Few-Shot Learning for Medical Image Classification

Aihua Cai, Wenxin Hu(✉), and Jun Zheng

East China Normal University, Shanghai, China
cah0231@163.com, {wxhu,jzheng}@cc.ecnu.edu.cn

Abstract. Rapid and accurate classification of medical images plays an important role in medical diagnosis. Nowadays, for medical image classification, there are some methods based on machine learning, deep learning and transfer learning. However, these methods may be time-consuming and not suitable for small datasets. Based on these limitations, we propose a novel method which combines few-shot learning method and attention mechanism. Our method takes end-to-end learning to solve the problem of artificial feature extraction in machine learning and few-shot learning method is especially to fulfill small datasets tasks, which means it performs better than traditional deep learning. In addition, our method can make full use of spatial and channel information which enhances the representation of models. Furthermore, we adopt 1×1 convolution to enhance the interactions of cross channel information. Then we apply the model to the medical dataset Brain Tumor and compare it with the transfer learning method and Dual Path Network. Our method achieves an accuracy of 92.44%, which is better than the above methods.

Keywords: Medical image classification · Few-shot learning · Attention mechanism · Transfer learning

1 Introduction

The classification of medical images such as tumor types is important for the diagnosis and subsequent treatment of diseases. However, classifying medical images with similar structures manually is a difficult and challenging task that requires a lot of time for experienced experts. In order to improve the efficiency and accuracy of classification, researchers propose plenty of methods, such as machine learning [8], deep learning [14] and transfer learning [7]. However, these methods have some shortcomings. As the most important step in machine learning, feature extraction requires experts to spend much time determining the features. Deep learning is more suitable for huge datasets, which means the small amount and unbalanced categories problems will limit the efficiency of deep models. Transfer learning can use a pre-trained model to address the problems of small datasets, while there is a great difference between natural and medical images. So we need to explore new methods to solve these problems.

I. Farkaš et al. (Eds.): ICANN 2020, LNCS 12396, pp. 441–452, 2020.
https://doi.org/10.1007/978-3-030-61609-0_35

The emergence of few-shot learning provides new directions for medical image classification tasks. The few-shot learning [17,18] is proposed to solve the problem of overfitting and aims to recognize novel categories from very few labeled examples. To this day, it has also produced a variety of effective models and algorithms. The main methods are meta-learning [12], metric-based [20], data-augmented [5], semantic-based [19], and so on. The baseline used in the paper is Prototypical Network [17] which is based on metric learning. This model is outstanding in the classification of small samples, but there is still the problem that spatial and channel information is not considered in feature extraction. Due to the noise in the medical image, the features that contribute to the classification results should be found more accurately.

Based on this deficiency, we improve the original model. In the embedding module, we extract features through several convolution operations. After extracting many features, we consider "what" is meaningful in the image and "where" is an informative part of the classification task. So we add an attention module into the embedding module. Inspired by *Network in Network* [13], we add 1×1 convolution into the convolutional blocks to enhance the interactions of cross channel information. In this way, we can improve feature representations and suppress more useless information. The use of few-shot learning methods can effectively address the problem of overfitting. We conduct a series of comparative experiments to validate the effectiveness of our methods. At last, the results show that compared with pre-trained models, our model is more effective.

The main contributions of this paper can be summarized as follows:

1. We propose to add the attention mechanism into the network, which helps model extract features from spatial and channels simultaneously. Through the experiments, we find different placements of attention mechanism share different results. When the Convolutional Block Attention Module(CBAM) is put between the last two convolutional blocks, the result is the best.
2. We propose to add 1×1 convolution into the convolutional blocks. This operation can enhance the interactions of cross channel information and the results outperform the prior one.
3. We apply few-shot learning methods to the field of medical image classification to solve the problem of small datasets and achieve good results.

The rest of the paper is organized as follows: we first introduce the related work about methods to solve medical image classification in Sect. 2. Then we describe our methods in detail in Sect. 3. In Sect. 4, we mainly present our datasets, experimental setup and results. Finally, in Sect. 5, we conclude our work and indicate future directions.

2 Related Work

2.1 Method Based on Machine Learning

In the medical field, doctors need to judge the type of tumors according to CT or MRI images [8]. However, it costs much time and energy of doctors and

experts that they need to identify the location of tumors, compare the location and shape, make more accurate judgment and conclusion. In the beginning, researchers adopt machine learning methods, which involve data preprocessing, image segmentation, feature extraction, selection and classification. Before feature extraction, experts are required to select a series of features for calculation, which can be gray value, texture features, etc. Due to too many features, feature selection is also required. From these processes, we can see that these features for classification are very important and these processes have great flexibility and complexity, so it may lead to the result is not ideal, unstable and has weak generation. These methods cannot apply to several datasets well.

2.2 Method Based on Deep Learning

Unlike machine learning, deep learning does not require manual feature extraction. The image can be directly classified without image segmentation because most methods are end-to-end which greatly save time and energy. Deep learning is also widely used in the medical field, from image segmentation to recognition to classification. [11] classified diabetic retinopathy images by combining Inception-v3, ResNet152 [9] and Inception-ResNet-v2, which achieved great results. [23] proposed the siamese-like structure and used two pictures as input to learn their correlation to help classify. [3] proposed an improved capsule network to classify brain tumor types. Through a large number of experiments, deep learning has been proved to be effective in processing medical images. However, at present, most deep learning methods extract features through convolution operation. Convolution operation can extract information such as edges and textures of images, but spatial and channel information cannot be used well.

2.3 Method Based on Transfer Learning

There are still some problems in the application of deep learning to the medical field, mainly due to the small medical datasets. Since the datasets are small, it is probable to be overfitting when applying deep learning models. In view of the small amount of data, researchers proposed to use transfer learning methods. Most models are trained on big datasets such as ImageNet and use medical datasets to fine-tune some layers of the model so that the model can better adapt to medical datasets [21]. However, there are also some papers question whether the effect of transferring the pre-trained model learned on ImageNet to the medical dataset is good and experiments have proved that the use of a smaller and simpler model can achieve comparable results as the use of pre-trained models.

2.4 Method Based on Few-Shot Learning

Few-shot learning [15] is also applied to fulfill the task of medical image classification. [14] proposed an AffinityNet, which used the k-Nearest-Neighbor attention pooling layer to extract abstract local features. This model based on semi-supervised few-shot learning shows great performance on disease type prediction.

[6] used Siamese Network, one of the classical network in few-shot learning, to retrieve images in medical datasets. [16] applied Triplet Network, which was improved by Siamese Network, to accomplish brain imaging modality recognition. So far, few cases use few-shot learning to solve the problem of medical image classification and the method above mainly uses CNN to extract features which can not find discriminative features to help classify.

3 Methodology

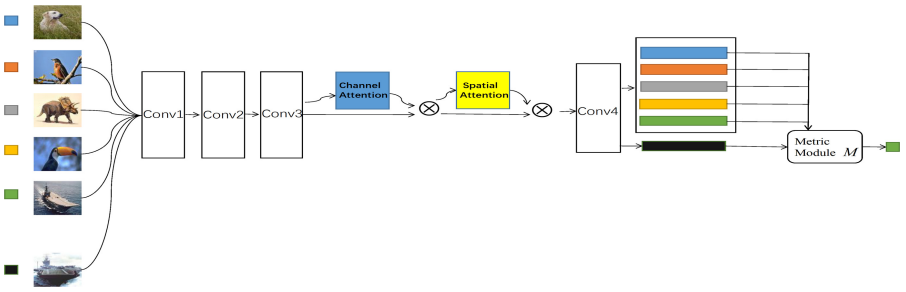


Fig. 1. The overview architecture of our method.

3.1 Overview

In this paper, we propose a novel method to solve the medical image classification task. As the original convolutional block can only extract the basic information and it focuses on the whole image, while the features of medical images are usually much noisier and heterogeneous, we choose Prototypical Network and improve it with the attention mechanism. With this improvement, our model can pay more attention to the part that contributes to the classification. The specific structure is demonstrated in Fig. 1. When the sample is mapped to the feature space, we mainly use an embedding function $f(x)$, which is a neural network, and it is composed of four convolutional blocks as shown in Fig. 2. Since the convolution operation can only extract partial edge, texture, direction and other information, we use the attention mechanism to increase the representation ability of the model. We add spatial and channel attention module between the third and fourth convolutional blocks respectively to extract the spatial and channel information of the image. Furthermore, we adopt 1×1 convolution kernels in the convolutional block. This convolution can greatly increase the nonlinear characteristics and deepen the network without losing the resolution. Next, we will introduce the basic model.

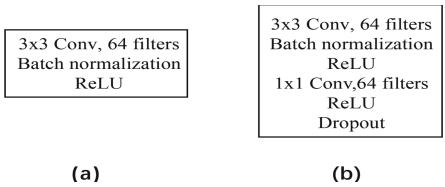


Fig. 2. Two types of convolutional blocks used in the experiment. (a) is the classical one which is widely used in few-shot learning models. (b) is the modified convolutional blocks, which is added 1×1 convolution kernel and dropout.

3.2 Prototypical Network

Prototypical Network is a classical model in few-shot learning. It was proposed by [17] in 2017, and was mainly used to solve the problem of overfitting in few-shot datasets. It projects the sample into the feature space, where the homogeneous samples are closer together and the heterogeneous samples are farther apart. In the Prototypical Network, the prototype is obtained by calculating the mean value of the features from the same class. In the subsequent testing process, the distance between the test sample and each prototype is calculated to see which prototype is close to it. The class is decided by which class the prototype belonging to. As in Fig. 1, we present a 5-way 1-shot condition. The input is 5 images from 5 classes, so we do not calculate the mean of every embedding value. When we use 5-way 5-shot, we need to calculate the mean of embedding values as a prototype. Our method is based on the Prototypical Network and then we will introduce the attention mechanism.

3.3 Attention Mechanism

CBAM was proposed by [22] in 2018. The main purpose is to increase the representation ability of models through the attention mechanism. It considers the information in channels and spatial. In the channel attention module, it can generate a channel attention map and it mainly mines "what" is meaningful in the input image. In the spatial attention module, it uses the information on the relationship in feature space to explore "where" is important. Especially in medical fields, doctors determine the type of tumor mainly from the position, size, or color in medical images. Therefore, it is very important to explore the spatial and channel information in medical images.

3.4 Network Architecture

The model based on metric learning includes two components, one is the embedding module and the other is the classifier. The image x_i from support set and the image x_j from query set are fed into the embedding module respectively and obtain their feature maps $f(x_i)$ and $f(x_j)$. Then the feature map will be input

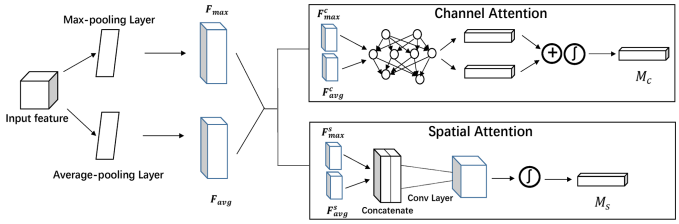


Fig. 3. CBAM module

into the classifier to calculate the distance between them. The classifier in Prototypical Network is Euclidean distance. In prior works, four convolutional blocks are utilized for embedding modules, which can be seen in Fig. 2(a). Every block contains 64 filters 3×3 convolution, batch normalization and a ReLU nonlinearity layer. We also conduct experiments with the modified convolutional blocks (Fig. 2(b)). It differs in that we add 1×1 convolution and dropout. By adding these two operations we can further enhance the representation of models and the effect of dropout is to reduce the number of intermediate features and reduce redundancy. Between convolutional blocks, there is a 2×2 max-pooling layer. The CBAM module is shown in Fig. 3. According to the conclusions in [22], we choose to place the module in a sequential arrangement and put the spatial attention behind the channel attention. The two modules are both use max-pooling and average-pooling. In channel attention module, max-pooling and average-pooling are first adopted to generate respect features F^c_{max} and F^c_{avg} and then put them into a shared network. After that, we merge the output feature vectors through element-wise summation. Finally, the sigmoid function is used to calculate the channel attention M_c in the following formula:

$$M_c = \sigma(MLP(F^c_{avg}) + MLP(F^c_{max})) \tag{1}$$

where σ denotes the sigmoid function, MLP is a shared network with a hidden layer.

In spatial attention module, after max-pooling and average-pooling, we do concatenate operation to generate feature descriptors. Then we apply a convolution layer to generate a spatial attention map M_s. The whole process can be described by the following formula:

$$M_s = \sigma(Conv([F^s_{avg}; F^s_{max}])) \tag{2}$$

where σ denotes the sigmoid function, $Conv$ represents a convolution operation, $[;]$ means the concatenation operation.

4 Experiments

4.1 Datasets and Preprocessing

The brain tumor dataset [1] consists of three types: meningioma, glioma and pituitary tumor, which is shown in Fig. 4. The number of these three brain

tumor images is 708, 1426 and 930, which is quite different, so we need to do data augmentation. We rotate these images at 90, 180 and 270 degrees. We also add Gaussian noise and Salt-Pepper noise to images to enhance the robustness of the model. Finally, the amount of data can reach 2832, 2852, 2790, which achieves the class balance generally. All images are 512×512.

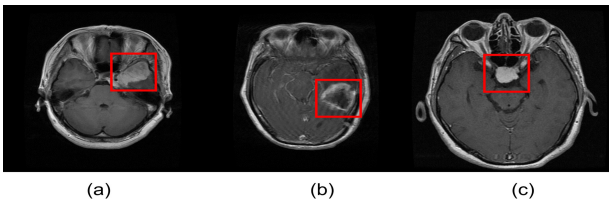


Fig. 4. (a) Meningioma, (b) Glioma, (c) Pituitary tumor. Tumors are localized inside a red rectangle.

4.2 Experimental Settings and Training Details

In the following experiments, we adopt Adam as the optimizer for model training and the initial learning rate is 10^{-3}. The number of training and testing episodes per epoch is 100. The batch size is 8. The number of epochs to train is 100. To alleviate overfitting, we also adopt dropout and the value is 0.5. The dimensionality of hidden layers is 64.

In the few-shot learning, episode training strategy is widely used. We use 5-way 20-shot with 20 query images for each class in the training episode. Firstly, we sample 5 classes in the training set and then sample 20 images from these 5 classes. The 20 query image is selected from the rest images of the 5 classes. Finally, the $5 \times 20 + 5 \times 20 = 200$ images compose a training episode. The training and testing conditions are the same. In our experiments, there are only three classes, so we have to use prior knowledge to train the model. We use other medical datasets [2] to augment the class of training set, while the dataset is not very similar to the brain tumor dataset, then we combine 1500 images each class in the brain tumor dataset to help fine-tune the model. We considered that if we involve the images of novel classes at the beginning of the training, it would make the model more suitable to solve the existing classification problem.

4.3 The Effect of Adding CBAM Module

In Sect. 3, we discuss the CBAM module can exploit the information from channel and spatial to strengthen the representation ability. In this section, we conduct several experiments to validate this idea. Some network architectures can be seen in Table 1. P1 is the network that uses four convolutional blocks (Fig. 2(a)) and P2 is the network that we add a CBAM module between the third and fourth

Table 1. The network architectures used in the experiment

Name	P1	P2	P3	P4
Input	Brain Tumor Dataset: 512×512 gray-scale images			
Network architecture	conv3-64	conv3-64	conv3-64 conv1-128	conv3-64 conv1-128
	maxpool: 2×2			
	conv3-64	conv3-64	conv3-64 conv1-128	conv3-64 conv1-128
	maxpool: 2×2			
	conv3-64	conv3-64	conv3-64 conv1-128	conv3-64 conv1-128
	conv3-64	CBAM	CBAM	conv3-64 conv1-128
		conv3-64	conv3-64 conv1-128	
	Flatten			
Results	83.27%	87.35%	**92.44%**	89.48%

convolutional blocks. From their results, we can see that P2 is better than P1 by an increased accuracy of 4%.

To further investigate the effect of CBAM, we try to add the CBAM module between different convolutional blocks. Such as P2, CBAM is added between the third and fourth blocks. We also try to put it in other places. Through the experiment, we find that different locations and numbers of CBAM modules generate different results. The results are shown in Table 2. In the experiment, it is found that 4 convolutional blocks are more effective than 3 or 5 convolutional blocks. Placing the CBAM module between the third and fourth convolutional blocks (Number 2 in Table 2) is better than others. Using several CBAM modules (Number 5 in Table 2) is no better than using a single module.

Table 2. Comparison with different placements of CBAM modules

Number	Method	Accuracy
1	3_CNN_12_CBAM	62.24%
2	4_CNN_34_CBAM	**87.35**%
3	4_CNN_23_CBAM	77.06%
4	4_CNN_12_CBAM	70.40%
5	4_CNN_2CBAM	79.26%
6	5_CNN_45_CBAM	68.73%

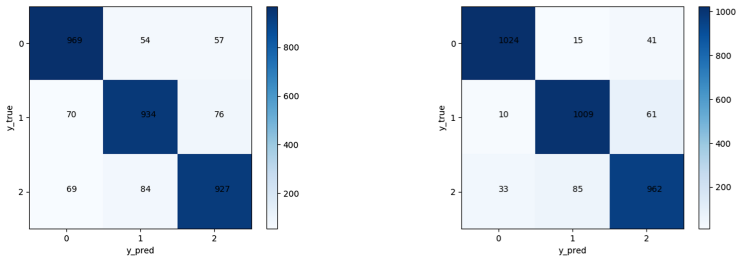
Different convolutional blocks convey different information, the lower convolutional blocks extract low-level features, the upper extract high-level features.

And the closer to the lower layer, the vaguer the information CBAM extracts, the less contribution to the classification task. The higher convolutional blocks, especially the third and fourth convolutional block, extract more specific and useful information. And these convolutional blocks contribute more to classification tasks. Therefore, the CBAM module placed between the last two convolutional blocks performs best. The information extracted using three convolutional blocks cannot accomplish classification, so it is the least effective. The above experiments also show that different placements are very important for the results.

4.4 The Effect of Adding 1×1 Convolution Kernel

In this part, we conduct a series of experiments to validate the effect of adding 1×1 convolutions. In Table 1, P3 and P4 are used to compare the effectiveness. When we compare P1 with P4, we can find that adding 1×1 convolution is definitely effective and it performs better than P2. We use 1×1 convolution and it can increase nonlinear characteristics without losing resolution. It can also reduce dimension and increase dimension. In P3 and P4, we change the number of output channels to strengthen the information interaction between channels.

So we add both CBAM and 1×1 convolution into the model which is shown as P3. Figure 5(a) is the confusion matrix of P2 and Fig. 5(b) is the confusion matrix of P3. During the test process, we use 3240 images. From the confusion



(a) (b)

Fig. 5. Two types of convolutional blocks used in the experiment. (a) is the classical one which is widely used in few-shot learning. (b) is the modified convolutional blocks, which is added 1×1 convolution kernel and dropout.

Table 3. Classification performance of the proposed method on brain tumor dataset. The values in parentheses are the results of adding the 1×1 convolution kernel and CBAM module. The values outside parentheses are the results of adding CBAM module.

Type	Accuracy	Precision	Recall	F1-score
Meningioma	0.90(**0.95**)	0.87(**0.96**)	0.90(**0.95**)	0.89(**0.95**)
Glioma	0.86(**0.93**)	0.87(**0.91**)	0.86(**0.93**)	0.87(**0.92**)
Pituitary tumor	0.86(**0.89**)	0.87(**0.90**)	0.86(**0.89**)	0.87(**0.90**)

matrixes, we can easily calculate the accuracy, precision, recall and f1-score, which is list in Table 3. From the table, we can see that the effect of the model is greatly improved after the addition of 1×1 convolution, thus further demonstrating that information interaction between channels is very useful for image classification.

4.5 Comparison of Different Methods

In the experiment, we compare the transfer learning method with our method. We use AlexNet, VGG16, ResNet101 [9] and DenseNet169 [10] pre-trained models which are modified with the last classification layer.

The results are presented in Table 4. We use the Prototypical Network to train and test on the brain tumor dataset and achieve the accuracy of 92.44%. The experimental results show that compared to models with pre-trained, our method has obvious advantages. The parameters of the pre-trained models are usually trained on the ImageNet dataset, while the natural images on ImageNet are quite different from the medical images, which often lead to unsatisfactory results. The performance of Dual Path Network [4] which combines ResNeXt and DenseNet is also not good and we find that when the embedding module is modified to ResNet, DenseNet or other deeper networks, the results are not ideal, which may be caused by the characteristics of the medical image itself. Therefore, our method performs better than these deeper networks.

Table 4. Results on brain tumor dataset

Method	Accuracy
AlexNet	79.80%
VGG16	82.08%
ResNet101	82.57%
DenseNet169	83.06%
Dual Path Network	82.74%
Our method	**92.44%**

5 Conclusion

In this paper, we adopt the few-shot learning model to solve the problem of medical image classification. We concretely choose Prototypical Network because this approach is far simpler and more efficient than other meta-learning approaches. We improve it and add a CBAM module based on the attention mechanism and 1×1 convolution kernel into the embedding module, which greatly strengthens the presentation ability of the model. We perform several comparative experiments on the brain tumor dataset. The results prove that these combinations are

applicable to the classification of medical images. As the interpretability of models and results is very important, in future studies, we will continue to explore the interpretability of the model to explain why different positions of CBAM have different effects on the final results. In our method, we use Euclidean distance to calculate the distance between the prototype and query image and we will try other distance measures for further study.

Acknowledgments. We thank all viewers who provided the thoughtful and constructive comments on this paper. This research is funded by Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China. The computation is supported by ECNU Multifunctional Platform for Innovation (001).

References

1. Brain tumor. https://figshare.com/articles/brain_tumor_dataset/1512427
2. OCT and Chest X-Ray. https://data.mendeley.com/datasets/rscbjbr9sj/2
3. Afshar, P., Plataniotis, K.N., Mohammadi, A.: Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries. In: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, United Kingdom, 12–17 May 2019, pp. 1368–1372. IEEE (2019)
4. Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., Feng, J.: Dual path networks. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017, pp. 4467–4475 (2017)
5. Chen, Z., Fu, Y., Wang, Y., Ma, L., Liu, W., Hebert, M.: Image deformation meta-networks for one-shot learning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, 16–20 June 2019, pp. 8680–8689. Computer Vision Foundation/IEEE (2019)
6. Chung, Y., Weng, W.: Learning deep representations of medical images using Siamese CNNs with application to content-based image retrieval. CoRR abs/1711.08490 (2017)
7. Deepak, S., Ameer, P.M.: Brain tumor classification using deep CNN features via transfer learning. Comput. Biol. Med. **111**, 103345 (2019)
8. García-Floriano, A., Ferreira-Santiago, Á., Camacho-Nieto, O., Yáñez-Márquez, C.: A machine learning approach to medical image classification: detecting age-related macular degeneration in fundus images. Comput. Electr. Eng. **75**, 218–229 (2019)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016, pp. 770–778. IEEE Computer Society (2016)
10. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017, pp. 2261–2269. IEEE Computer Society (2017)
11. Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., Qian, W.: An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2019, Berlin, Germany, 23–27 July 2019, pp. 2045–2048 (2019)

12. Kim, J., Kim, T., Kim, S., Yoo, C.D.: Edge-labeling graph neural network for few-shot learning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, 16–20 June 2019, pp. 11–20. Computer Vision Foundation/IEEE (2019)
13. Lin, M., Chen, Q., Yan, S.: Network in network. In: Bengio, Y., LeCun, Y. (eds.) 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, 14–16 April 2014, Conference Track Proceedings (2014)
14. Ma, T., Zhang, A.: Affinitynet: semi-supervised few-shot learning for disease type prediction. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, 27 January–1 February 2019, pp. 1069–1076. AAAI Press (2019)
15. Munkhdalai, T., Yu, H.: Meta networks. In: Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6–11 August 2017, pp. 2554–2563 (2017)
16. Puch, S., Sánchez, I., Rowe, M.: Few-shot learning with deep triplet networks for brain imaging modality recognition (2019)
17. Snell, J., Swersky, K., Zemel, R.S.: Prototypical networks for few-shot learning. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017, pp. 4077–4087 (2017)
18. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H.S., Hospedales, T.M.: Learning to compare: relation network for few-shot learning. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018, pp. 1199–1208. IEEE Computer Society (2018)
19. Wang, X., Yu, F., Wang, R., Darrell, T., Gonzalez, J.E.: Tafe-net: task-aware feature embeddings for low shot learning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, 16–20 June 2019, pp. 1831–1840. Computer Vision Foundation/IEEE (2019)
20. Wang, X., Hua, Y., Kodirov, E., Hu, G., Garnier, R., Robertson, N.M.: Ranked list loss for deep metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, 16–20 June 2019, pp. 5207–5216. Computer Vision Foundation/IEEE (2019)
21. Wong, K.C.L., Moradi, M., Wu, J.T., Syeda-Mahmood, T.F.: Identifying disease-free chest x-ray images with deep transfer learning. CoRR abs/1904.01654 (2019). http://arxiv.org/abs/1904.01654
22. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
23. Zeng, X., Chen, H., Luo, Y., Ye, W.B.: Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network. IEEE Access **7**, 30744–30753 (2019)