Few-shot Learning Framework Based on Adaptive Subspace for Skin Disease Classification

Chuan Zhou[1], Mengqi Sun[1], Leiting Chen[1,2]*, Anping Cai[1], Jiahao Fang[1],
[1]*Key Laboratory of Digital Media Technology of Sichuan Province, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan Province, 611731, China*
[2]*Institute of Electronic and Information Engineering of UESTC in Guangdong, Dongguan, Guangdong Province, China*
zhouchuan@uestc.edu.cn, 202022080805@std.uestc.edu.cn, richardchen@uestc.edu.cn,
202022080807@std.uestc.edu.cn, 202022080806@std.uestc.edu.cn

2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-6819-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIBM55620.2022.9995042

***Abstract*—Skin disease classification from images is crucial to dermatological diagnosis. It is an important task to develop Computer Aided Detection(CAD) systems that can help dermatologists improve classification performances. One challenge limits the adoption of such systems so far: traditional CAD can't handle new emerging diseases. The reason is that standard deep learning models can only identify the categories that appear in the training set. However, this limitation can be addressed with Few-shot learning, a machine learning paradigm where a classifier has to generalize to a new category not seen in training, only given a few examples of this category. However, because of the complexity of lesions, most Few-shot learning methods do not work well in medical tasks. We find that most methods assume a single similarity measure and only obtain a single feature space. Motivated by this, we propose a three-stage learning paradigm. In the second stage, we introduce the subspace method to construct a symmetric function. In the third stage, we propose a metric module that consists of two similarity measures . In this way, the model enables to learn more discriminative features from few shots of skin disease images and has better generalization ability. The results demonstrate that our method produces a substantial improvement on the ISIC-2019 dataset.**

***Index Terms*—few-shot learning, skin disease classification, metric space, adaptive subspace**

I. Introduction

Skin disease is one of the most common diseases in the world, and the number of deaths by skin diseases has increased in recent years, which aroused public attention[1]. Skin diseases cover all cultural regions and occur in all ages. According to statistics, about 30% to 70% of the global population belongs to high-risk groups of skin diseases [2]. Diagnosis time is very crucial for skin diseases. Among malignant tumors, melanoma is the most serious and fatal. If it can be diagnosed and found in the early stage, the survival rate is as high as 96%, while if it is diagnosed in the late stage, the survival rate is only 5% [3]. To decrease the associated consequences, cost, mortality ,and morbidity rate, skin diseases should be treated in their initial stages. As a key step in dermatological diagnosis, skin disease classification is quite challenging. Such complexity in skin disease taxonomy requires a great deal of expertise. In addition, there is a shortage of dermatologists in most underdeveloped areas, which necessitates research for computer-aided diagnosis.[4-5]

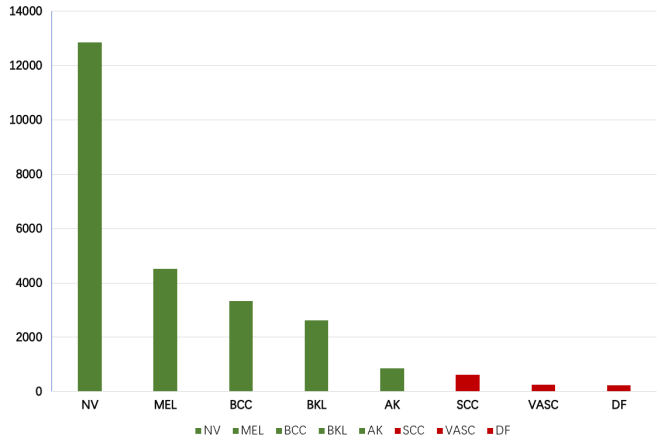

Fig. 1. The skin diseases dataset has a long-tailed data distribution.

Computer-aided diagnosis systems based on machine learning have been widely used in the early stage. Many researchers have done investigations to broaden the availability of dermatology expertise with machine learning techniques. Subsequently, deep learning models have achieved great success in visual recognition[6-7]. In particular, convolutional neural networks (CNNs) have shown excellent performance in classifying images and achieved human-level performance in various medical image diagnosis tasks [8-10]. Motivated by these successes, more recently, deep learning has been increasingly used in medical applications. It is of tremendous significance in the medical field. Despite the current research

*Leiting Chen is the corresponding author.

achievement, skin lesions classification is still a challenging task due to the following reasons:

(1) Deep networks often require a large number of labeled instances to learn their parameters. However, skin lesion datasets often has long-tailed class distribution, many categories only contain very few samples. The data distribution if shown in Figure 1. Various studies show that many deep learning techniques will fail to produce reliable models that generalize well if limited annotations are available.

(2) New skin diseases often appear in the field of Dermatology. The training mechanism of deep learning makes it unable to recognize new categories unless trained again. It's very time-consuming and labor-intensive.

In contrast to the current trend in deep learning, human can adapt fast based on transferable knowledge from previous experience and learn new concepts from only few observations. This in turn provides humans with lifelong learning abilities [11]. It is very important to design deep learning algorithms to have such abilities, this type of learning, known as Few-Shot Learning (FSL). FSL focus on maintaining the good generalization of the target task by acquiring transferable knowledge from a lot of auxiliary tasks [12]. Inspired by the recent success of Few-Shot Learning (FSL) in natural image classification, we propose to apply FSL to skin disease classification to solve the above two problems: the extreme scarcity of training samples and the generalization ability of ordinary deep learning models in new categories. We found that the direct application of FSL to skin disease classification tasks will not achieve the same good results as natural images. After our analysis, we find that the problem should be that compared with natural images, skin disease datasets have many similar sub-categories and the lesions are complicated, which bring great challenges to FSL classification task.

In this paper, we first formulate FSL as a three-stage learning paradigm, namely, 1. learning a universal feature extractor followed by 2. learning a symmetric function from high-dimensional data followed by 3. learning to classify through similarity measures. Once we establish the three-stage learning paradigm, we turn our attention to the second and third stages.

In the second stage, we propose to construct the symmetric function using subspaces that have a long history in modeling visual data[13-15]. As far as we know, this is the first application in few-shot skin diseases classification task. Using subspaces differs in large from previous studies where the symmetric function is realized through a form of average pooling [16]. The comparison between the average pooling method and the subspace method is shown in Figure 2.

In the third stage, we use two similarity measures, the Euclidean distance measure ,and the cosine distance measure. The obtained features can simultaneously adapt two similarity measures of diverse characteristics and the samples within one class can be mapped more compactly into a smaller feature space. In this way, our network can be more discriminative.

The main contributions of our work are as follows:

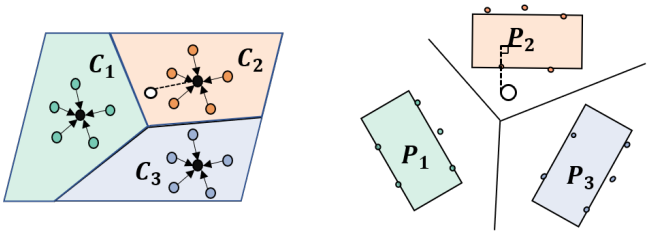

Fig. 2. The comparison between the average pooling method and the subspace method in constructing symmetric functions

- We propose a novel three-stage learning paradigm for skin disease classification as a FSL problem. And this learning paradigm is universal to all FSL problems.
- We use adaptive subspace to construct the symmetric function. As far as we know, no work applies subspace to skin disease classification.
- Our proposed framework effectively combines two similarity measures: the Euclidean distance and the cosine distance, which could generate more discriminative features than using a single measure.
- We conducted comparison experiments between many excellent metric-learning methods and our method. The results demonstrate the effectiveness of our framework.

II. Related Work

In this section, we review the literature on few-shot learning and subspace methods for classification tasks. Many few-shot classification algorithms can be grouped into two branches: meta-learning based methods [17-20] and metric-based methods [21-24]. Then we will also introduce the application of subspace .

A. Meta-learning based methods

Bertinetto *et al.* proposed that through real-time, efficient and one-time calculation, a complete depth discrimination model can be obtained from a single sample, which can identify other targets belonging to the same category [17]. Finn *et al.* proposed MAML, a probabilistic meta-learning algorithm, which can sample the model for a new task from a model distribution. The method extends the model-agnostic meta-learning, which adapts to new tasks via gradient descent, and combines the parameter distribution trained by variational lower bound [19]. Qi *et al.* proposed weight imprinting, it directly sets weights for a new category based on an appropriately scaled copy of the embedding layer activations for that training example. Only a single imprinted weight vector is learned for each novel category, rather than relying on a nearest-neighbor distance to training instances as typically used with embedding methods[20]. Zhang *et al.* combined spatial transformer networks with meta-learning applied it to the classification of skin diseases [25].

B. Metric-based methods

Siamese Neural Networks restrict the input structure and automatically find features that can be generalized from new

samples. It is trained by supervised metric learning based on twin network, and then reuse the extracted features of that network for few-shot learning [26]. Vinyals *et al.* proposed matching nets based on memory and attitude, which makes it possible to learn quickly. The training and testing conditions must match. They train the network by showing only a few examples per class, switching the task from minibatch to minibatch [22]. Sung *et al.* proposed the Relation Network to calculate similarity scores to measure the similarity between two images [24]. Snell *et al.* [16] proposed the Prototype Network. In the metric learning algorithm, a metric space is learned. After extracting features from the image, a prototype is calculated for each class. Li *et al.* they designed a new cosine measurement module, which combined with metric networks to achieve better results on natural image datasets [27].

C. Application of subspace

Basri *et al.*[28] they proved that the set of images of a convex Lambertian object obtained under a wide variety of lighting conditions can be approximated accurately by a low-dimensional linear subspace. And they apply these algorithms to perform face recognition. DASC[29] model learns more favorable sample representations by deep learning for subspace clustering, it performs well on real data with complex underlying subspaces. For one-class learning problem,Wang *et al.*[30] their key idea is to use a pair of orthonormal frames as subspaces and optimizing for two objectives: minimize the distance between the origins of the two subspaces, and maximize the margin between the hyperplanes and the data. Simon *et al.*[31] they introduce a subspace classifier and develop a discriminative form to constructe a few-shot learning framework.

Inspired by previous work, we first apply subspace to the metric learning module and improve the classification results on skin disease images.

III. Methodology

A. Problem Formulation

Unlike traditional classification problems, the label set in the training set is identical to those in the validation and test sets. Few-shot learning is needed to classify the novel classes. Few-shot learning is often formalized as C-way K-shot classification problems, in which models are given K labeled images from each of C classes. There are also two concepts in the field of Few-shot: support set S_i and query set Q_i. Both support set S_i and query set Q_i are from the training set, and they contain the same category, but the number of samples can be different. *eg.* In Prototype Network [16], support set is used to train the prototype of each category, and query set is used to evaluate the quality of the prototype.

B. Embedding module.

The complete structure of our network is shown in Figure 3. The network has one embedding module f_θ, followed by a bi-similarity metric module consisting of two similarity measurement branches g_φ and h_γ. Following [32], we adapt convolution-based architecture as our embedding module f_θ. The embedding module has four convolutional layers. each convolution block of Conv4 has 3×3 convolution of 64 filters, followed by batch normalization and a ReLU activation function. The embedding module will generate representations $f_\theta(x_i)$ and $f_\theta(q_i)$ for the image samples x_i and q_i, here x_i denotes the image samples from S_i and q_i denotes the image samples from Q_i.

C. Subspace module

In Few-shot learning, the essence of learning a classifier is to learn a symmetric function, we use subspaces to construct the symmetric function. To this end, the final layer of a neural network along the softmax layer implements:

$$p(c|q) = \frac{exp(W_C^T f_\theta(q))}{\sum_{c'} exp(W_{c'}^T f_\theta(q))} = \frac{exp(d_c(q))}{\sum_{c'} exp(d_{c'}(q))}, \quad (1)$$

where W_c is a weight of class C. Then the key to Few-shot problems is how to generate the $d_c(q)$ when solving a new task. We define M_c as the subspace of the C^{th} class. Inspire by the method of constructing subspace in points model, each subspace Z_i has a basis represented by $R^{D\times n} \ni B_i = [b_1, \cdots, b_n]$, Our goal is to find a basis represented P_c, then use P_c to generate subspace M_c. Below, we describe how to create a subspace classifier.

$$\mu_c = \frac{1}{k} \sum_{x_i \in X_c} f_\theta(x_i) \quad (2)$$

where $f_\theta(x_i) : \chi \to R^d$ represents the feature extraction process for the image sample x_i, k represents K shot, each time we choose k labeled images from each of C classes.

A new set of samples can be expresses as $\tilde{X}_c$:

$$\tilde{X}_c = [f_\theta(x_{c,1}) - \mu_c, ..., f_\theta(x_{c,K}) - \mu_c] \quad (3)$$

A basis for the subspace representing class C can be obtained by truncated singular value decomposition (SVD).

$$\tilde{X}_c = U \sum V^T \quad (4)$$

P_c is calculated from $\tilde{X}_c$, we choose the first N dimension of U to construct P_c.

$$M_c = P_c P_c^\top \quad (5)$$

Now a query q_i can be projected onto P_c and the classification based on the shortest distance from the query to its projection onto P_c. And we use softmax function to calculate the probability of q_i to each class.

$$d_c(q) = -||(I - M_c)(f_\theta(q) - \mu_c)||^2 \quad (6)$$

$$p_{c,q} = p(c|q) = \frac{exp(d_c(q))}{\sum_{c'} exp(d_{c'}(q))} \quad (7)$$

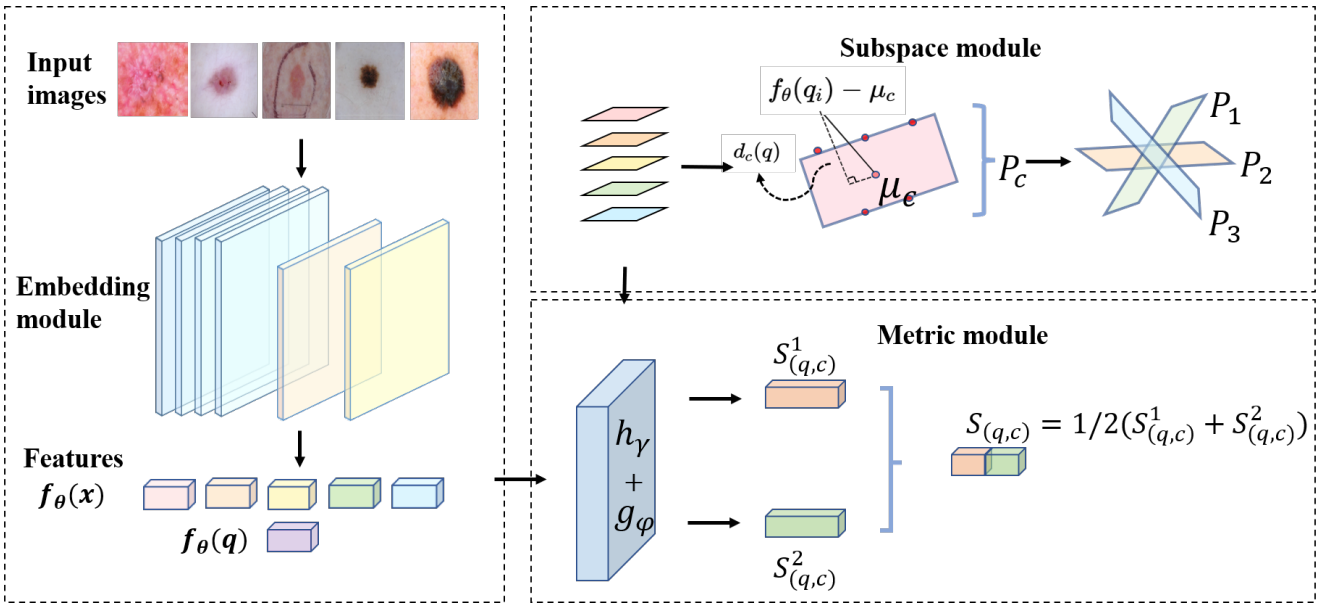

Fig. 3. The overall framework of our approach. The first stage include one embedding module, which consists of four convolutional layers followed by a batch normalization layer and a ReLU activation function. For the second stage, learning a symmetric function, we take subspace as an example. The third stage include one metric module, g_φ and h_γ using Euclidean distance measure and cosine distance measure respectively.

D. Bi-similarity metric module.

The bi-similarity metric module has two measurement modules h_γ and g_φ. During the training process, we will calculate two similarity scores $S^1_{(q,c)}$ and $S^2_{(q,c)}$ for the q^{th} query image, then we can get two one-hot vectors $\hat{y}_q^{(i,1)}$ and $\hat{y}_q^{(i,2)}$, as two predictions. Specific construction methods and calculation rules are illustrated by examples. Here we take the prototype module and the cosine similarity module as examples. Euclidean distance algorithm embedded in the prototype module. The specific calculation formula of the distance between $f_\theta(q_i)$ and $f_\theta(x_i)$ is $||f_\theta(q_i) - f_\theta(x_i)||^2$. For the K -shot scenario, the mean of support images features is the prototype for each class.

$$S^1_{(q,c)} = g_\varphi(\frac{1}{K}\sum_{i=1}^{K} f_\theta(x_{(i,c)}), f_\theta(q_i)). \tag{8}$$

The cosine similarity module has two parts, the first is a cosine similarity layer h^{cos}_γ , and the second is two convolution blocks h^{em}_γ. The cosine module fits well with the prototype module, relation module, matching module ,and subspace module, specific experimental results are shown in Sec. IV.

$$S^2_{(q,c)} = h^{cos}_\gamma(h^{em}_\gamma(\frac{1}{K}\sum_{i=1}^{K} f_\theta(x_{(i,c)})), h^{em}_\gamma(f_\theta(q_i))), \tag{9}$$

One-hot vectors $\hat{y}_q^{(i,1)}$ and $\hat{y}_q^{(i,2)}$ are calculated as follows:

$$\hat{y}_q^{(i,1)} = \left[0, \cdots, \underbrace{1}_{argmax_c S^1_{(q,c)}, c\in\{1,\cdots,C\}}, \cdots, 0\right], \tag{10}$$

$$\hat{y}_q^{(i,2)} = \left[0, \cdots, \underbrace{1}_{argmax_c S^2_{(q,c)}, c\in\{1,\cdots,C\}}, \cdots, 0\right] \tag{11}$$

By comparing the two assignments $\hat{y}_q^{(i,1)}$ and $\hat{y}_q^{(i,2)}$ with the $\hat{y}_q^i$, we can get the two loss values of the query image x_q^i, and take the average of the two values for back-propagation. $\hat{y}_q^i$ denotes the one-hot vector of the query image's ground-truth label. The total loss value $Loss_q^i$ is calculated from the two parts:

$$\begin{cases} Loss_q^{(i,1)} = \sum_j^C (\hat{y}^{(i,1)}_{(q,j)} - \hat{y}^i_{(q,j)})^2, q = 1, \cdots, |Q_i|, \\ Loss_q^{(i,2)} = \sum_j^C (\hat{y}^{(i,2)}_{(q,j)} - \hat{y}^i_{(q,j)})^2, q = 1, \cdots, |Q_i|, \\ Loss_q^i = \alpha \times Loss_q^{(i,1)} + \beta \times Loss_q^{(i,2)}, \end{cases} \tag{12}$$

The matching module builds different encoders for the S_i and the Q_i, and the output of the final classifier is the weighted sum of the predicted values between the x_s^i and x_q^i. The model uses recent advances in attention and memory that enable rapid learning. The memory network uses a simple bidirectional LSTM, which further processes the x_s^i from the embedding module f_φ, and a softmax layer is used in the attention module. The matching module can generate labels for unknown types without changing the model.

The relation module judges whether two images come from the same type by calculating the similarity score between the two images. The query image's representation $f_\varphi(x_q^i)$ is concatenated to every support image's representation $f_\varphi(x_s^i)$, for each query image x_q^i , the concatenated feature is directly

fed forward into two further convolution blocks and two fully connected layers to get similarity scores.

IV. Experiment

A. Experiment Settings

Dataset. In order to meet the training requirements, the data set should have sufficient classes and the data distribution should be unbalanced. We choose ISIC 2019 Skin Lesion Analysis Towards Melanoma Detection dataset [33] for experiment. The data set contains 8 types of images, with a total of 25331 images, including Melanoma(MEL), Melanocytic nevus(NV), Basal cell carcinoma(BCC), Actinic keratosis(AK), Benign keratosis(BKL), Dermatofibroma(DF), Vascular lesion(VASC) and Squamous cell carcinoma(SCC). The quantity distribution of images of each type is shown in Table IV. To simulate the rare disease classification scenario, we adopted the three classes with the least number: SCC, DF, VASC as the test set, the rest are used as training set.

Training Details. Our proposed method is implemented using PyTorch 1.6.0 on Ubuntu 18.04, and is trained on NVIDIA GeForce RTX 2060, which has 6 GB memory. In the training phase, we use Adam optimizer, with the momentum of 0.9 and the initial learning rate of 1e-3. Limited by the memory, all images are resized to 84×84 and $batchsize = num_{support} + num_{query}$. We found that a large number of FSL papers only choose ACC as the evaluation index, for better comparison, we also use ACC to assess the classification performance of our method.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}. \tag{13}$$

B. Parameter analysis

The formula proposed in section III-D shows that the total loss value is composed of two parts $Loss_q^{(i,1)}$and $Loss_q^{(i,2)}$. Before comparison experiments, we must first determine the contribution of $Loss_q^{(i,1)}$and $Loss_q^{(i,2)}$ in the process of back propagation. In this section, we will conduct parameter analysis experiment on S&C metric module ,we constantly adjust the two weights α and β of $Loss_q^{(i,1)}$and $Loss_q^{(i,2)}$ from 0 to 1. The experimental results are shown in Table I . From the results, we can see that $\alpha = 0.5$ and $\beta = 0.5$ are the most suitable weights setting. We know that the most suitable weight settings for different modules should be slightly different, but we need fixed weights with high universality, so that the training efficiency will be higher.

C. Comparison experiments

After determining the weights of α and β, in order to verify the ideas described in the methodology section, we divide the comparative experiments into two parts . The first part, we verify the effectiveness of adaptive subspace method in constructing symmetric functions. The second part, we verify the effectiveness of our metric module.

subspace method. In order to prove the effectiveness of the dynamic subspace method, we compare several methods of constructing symmetric functions. The results are shown in Table II, it can be seen from the experimental results that adaptive subspace method always performs well in the k-shot scenario, and the classification accuracy is better than other methods.

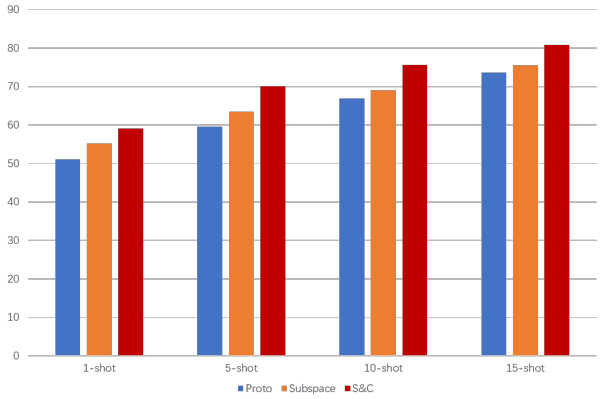

Fig. 4. The accuracy of prototype module, subspace module and S&C module in different shot scenario.

Metric module. In this part, in order to prove the effectiveness of our metric module, we keep g_φ using cosine distance measure and change h_γ with subspace module, prototype module, matching module and relation module. Euclidean distance algorithm embedded in these modules. The experimental results are shown in Table III. By comparing the results, the classification effect of the combination of two similarity measures is obviously better than one.

As can be seen from Figure 4, through our improvement on the second and third stages of the three-stage learning paradigm, the classification effect is gradually improved.

V. Conclusion

In this paper, we formulate FSL as a three-stage learning paradigm, and apply it to skin disease classification. In medical filed, new diseases emerge in endlessly, so it's very important for model to have the learning ability like human beings. Compared with natural images, because of the complexity of the lesion, it is more difficult to train the FSL model on medical images. We make improvements in two aspects. First we introduce the subspace method to construct a symmetric function, and second we defining a new metric strategy, which combines two similarity measures. Experiments on ISIC-2019 demonstrate the effectiveness of our method, and our three-stage learning paradigm can be applied on other FSL module in future work.

Acknowledgment

This study was supported by Sichuan Science and Technology Program (No.2019YFQ0005).

TABLE I

PARAMETER ANALYSIS ON WEIGHTS OF h_γ AND g_φ. HERE α DENOTES THE WEIGHT OF h_γ AND β DENOTES THE WEIGHT OF g_φ. THE BEST RESULTS ARE SHOWN IN BOLD IN THE TABLE.

α	β	3-way ACC			
		1-shot	5-shot	10-shot	15-shot
0	1	$47.83\% \pm 0.97$	$62.82\% \pm 0.68$	$67.44\% \pm 0.60$	$72.34\% \pm 0.56$
0.1	0.9	$56.88\% \pm 0.98$	$64.93\% \pm 0.72$	$71.65\% \pm 0.71$	$75.48\% \pm 0.5$
0.3	0.7	$57.96\% \pm 0.98$	$64.97\% \pm 0.63$	$72.89\% \pm 0.71$	$77.08\% \pm 0.84$
0.5	0.5	**59.13**$\% \pm$ **0.97**	**70.12**$\% \pm$ **0.75**	**75.67**$\% \pm$ **0.73**	**80.87**$\% \pm$ **0.69**
0.7	0.3	$56.68\% \pm 0.93$	$64.87\% \pm 0.75$	$73.02\% \pm 0.87$	$76.12\% \pm 0.63$
0.9	0.1	$57.73\% \pm 0.94$	$67.04\% \pm 0.64$	$69.14\% \pm 0.50$	$75.09\% \pm 0.66$
1	0	$55.21\% \pm 0.98$	$63.51\% \pm 0.76$	$69.02\% \pm 0.76$	$75.49\% \pm 0.71$

TABLE II

THE CLASSIFICATION RESULTS OF SUBSPACE MODULE, PROTOTYPE MODULE, RELATION MODULE, MATCHING MODULE AND COSINE MODULE IN MULTIPLE K-SHOT SCENARIOS. THE BEST RESULTS ARE SHOWN IN BOLD IN THE TABLE.

Model	3-way ACC			
	1-shot	5-shot	10-shot	15-shot
Subspace	**55.21**$\% \pm$ **0.98**	**63.51**$\% \pm$ **0.76**	**69.02**$\% \pm$ **0.76**	**75.49**$\% \pm$ **0.71**
Prototype	$51.04\% \pm 0.98$	$59.52\% \pm 0.85$	$66.94\% \pm 0.51$	$73.67\% \pm 0.39$
Relation	$47.28\% \pm 0.98$	$60.07\% \pm 0.72$	$67.57\% \pm 0.60$	$72.99\% \pm 0.56$
Matching	$52.12\% \pm 0.92$	$62.17\% \pm 0.83$	$66.31\% \pm 0.55$	$72.01\% \pm 0.37$
Cosine	$47.83\% \pm 0.97$	$62.82\% \pm 0.68$	$67.44\% \pm 0.60$	$72.34\% \pm 0.56$

TABLE III

COMPARISON EXPERIMENTS TO PROVE THE EFFECTIVENESS OF THE BI-SIMILARITY METRIC MODULE. KEEP g_φ UNCHANGED, ONLY CHANGE THE SYMMETRIC FUNCTION USED IN h_γ , S&C REPRESENTS h_γ USE SUBSPACE MODULE TO CALCULATE THE SYMMETRIC FUNCTION,g_φ USE COSINE DISTANCE MEASURE. THE BEST RESULTS ARE SHOWN IN BOLD IN THE TABLE.

Model	3-way ACC			
	1-shot	5-shot	10-shot	15-shot
Subspace	$55.21\% \pm 0.98$	$63.51\% \pm 0.76$	$69.02\% \pm 0.76$	$75.49\% \pm 0.71$
S&C	**59.13**$\% \pm$ **0.97**	**70.12**$\% \pm$ **0.75**	**75.67**$\% \pm$ **0.73**	**80.87**$\% \pm$ **0.69**
Prototype	$51.04\% \pm 0.98$	$59.52\% \pm 0.85$	$66.94\% \pm 0.51$	$73.67\% \pm 0.39$
P&C	**55.57**$\% \pm$ **0.97**	**65.33**$\% \pm$ **0.83**	**70.32**$\% \pm$ **0.56**	**76.62**$\% \pm$ **0.34**
Relation	$47.28\% \pm 0.98$	$60.07\% \pm 0.72$	$67.57\% \pm 0.60$	$72.99\% \pm 0.56$
R&C	**49.18**$\% \pm$ **0.97**	**67.13**$\% \pm$ **0.70**	**69.10**$\% \pm$ **0.63**	**75.67**$\% \pm$ **0.66**
Matching	$52.12\% \pm 0.92$	$62.17\% \pm 0.83$	$66.31\% \pm 0.55$	$72.01\% \pm 0.37$
M&C	**53.51**$\% \pm$ **0.94**	**64.07**$\% \pm$ **0.73**	**70.65**$\% \pm$ **0.53**	**75.78**$\% \pm$ **0.33**

TABLE IV

DISTRIBUTION OF SKIN DISEASES IN THE ISIC-2019 DATASET

Disease type	AK	BCC	BKL	MEL	NV	**SCC**	**DF**	**VASC**
Samples	867	3323	2624	4522	12875	**628**	**239**	**253**

REFERENCES

[1] Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. JAMA Dermatol. 151(10), 10811086 (2015).

[2] M. T. Johnson and J. Roberts, "Skin conditions and related need for medical care among persons 174 years. United States 19711974", Vital Health Stat., vol. 11, no. 212, pp. i-v, Nov. 1978.

[3] L. -F. Li, X. Wang, W. -J. Hu, N. N. Xiong, Y. -X. Du and B. -S. Li, "Deep Learning in Skin Disease Image Recognition: A Review," in IEEE Access, vol. 8, pp. 208264-208280, 2020

[4] Okuboyejo, D.A., Olugbara, O.O., Odunaike, S.A.: Automating skin disease diagnosis using image classification. In: Proceedings of the World Congress on Engineering and Computer Science, vol. 2, pp. 850854 (2013)

[5] Sumithra, R., Suhil, M., Guru, D.: Segmentation and classification of skin lesions for disease diagnosis. Procedia Comput. Sci. 45, 7685 (2015).

[6] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature, vol. 521, no. 7553, p. 436, 2015

[7] N. Dvornik, J. Mairal, and C. Schmid, Diversity with cooperation: Ensemble methods for few-shot classification, in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 37233731.

[8] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern and J. R. Smith, "Deep learning sparse coding and SVM for melanoma recognition in dermoscopy images", In International Workshop on Machine Learning in Medical Imaging, pp. 118-126, 2015.

[9] J. Kawahara, A. BenTaieb and G. Hamarneh, "Deep features to classify skin lesions. In Biomedical Imaging (ISBI)", 2016 IEEE 13th International Symposium on, pp. 1397-1400, 2016, April, 2016.

[10] T. Majtner, S. Yildirim-Yayilgan and J. Y. Hardeberg, "Combining deep

learning and hand-crafted features for skin lesion classification," 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2016, pp. 1-6.

[11] X. Li, Z. Sun, J.-H. Xue, and Z. Ma, A concise review of recent fewshot meta-learning methods, Neurocomputing, to be publisshed, doi:10.1016/j.neucom.2020.05.114.

[12] L. Bertinetto, J. F. Henriques, J. Valmadre, P. Torr and A. Vedaldi, "Learning feed-forward one-shot learners", Proc. Adv. Neural Inf. Process. Syst., pp. 523-531, 2016.

[13] M. A. Turk and A. P. Pentland, Face recognition using eigenfaces, in Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586591.

[14] R. Basri and D. W. Jacobs, Lambertian reflectance and linear subspaces, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 218233, 2003.

[15] P. Zhou, Y. Hou, and J. Feng, Deep adversarial subspace clustering. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 15961604.

[16] J. Snell, K. Swersky and R. Zemel, "Prototypical networks for few-shot learning", Proc. Adv. Neural Inf. Process. Syst., pp. 4077-4087, 2017.

[17] L. Bertinetto, J. F. Henriques, J. Valmadre, P. Torr, and A. Vedaldi. Learning feed-forward one-shot learners in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 523531.

[18] C. Finn, P. Abbeel, and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in Proc. Int. Conf. Mach. Learn.,2017, pp. 11261135.

[19] C. Finn, K. Xu, and S. Levine, Probabilistic model-agnostic metalearning, in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 95169527.

[20] H. Qi, M. Brown, and D. G. Lowe, Low-shot learning with imprinted weights, in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.,Jun. 2018, pp. 58225830.

[21] H. Li, W. Dong, X. Mei, C. Ma, F. Huang, and B.-G. Hu, LGM-net: Learning to generate matching networks for few-shot learning, 2019,arXiv:1905.06331. [Online]. Available: http://arxiv.org/abs/1905.06331

[22] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu and D. Wierstra, "Matching networks for one shot learning", Proc. Adv. Neural Inf. Process. Syst., pp. 3630-3638, 2016.

[23] W. Li, L. Wang, J. Xu, J. Huo, Y. Gao and J. Luo, "Revisiting local descriptor based image-to-class measure for few-shot learning", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 7260-7268, Jun. 2019.

[24] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 1199-1208, Jun. 2018.

[25] D. Zhang, M. Jin and P. Cao, "ST-MetaDiagnosis: Meta learning with Spatial Transform for rare skin disease Diagnosis," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2153-2160

[26] S. Chopra, R. Hadsell and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 539-546

[27] X. Li, J. Wu, Z. Sun, Z. Ma, J. Cao and J. -H. Xue, "BSNet: Bi-Similarity Network for Few-shot Fine-grained Image Classification," in IEEE Transactions on Image Processing, vol. 30, pp. 1318-1331, 2021

[28] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 2, pp. 218-233, Feb. 2003, doi: 10.1109/TPAMI.2003.1177153.

[29] P. Zhou, Y. Hou and J. Feng, "Deep Adversarial Subspace Clustering," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1596-1604, doi: 10.1109/CVPR.2018.00172.

[30] J. Wang and A. Cherian, "GODS: Generalized One-Class Discriminative Subspaces for Anomaly Detection," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 8200-8210, doi: 10.1109/ICCV.2019.00829.

[31] C. Simon, P. Koniusz, R. Nock and M. Harandi, "Adaptive Subspaces for Few-Shot Learning," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 4135-4144, doi: 10.1109/CVPR42600.2020.00419.

[32] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang and J.-B. Huang, "A closer look at few-shot classification", Proc. Int. Conf. Learn. Representations, pp. 1-17, 2019.

[33] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC), 2017