature model with richer multi-scale feature information is obtained with the query set image feature cascade and input relationship model. As the output of the feature map.

$$out = cat([Y_0, Y_1, \dots, Y_{s-1}])$$

III. EXPERIMENT AND ANALYSIS

The comparison of the proposed MAF method and the most recent state-of-the-art small sample learning methods on Mini-ImageNet and Tiered-ImageNet yielded the results, which are presented in Table 1 in the appropriate order.This model can be used more effectively for the classification task of small samples because the experimental results demonstrate that it performs better on the classification task than other current methods.

A. Dataset

The MiniImageNet dataset is a benchmark dataset in the field of meta-learning and small samples, extracted and developed by the Google DeepMind team. on the basis of ImageNet, a complex dataset containing 60,000 color images in 100 categories, with 600 samples per category, each of 84 × 84 size. 84, suitable for prototype design and experimental research There are three sections to the Tiered-ImageNet dataset, which contains 608 classes and 1,281 pictures per class: 351 for the training, 97 for the validation, and 160 for the testing of small sample learning. Each image in the Mini-ImageNet and Tiered-ImageNet databases has a dimension of 84 by 84 pixels.

B. Experimental environment

All experiments were tested in the Ubuntu18 environment.Due to the characteristics of deep learning, the performance of the model depends heavily on the design of the network structure and the initialization of the parameters. To be fair, Conv64 was chosen as the feature extractor. The standard meta-learning approach is followed during the training and testing phases, i.e., the data are strictly in the form of N-way k-shot for each task in training and testing. The goal of the small-sample classification task is to determine which image in the query set and the support set belong to the same class. As in other papers, the accuracy of the model is measured in the form of 5-way 1 shot and 5-way 5 shots. In the training process, Adam is used as the optimizer, and the learning rate is set to 0.001, and the learning rate is halved every 10,000 training sessions. The parameters are initialized in the normal way, and the rest of the parameters are used in the default way. Each new task randomly selects 5 categories, and each category has only 1 sample of training data, then randomly selects 15 images from each category as the support set, and the total 75 support sets form a 5-way 1-shot task. The 5-way 5-shot and 5-way 1-shot tasks are performed to determine which of the 5 categories the 75 images belong to. Similar to the 5 way 1 shot, the 5 way 5 shot uses 600 tasks randomly chosen from the test dataset and uses the average accuracy of top1 as the model accuracy for that time. The average of the 5 times is repeated 5 times, and the final model accuracy is recorded as the final model accuracy.

C. Experimental results and analysis

The performance of the MAF model is compared with other state-of-the-art small-sample classification methods using the Mini-ImageNet dataset. MAF models are compared with different approaches to evaluate their performance in five-direction one-shot and five-direction five-shot tasks. The results show that the MAF model performs well on a variety of tasks and helps identify very small samples.

Table 1. Task Classification Precision on Dataset

Method	Mini-ImageNet		Tiered-ImageNet	
	5way-1shot	5way-1shot	5way-1shot	5way-5shot
MAML [19]	48.71±0.60	63.10 ±0.92	49.21±0.81	66.22 ±0.47
L2F [16]	52.14±0.51	69.34±0.47	54.48±0.58	73.34±0.44
ADM [20]	56.71±0.66	72.54±0.52	56.11±0.69	75.19±0.56
MAF (the proposed method)	58.54±0.63	74.76±0.48	59.74±0.29	78.65±0.51

On the Mini-ImageNet dataset, the proposed MAF network performs better than the small sample classification methods currently in use. In comparison to MAML, the accuracy in the 5-way 1-shot task is nearly 10% higher. Accuracy is 11.66 percent greater than with MAML on the 5-way, 5-shot job. The effectiveness of the model proposed in this research may be utilised to confirm its robustness because Mini-ImageNet is a multispecies dataset.

On the Tiered-ImageNet dataset, the MAF network beat rival methods for small sample classification issues. The MAF network outperformed MAML in accuracy on both the 5-way 1-shot test and the 5-way 5-shot job by 10.53 percent and 12.43 percent, respectively. The MAF network's multi-scale attention mechanism is credited with the improved performance. Overfitting is reduced throughout the training phase by utilising channel attention and expansion coefficients, which enables the network to gather more thorough feature data from diverse angles. The result is a model that generalises well as shown by its strong representational capabilities across many datasets. Additionally, the MAF network's ability to extract data allows for a more comprehensive and accurate representation of the data.

D. Ablation experiments

In this section, the ablation of the proposed multi-angle feature module and attentional association classifier under different data volume scenarios is studied experimentally. In addition, this section also introduces the baseline model MAML to add attention, so as to verify the effectiveness of adding attention and further improve the performance of the model under the condition of limited samples.

Table 2. MAF Under Single Attention Network

Method	Accuracy	
	5way-1shot	5way-5shot
MAML with Attention	51.23±0.60	69.36±0.52
MAML	49.21±0.81	66.22 ±0.47
Only Attention	52.88±0.62	69.76±0.50
MAF without Attention	48.13±0.60	67.76±0.52
MAF	58.54±0.63	74.76±0.48

The results of the ablation experiments under 5-way 1-shot and 5-way 5-shot conditions are shown in Table 2. It can be seen that after adding the multi-scale feature extraction and introducing the attention module proposed in this paper, the recognition accuracy is improved by about 10% on the 5-way 1-shot task and 7% on the 5-way 5-shot task.

IV.CONCLUSION

In order to solve the problem of low readiness for image recognition under a small sample training dilemma, a multi-scale attentional feature network is proposed in this paper. The multi-scale feature mapping network generates multi-scale

features by convolutional operations with different unfolding coefficients, which enhances data diversity and overcomes the limitation of small sample size. A multi-scale feature mapping network and a multi-scale feature classification network make up the majority of the MAF network. The multi-scale feature classification network of this network uses attention weights to extract features from the multi-scale feature map for prediction, and each sub-network has a scalar related to the size of the feature mapping and contains multiple multiscale classification subnetworks, and the knowledge present in the multiscale classification network is used to improve the network performance by calculating the KL scatter between the softmax outputs of each subnetwork. The experimental results show that the MAF network outperforms all the baseline networks in this experiment, and it can be anticipated that the network has great efficiency and potential for generalization in various fields.

ACKNOWLEDGEMENT

The authors would like to thank the two anonymous reviewers who have provided valuable suggestions to improve the article.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", International Conference on Neural Information Processing Systems, pp. 1097, 2012. https://doi.org/10.1145/3065386

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, pp. 211-252, 2015. https://doi.org/10.1007/s11263-015-0816-y

[3] Y. Wang, Q. Yao, J. T. Kwok, L. M. Ni, "Generalizing from a Few Examples: A Survey on Few-Shot Learning", ACM Computing Survey, vol. 53, no. 3, pp 1–34, 2020. https://doi.org/10.1145/3386252

[4] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, T. Darrell, "Few-Shot Object Detection via Feature Reweighting", IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8419-8428, 2019. https://doi.org/10.48550/arXiv.1812.01866

[5] M. Y. Xin, L. W. Ang, S. Palaniappan, "A Data Augmented Method for Plant Disease Leaf Image Recognition based on Enhanced GAN Model Network," Journal of Informatics and Web Engineering, vol. 2, no. 1, pp 1-12, 2023. https://doi.org/10.33093/jiwe.2023.2.1.1

[6] Z. Shen, Z. Liu, J. Li, Y. G. Jiang, Y. Chen, X. Xue, "DSOD: Learning Deeply Supervised Object Detectors from Scratch", IEEE International Conference on Computer Vision (ICCV), pp. 1937-1945, 2017. https://doi.org/10.1109/ICCV.2017.212.

[7] K. Saito, Y. Ushiku, T. Harada, K. Saenko, "Strong-Weak Distribution Alignment for Adaptive Object Detection", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6949-6958, 2019. https://doi.org/10.1109/CVPR.2019.00712.

[8] Y. Z. Liu, K. M. Shi, Z. X. Li, G. F. Ding, Y. S. Zou, "Transfer Learning Method for Bearing Fault Diagnosis Based on Fully Convolutional Conditional Wasserstein Adversarial Networks", Measurement, vol. 180, pp. 109553, 2021. https://doi.org/10.1016/j.measurement.2021.109553

[9] Z. Chen, Y. Fu, K. Chen, Y. G. Jiang, "Image Block Augmentation for One-Shot Learning," Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, pp. 3379–3386, 2019. https://doi.org/10.1609/aaai.v33i01.33013379

[10] Q. Lyu, D. Xia, Y. Liu, X. Yang, R. Li, "Pyramidal Convolution Attention Generative Adversarial Network with Data Augmentation for Image Denoising", Soft Computing, vol. 25, pp. 9273–9284, 2021. https://doi.org/10.1007/s00500-021-05870-7

[11] S. Lavania and P. S. Matey, "Novel Method for Weed Classification in Maize Field Using OTSU and PCA Implementation", IEEE International Conference on Computational Intelligence & Communication Technology, pp. 534-537, 2015. https://doi.org/10.1109/CICT.2015.71

[12] H. Altae-Tran, B. Ramsundar, A. S. Pappu, V. Pande, "Low Data Drug Discovery with One-Shot Learning", ACS Central Science, vol. 3, no. 4, pp. 283–293, 2017. https://doi.org/10.1021/acscentsci.6b00367

[13] G. Daras, A. Odena, H. Zhang, A. G. Dimakis, "Your Local GAN: Designing Two Dimensional Local Attention Mechanisms for Generative Models", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14519-14527, 2020.

[14] B. Liu, Z. Ding, L. Tian, D. He, S. Li, H. Wang, "Grape Leaf Disease Identification Using Improved Deep Convolutional Neural Networks", Frontier in Plant Science, vol. 11, 2020. https://doi.org/10.3389/fpls.2020.01082

[15] Y. Xiao, X. Huang, K. Liu, "Model Transferability from ImageNet to Lithography Hotspot Detection," Journal of Electronic Testing, vol. 37, no.1, pp. 141–149, 2021.

[16] S. Baik, S. Hong, K. M. Lee, "Learning to Forget for Meta-Learning", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2376-2384, 2020.

[17] H. Fukui, T. Hirakawa, T. Yamashita, H. Fujiyoshi, "Attention Branch Network : Learning of Attention Mechanism for Visual Explanation", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10697-10706, 2019. https://doi.org/10.48550/arXiv.1812.10025

[18] J. Hu, L. Shen, G. Sun, "Squeeze-and-Excitation Networks", IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132-7141, 2018. https://doi.org/10.1109/CVPR.2018.00745.

[19] C. Finn, P. Abbeel, S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks", International Conference on Machine Learning, vol. 70, pp. 1126–1135, 2017. https://doi.org/10.48550/arXiv.1703.03400

[20] W. Li, L. Wang, J. Huo, Y. Shi, Y. Gao, J. Luo, "Asymmetric Distribution Measure for Few-Shot Learning", Twenty-Ninth International Joint Conference on Artificial Intelligence, pp. 2957–2963, 2020. https://doi.org/10.48550/arXiv.2002.00153