FS3DCIoT: A Few-Shot Incremental Learning Network for Skin Disease Differential Diagnosis in the Consumer IoT

Junsheng Xiao, Jirui Li, and Honghao Gao, *Senior Member, IEEE*

***Abstract*—The computer-aided diagnosis (CAD) method based on few-shot learning (FSL) effectively reduces the dependence on labelled medical images. However, the catastrophic forgetting defect of neural networks seriously decreases the ability of CAD methods to meet real diagnosis needs. Moreover, the low classification accuracy and limited categories make FSL classification difficult. A few-shot class incremental learning (FSCIL) method for skin disease diagnosis (SDD) based on the consumer Internet of Things (CIoT) is designed to solve these problems in this paper. First, dermoscopic images and clinical images obtained from CIoT nodes are used to cover more skin disease categories, and a dual-flow modal alignment module is designed to mitigate the modal misalignment of different modal images. Second, a queue of gradient episodic memory (Q-GEM) method is designed to solve the catastrophic forgetting problem. Third, a differential diagnosis method (DDM), which can effectively improve the low classification accuracy of the few-shot learning (FSL) classification network, is designed. Experiments show that the top-3 diagnostic accuracy of the proposed method can match the accuracy level of dermatologists, the accuracy is 11.2% improvement over the SOTA method.**

***Index Terms*—Few-shot learning, skin disease, computer-aided diagnosis, consumer Internet of Things.**

I. Introduction

DEEP convolutional neural networks (DCNN) have made progress in various image understanding tasks, e.g., medical image analysis and telemedicine [1], [2]. However, these methods are dependent on many high-quality labelled samples [3], [4], [5]. FSL method can mitigate the difficulty of collecting and labelling skin disease images and it requires a data collection methods for obtaining more categories of samples [6], [7]. CIoT refers to the Internet of Things in the context of consumer applications and devices, a variety of medical sensors have been used to collect medical information by CIoT, providing new opportunities for obtaining samples [8]. Therefore, the consumer medical sensors will help to obtain more categories of disease samples and the FSL method based on CIOT will provide new opportunities for CAD. However, the FSL method based on a DCNN leads to more serious catastrophic forgetting since the DCNN has the inherent defect of the catastrophic forgetting problem. Catastrophic forgetting in a DCNN means that when the neural network is generalized to a new task, it will forget the knowledge learned from the old tasks [9]. Compared with machine learning systems, humans can easily learn new concepts with few examples without forgetting previously learned knowledge. Therefore, it is necessary to study an incremental learning network with limited samples that can continuously learn new knowledge without forgetting old knowledge. It is key to solving many practical problems, especially the classification of skin disease images with more than 2,000 categories, where most categories only have several available labelled samples. Fortunately, the few-shot class incremental learning (FSCIL) network can continuously learn new knowledge without forgetting old knowledge. FSCIL can solve many practical problems, especially regarding the classification of medical images in cases with only a small number of labelled samples.

The scarcity of new samples with different categories is a challenge for FSCIL, and it not only leads to serious overfitting but also exacerbates the catastrophic forgetting of old categories. Recently, different kinds of CIoT devices have conveniently provided samples, which has effectively alleviated scarce sample problem, and samples that cover more categories have been provided. Kiranmayi and Sharma proposed mobile apps based on CIoT for fisheries and aquaculture sectors [10]. Carter et al. proposed a health monitoring system based on CIoT for COVID-19 telemedicine [11]. Tao et al. proposed a method based on metalearning to solve the problem of incremental learning [12]. However, when this method is used in CAD, its accuracy is seriously reduced. Mohammad et al. studied the incremental learning problem in CAD [13]. However, they did not consider situations with insufficient training samples or samples in the medical field. Moreover, different kinds of CIoT devices usually generate different modes of samples. As shown in Fig. 1, an online smart dermatology diagnosis system is designed based on a CIoT platform. Support and query images are collected by medical equipment, e.g., a dermatoscope, a digital camera, or a confocal laser scanning microscope, on a CIoT platform,

Manuscript received 5 December 2022; revised 29 January 2023, 7 April 2023, and 30 May 2023; accepted 28 July 2023. Date of publication 22 August 2023; date of current version 21 February 2024. This work was supported in pert by the Association of Fundamental Computing Education in Chinese Universities under Grant 2023-AFCEC-242, and in pert by the Nursery Project of Henan University of Chinese Medicine under Grant MP2023-10. *(Corresponding author: Honghao Gao.)*

Junsheng Xiao and Jirui Li are with the School of Information Technology, Henan University of Chinese Medicine, Zhengzhou 450046, China (e-mail: xiaojunsheng@hactcm.edu.cn; ljrokyes@163.com).

Honghao Gao is with the School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China, and also with the College of Future Industry, Gachon University, Seongnam 461-701, Gyeonggi, South Korea (e-mail: gaohonghao@shu.edu.cn).

Digital Object Identifier 10.1109/TCE.2023.3301874

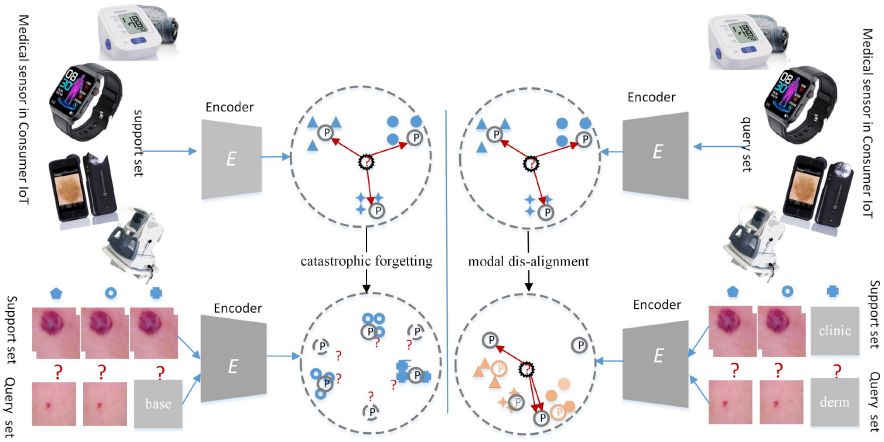

Fig. 1. Catastrophic forgetting in a few-shot dermatology classification network.

and the images have great differences in appearance so the different platforms are considered to be different modes. Then, support and query images are first extracted by the encoder, and their features are matched in the high-dimensional space by a metric function. The loss function evaluates the matching results and guides the network during training. However, since the parameter weights are continuously adjusted during training, the trained model cannot correctly classify the old categories when the parameters are generalized to new categories, which results in catastrophic forgetting. Since both dermoscopic images and clinical images are used in our method, the image feature distributions exhibit significant differences since they are extracted from different modalities. In particular, dermoscopic and clinical images do not necessarily appear in pairs, i.e., images from the same modality are also accepted.

The few-shot incremental learning method for multimodal images (FCILOMI) is proposed in this paper based on the CIoT platform. The proposed method tries to solve the catastrophic forgetting problem encountered when classifying skin diseases with few labelled samples. Multimodal images obtained from different CIoT devices are used to alleviate the long-tailed category distribution of skin disease image datasets. First, clinical images and dermoscopic images are included in the training and testing sets as multimodal samples to improve the generalization ability of the model. Moreover, the coverage of skin disease categories is increased. This approach imitates doctors simultaneously collecting dermoscopic and clinical images to make a diagnosis through a comprehensive comparative analysis. From the perspective of computer image processing, dermoscopic images generally contain detailed information about a certain diseased area, while clinical images may contain multiple diseased areas and large areas of normal skin. The combination of local information and overall information is beneficial for skin disease classification. However, from the perspective of image feature characteristics, the two kinds of images mentioned above are quite different. To align the features of dermoscopic images and clinical images, a dual-flow multimodal alignment network (DFMAN) is designed. The DFMAN is based on the characteristics of the FSL framework, in which the support branch and the query branch correspond to specific domain features. Then, a sharing layer converts the specific domain features into the embedding space. Second, the queue of gradient episode memory (Q-GEM) method is designed; this approach is based on metalearning to achieve few-shot incremental learning. Q-GEM integrates the GEM algorithm into the metalearning task with the metalearning training framework as the basic algorithmic framework. The method maintains the characteristics of metalearning and reasonably initializes new tasks, so it can also control the update speed of the parameters learned from the previous tasks. Third, the differential diagnosis method (DDM) is applied instead of the single classification method to provide doctors with more information to prevent misdiagnosis, which is an important feature of computer-aided diagnosis (CAD) systems. Experiments show that the method improves the performance of the network and effectively reduces the number of misdiagnoses. The contributions of this paper are summarized as follows.

1. The CIoT devices are used to obtain dermatology images and clinical images to train the FSL method, which covers more skin disease categories. A dual-flow modal alignment network (DFMAN) is designed to align dermoscopic images and clinical images in high-dimensional space.

2. The queue of gradient episode memory (Q-GEM) method is designed to solve the catastrophic forgetting problem based on the FSCIL module, which mostly maintains the generalization ability of the FSCIL network when multimodal images cover more skin disease categories.

3. The differential diagnosis method (DDM) effectively overcomes the problem of low classification accuracy when there are only a limited number of labelled samples and makes our method more practical in clinical applications.

II. Related Works

A. Incremental Learning

Neural networks are good at obtaining generalized knowledge through different training stages to solve classification tasks. However, the trained networks are built on static knowledge. When a network adapts to a new task, its

performance on the old task is reduced, and this is known as catastrophic forgetting. Therefore, it is necessary to restart the network training process when a new category of samples is available. However, this is difficult to realise in real scenarios, and it is necessary to design a system that can constantly adapt and learn to address new sample categories. Incremental learning, i.e., continuous learning or lifelong learning, refers to learning from continuous data streams and gradually expanding the learned knowledge to generalize it to new tasks [14]. Early incremental learning methods mainly solved the catastrophic forgetting problem with different input modes; these approaches include reducing the overlap among sample representations, replaying samples [15], and introducing two-way architectures [16]. These methods are based on shallow network architectures and generally only consider small numbers of samples. Recently, incremental learning methods with more samples and longer task sequences have attracted attention and are mainly divided into three types: sample playback [15], regularization [17], and parameter-independent methods [18]. A method based on sample playback stores samples in their original format or generates dummy samples using a generation model. These samples from the previous task are reused when the network is trained on a new task to mitigate catastrophic forgetting. For example, the GEM incremental learning method limits only new task updates, aiming to not interfere with the previous tasks [19]. This method uses a first-order Taylor series approximation to project the estimated gradient direction into the feasible region produced by the previous task. An additional regularization function is introduced to the loss function, which consolidates the previous knowledge when learning new tasks. The A-GEM method relaxes the constraint to a direction estimated by samples that are randomly selected from the data buffer of the previous task [20].

B. Multimodal Learning

Modality is the way things occur or exist, and multimodality is a combination of two or more modalities in various forms. Humans obtain information from the world in a multimodal manner, e.g., humans can see, hear, feel, smell, and taste. For artificial intelligence (AI) to progress so that it can be used to understand the world in a way similar to humans, it is necessary to interpret multimodal signals simultaneously. Audio-visual speech recognition (AVSR) was the earliest multimodal research, and it applied the combination of hearing and vision information in the process of speech perception [21]. Schwartz et al. proposed that an infant's ability to recognize objects is correlated with their language skills [22]. Recently, multimodal learning based on deep learning has attracted much interest. Large-scale datasets, faster image processing units, and visual and linguistic features are the key driving factors of multimodal machine learning research in the era of deep learning. Lu et al. proposed a pedestrian rerecognition method based on RGB images and infrared images, effectively improving the accuracy of pedestrian rerecognition at night [23]. Since dermoscopic images and clinical images that contain features belonging to the same skin disease category may be quite different in terms of appearance, they are not consistent. The difference between them is mainly due to the inconsistency among the utilized observation methods. Therefore, we believe that dermoscopic and clinical images can be seen as different modes, and the multimodal alignment method is used to align the dimensions in the high-dimensional space.

C. Consumer IoT

The IoT is an extension and expansion of the Internet by combining various information sensing devices with the network to form a huge network, which has been used in various scenarios, including medical scenarios [8]. The CIoT is the integration of IoT into regular consumer applications and devices [10], [11]. Therefore, the CIoT is closer to people's lives and changing their lives at a faster pace. For example, it is used in Home Security and Smart Homes, Consumer IoT and Personal Security, and Smart Healthcare. Especially, we focus on the applications of CIoT in Smart Healthcare. Such as Consumer IoT applications have improved personal healthcare by introducing wearable IoT-connected devices that transmit data to people's doctors or loved ones. A few-shot learning method based on the CIoT platform is designed in this paper depending on its convenience for obtaining data.

III. FCILOMI Algorithm

A. Task Description of Few-Shot Class Incremental Learning

FSCIL aims to design a machine learning algorithm that can constantly learn new categories from a small number of labelled training samples without forgetting the old categories [24]. Once a learning phase ends, the training sample from the previous phase is no longer available. However, during the evaluation stage, FSCIL needs to evaluate all previous training categories. $\{D^0_{train}, D^1_{train}, \ldots, D^n_{train}\}$ denotes training sets from different stages, where the relevant label space D^i_{train} is labelled by category set C^i. The datasets at different stages do not have overlapping categories; i.e., $\forall i, i \neq j, C^i \cap C^j = \emptyset$. In the i-th training stage, only D^i_{train} can participate in the training and verification process of the network. However, the test dataset $D_{base+novel}$ includes not only the categories of the current stage but also the categories of all previous stages; i.e., the category space is $\{C^0 \cup C^1 \cup, \ldots, C^n\}$. Generally, the training dataset D^0_{train}, as the first training set, is called the basic training set, which is usually a relatively large dataset, and it contains sufficient training samples for each category. However, the datasets in all subsequent sessions only include limited numbers of samples. The dataset for a specific session D^i_{train} is usually described as an N-way, K-shot training set, where there are n classes in the dataset and each class has k training samples. FSCIL defines a strict problem setting in which data imbalance and scarcity further aggravate the knowledge forgetting issue.

B. Network Architecture

As shown in Fig. 2, the network framework proposed in this paper includes three parts: (1) a dual-stream feature alignment module; (2) an FSCIL module; and (3) the DDM. The

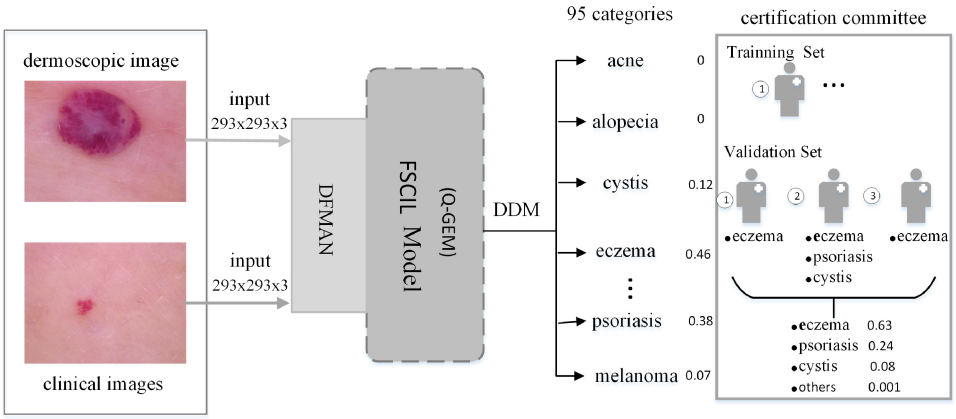

Fig. 2. Framework of our proposed method.

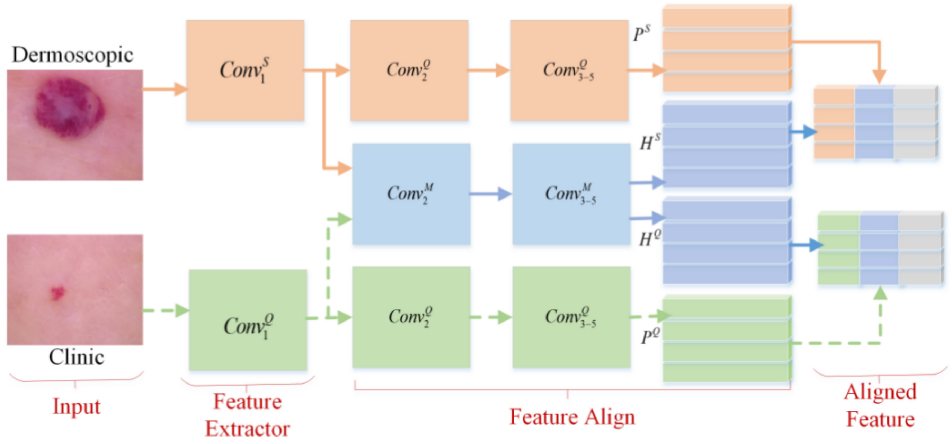

Fig. 3. Dual-flow feature alignment network.

relationship of the three parts can be expressed as follows:

$$ACC_t = d_\delta f_\theta[M(I(der), I(clc), Q_G EM(\tau))] \quad (1)$$

where Acc_t is the differential accuracy of the DDM, $t \in [1, 2, 3]$, the function $M(',')$ is implemented by the DFMAN, I_{der} and I_{clc} are the dermatology and clinical images, respectively, τ is the task in the Q-GEM method, f_θ is the FSCIL network, d_σ is the differential module, and θ and σ are their respective parameters. First, dermoscopic and clinical images are input into the DFMAN as the supporting and query images, respectively. The DFMAN aims to align different modal images with large appearance differences in the high-dimensional feature space. Then, the modal-aligned support and query images are sent to the FSCIL module. Third, the DDM module follows the FSCIL module to obtain the differential classification results, which are used to replace the single classification results.

C. Two-Stream Feature Alignment Network

The well-labelled skin disease images are only concentrated in the relatively fatal skin cancer categories. More than 400 common skin diseases have been studied for CAD systems, which is less than the more than 2,000 skin diseases found on the human body. Incremental learning aims to continuously expand its generalization ability through continuous learning to cover more image categories. Since the collection of clinical skin disease images does not require professional equipment, the clinical image dataset contains more skin disease categories. Therefore, although most deep learning methods train their networks using dermoscopic images to achieve classification [25], [26], some methods also utilize clinical images to train their networks for skin image classification [27]. However, no relevant study has been conducted on FSL with two sample modes of skin disease images. The proposed method combines dermoscopic images and clinical images to alleviate the lack of samples in most categories and to cover more skin disease categories. However, dermoscopic images and clinical images are different in terms of appearance and can be seen as two different modes of images. When two different modal samples are used together as training samples, achieving modal alignment is key. Based on the characteristics of the FSL framework composed of supporting image branches and query image branches, a DFMAN is designed to achieve modal alignment. Modal alignment is realized by the shared feature layer, which aims to embed images of any modality into the same feature space and achieve feature alignment in that same feature space. The dual-stream shared network structure is shown in Fig. 3.

The dual-stream feature sharing network is composed of three parts: a feature extraction module, a feature alignment

module and the output of the aligned features. First, the dermoscopic and clinical images are sent to the feature extractors of the FSL network. The feature extractors $Conv_1^S$ and $Conv_1^Q$ have the same structure, but their parameters are different so that they can extract different types of image features. Second, the image features are aligned by the feature alignment module. Third, the aligned features are sent to the FSCIL network to complete the classification task. In particular, the key process is the use of a shared network to model the affinity between and within modes. By calculating the affinity between the modes of different samples, the shared information is transmitted according to this affinity. Each sample accepts information from intermodal and intramodal sources. The information shared between them can mitigate the lack of specific information and enhance the robustness of the shared features to improve the overall representation ability of the model.

Each input image pair $X_m(m \in (S, Q))$ can generate shared modal features (blue blocks) and specific modal features (yellow blocks represent dermoscopic images, and green blocks represent clinical images) through the feature extractor. To achieve better performance, the shared features and specific modal features are separated in the shallow convolution layer as follows:

$$H^m = Feat^s\left(conv_2^s\left(conv_1^m\left(x^S\right)\right)\right) \tag{2}$$

$$P^m = Feat^m\left(conv_2^m\left(conv_1^m\left(x^Q\right)\right)\right) \tag{3}$$

Specifically, H^m denotes the shared features between the modes, P^m denotes specific modal features, and the function Feat() is a feature block to generate the shared feature and specific feature. The dual-flow network extracts the shared features and specific features of each mode. For the unified feature representation, a three-segment format is used to represent the features of each mode: [Derm-specific, Shared, Clin-specific]

$$z_i^C\left[P_i^C, H_i^C, 0\right], z_i^D\left[P_i^{dD}, H_i^D, 0\right] \tag{4}$$

where 0 denotes zero vector filling, which means there is no overlap between the dermoscopic and clinical images. During modal alignment, the specific mode features need to be converted into the shared modal space. Inspired by graph convolution, the proposed method uses a neighbour propagation network to spread information and retain the context structure of the whole network. The shared feature transmission network can promote modal alignment and enhance the feature representation ability and robustness of the model. The feature network first models the sample affinity between the two feature networks and calculates the shared intermodal and intramodal features using specific features:

$$A_{ij}^{m,m} = d\left(p_i^m, p_j^m\right) \tag{5}$$

$$A_{ij}^{m,m'} = d\left(H_i^m, H_j^{m'}\right) \tag{6}$$

where $A_{ij}^{m,m}$ is the affinity between the $i-th$ sample and the $j-th$ sample that belong to the same mode m, $A_{ij}^{m,m'}$ is the affinity between different categories, and $d(',')$ is the standard Euclidean distance:

$$(a, b) = 1 - \frac{1}{2} \times \left\| \frac{a}{\|a\|} - \frac{b}{\|b\|} \right\| \tag{7}$$

The similarities between different modes represent the modal relationships between different samples. The affinity matrix is defined as:

$$A = \begin{bmatrix} T\left(A^{C,C}, k\right) & T\left(A^{C,D}, k\right) \\ T\left(A^{D,C}, k\right) & T\left(A^{D,D}, k\right) \end{bmatrix} \tag{8}$$

where $T(., k)$ is the selected neighbour function, which guarantees that the top-k values in each row of the matrix can be obtained and set to 0. The shared features and specific features transmit information to each other via the affinity matrix. The affinity matrix represents the similarity between samples, and the feature propagation network propagates features on the matrix. Before propagation, the dermoscopic image features and clinical image features are linked through their dimensions, and each row stores the characteristics of the sample $z = [\frac{z^c}{z^d}]$. The affinity relationship of matrix D is obtained by a graph convolutional neural network:

$$\tilde{Z} = \begin{bmatrix} \tilde{Z}^c \\ \hline \tilde{Z}^d \end{bmatrix} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} Z W\right) \quad d_{ii} = \sum j A_{ij} \tag{9}$$

The feature matrix is first filled with the nearest neighbour structure ($D^{-\frac{1}{2}} A D^{-\frac{1}{2}} Z$) and then fused through a nonlinear variation function. After completing feature fusion, the transmitted features include the shared features and specific features of the two modes. The transmitted features $\tilde{Z}$ are defined as follows:

$$\tilde{Z} = \begin{bmatrix} \tilde{Z}^R \\ \hline \tilde{Z}^I \end{bmatrix} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} Z W\right) \tag{10}$$

where σ is the activation function, and the rectified linear unit (ReLU) function is used in this paper. W is the parameter of the transfer learning network. Finally, these propagation features are embedded into a feature space to optimize the whole learning process.

D. Q-GEM

The features of the support and query images are aligned and sent to the FSCIL module after model alignment. Most FSL frameworks are based on metalearning, which can effectively extend knowledge to samples outside the distribution that have been learned and quickly generalize to new categories. However, an FSL network based on metalearning is not able to overcome the catastrophic forgetting problem of neural networks. When the parameters of the network are updated during the training stage for the new categories, the overall distribution changes, resulting in a decrease in the recognition accuracy achieved for the old categories. Catastrophic forgetting is more significant in FSCIL tasks since the scarcity of samples makes it difficult to adapt to new categories and easily leads to the forgetting of old categories.

The GEM method is used to alleviate the catastrophic forgetting problem in deep learning, and we extend this theory to

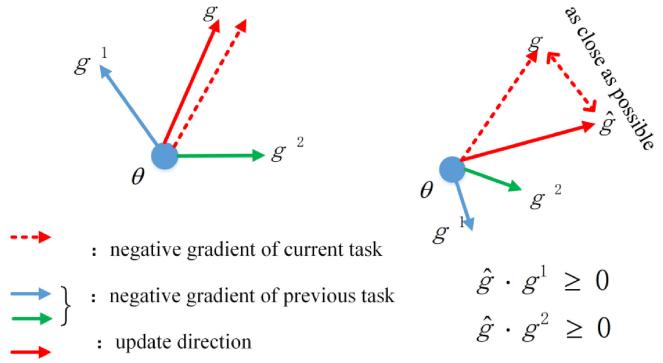

Fig. 4. Schematic diagram of gradient update direction calculation.

FSL in this paper [16]. In FSL, each training task is defined as D^i_{train}. Each sample impacts the whole network through the gradient g, and balancing the relationship between the gradient of the current task and the average gradient of all previous tasks can mitigate catastrophic forgetting. As shown in Fig. 4, the GEM method calculates the gradient angle change between the gradient loss g^i of each previous task and the gradient g of the current task. In particular, when the angle is greater than 90 degrees, the calculated angle is projected onto the closest edge by the L2 norm to keep the angle within the preset boundary. The formal definition is summarized as follows:

$$min_{\tilde{g}} \frac{1}{2}\|g - \tilde{g}\|_2^2 \quad s.t. \langle \tilde{g}, g_k \rangle \geq 0, \forall k < t \tag{11}$$

Although the GEM is very effective in a single training task, it needs to calculate the gradients of all previous tasks, and the calculation amount is large. Especially for FSL, whose number of learning tasks is relatively large and whose number of new categories is increasing dynamically, the computational and storage costs are further increasing. Chaudhry et al. proposed the A-GEM. method to improve the GEM method.

A-GEM randomly selects the gradients of a small batch of tasks in a gradient memory segment instead of calculating the gradients of all samples in episodic memories. The gradient subset is formed by samples that are randomly selected from the episodic memory to calculate the gradients. Then, the gradient subset replaces the gradients of all samples from the previous tasks. In other words, A-GEM replaces the $t-1$ constraint of the above algorithm with a single constraint by using the following constraint equation:

$$L\left(f_\theta^{t-1}, M\right) \leq L\left(f_\theta^{t-1}, M\right), where\ M = U_{k<t} M_k \tag{12}$$

However, experiments show that the randomly selected average gradient leads to the training process being unstable and nonconvergent. The FSCIL method aims to successfully recognize new categories while retaining the ability to recognize old categories. However, when the features of the samples are quite different, e.g., the features of the dermatological images and clinical images in this paper, the gradient in the memory changes substantially. This indicates that there is a large change in the new category compared to the old categories. Inspired by the GEM algorithm and by incorporating the characteristics of the FSL framework, the Q-GEM method is proposed to tackle this problem. The Q-GEM method allocates a queue with a fixed length as the memory space for training tasks. When it is necessary to check the change in the gradient of the loss function, the proposed method only needs to calculate the angle between the average gradient vector in the current queue and the gradient of the current task. Since the gradient vector is constantly leaving and entering the queue, the network can maintain continuous chain updates. Moreover, this process ensures that the gradient of the current task is updated in the same direction as the gradient of the task in the queue, which ensures that the gradient changes smoothly during the gradient change process.

We assume that the transfer of knowledge is a gradual process. However, the gradient update process of the new task needs to remain consistent with the updated direction within the previous visible distance. This ensures that the gradient update process of the whole network remains continuous, i.e., the learning curve remains flat without drastic changes. In practice, this represents the gradual transfer of knowledge and a reduction in catastrophic forgetting exhibited by the network. As shown in Fig. 5, the length of the memory gradient queue is Q_l, its value is 5 in this paper, and the initial gradient queue can be expressed as $\{M_{t-4}, M_{t-3}, M_{t-2}, M_{t-1}, M_t\}$. According to the rules of the queue operation, when a new gradient M_{t+1} enters the queue, the last gradient M_{t-4} exits the queue.

The proposed method greatly reduces the memory and calculation resources required by the whole calculation process. It also reduces the uncertainty of A-GEM, in which randomly selected samples represent the average gradients of previous tasks. Accordingly, the following new constraints are proposed:

$$L\left(f_\theta^{t-1}, M\right) \leq L(f_\theta((x_i, k), y_i), M), \quad where\ M = U_{k<q} M_k \tag{13}$$

The Q-GEM method should set a fixed length for queue Q. The queue is initialized to 0, and the size of the queue can be kept consistent with the number of support samples in FSL (i.e., k = 1, 5, 10, 20...). The corresponding optimization problem is adjusted as follows:

$$min \frac{1}{2}\|g - \tilde{g}\|_2^2 \quad s.t.\ \tilde{g} g_{\bar{q}} \geq 0 \tag{14}$$

Specifically, $g_{\bar{q}} = -(g_1, \ldots, g_q)$ is calculated in each gradient update step of the training process. However, since a relatively small number of tasks is to be recorded, the number of calculations is greatly reduced. Therefore, the optimization problem for the above constraint equations can be quickly solved. When the gradient g violates a constraint, it can be solved by the following equation:

$$\tilde{g} = g - \frac{g^T g^Q}{g_Q^T g^Q} g_Q \tag{15}$$

The proof process is as follows:

$$min \frac{1}{2}\|g - \tilde{g}\|_2^2 \quad s.t.\ \tilde{g}^T g_{\overline{Q}} \geq 0 \tag{16}$$

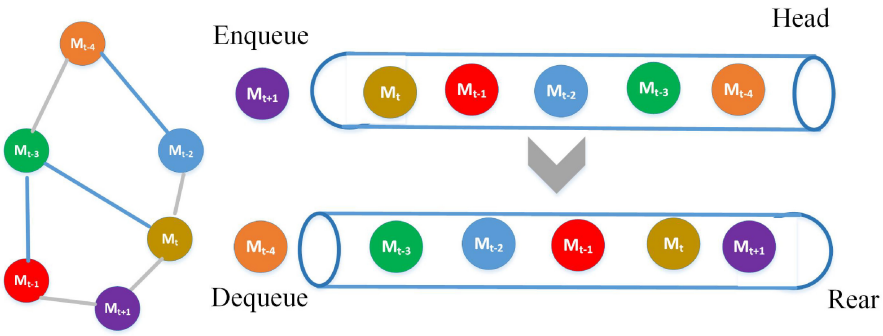

Fig. 5. Schematic diagram of the gradient queue update process.

The optimization objective $\tilde{g}$ is replaced by z, and the above formula can be simplified as:

$$min_z \frac{1}{2} Z^T Z - g^T Z \quad s.t. - Z^T g_{\overline{Q}} \leq 0 \tag{17}$$

In particular, this process discards $g^T g$ from the target and changes the signs of the inequality constraints. The Lagrangian equation of the constrained optimization problem defined above can be written as:

$$\mathcal{L}(z, \alpha) = \frac{1}{2} Z^T Z - g^T z - \alpha Z^T g_Q \tag{18}$$

The value z^* is found to obtain the minimized value of $\mathcal{L}(z, \alpha)$ by calculating its derivative:

$$\begin{aligned} \nabla_z \mathcal{L}(z, \alpha) &= 0 \\ z^* &= g + \alpha g_Q \end{aligned} \tag{19}$$

After using z^*, the dual formula can be simplified as:

$$\begin{aligned} \theta_D(\alpha) &= \frac{1}{2}\left(g^T g + 2\alpha g^T g_Q + \alpha^2 g_Q^T g_Q\right) \\ &\quad -g^T g - 2\alpha g^T g_Q - \alpha^2 g_Q^T g_Q \\ &= -\frac{1}{2} g^T g - \alpha g^T g_Q - -\frac{1}{2} \alpha^2 g_Q^T g_Q \end{aligned} \tag{20}$$

Then, the solution of the dual equation can be obtained as $\alpha^* = max_{\alpha>0} \theta_D(\alpha)$:

$$\begin{aligned} \nabla_\alpha \theta_D(\alpha) &= 0 \\ \alpha^* &= -\frac{g^T g_Q}{g_Q^T g_Q} \end{aligned} \tag{21}$$

When α^* is input into Equation (19), Q-GEM can be simplified as:

$$z^* = g - \frac{g^T g_Q}{g_Q^T g_Q} g_Q = \tilde{g} \tag{22}$$

E. Differential Diagnosis of Skin Diseases

The proposed FSCIL network is trained based on the metalearning framework. When the training process is finished, it not only is generalized to new skin disease categories but can also recognize old skin disease categories. In practical applications, doctors want to obtain high-accuracy and valuable suggestions from the CAD system. Therefore, the DDM is adopted to improve the diagnostic accuracy of the proposed method to the level of professional doctors. The FSCIL skin disease diagnosis method based on differential diagnosis has the following advantages. First, the system provides differential diagnoses of 95 diseases instead of separate classifications for a few diseases. It includes various skin diseases to help with clinical decision-making, e.g., pigmentary diseases, alopecia, and lesions. During training, the summary ranking list of diagnoses provided by dermatologists has relevant summary "confidence" scores for each diagnosis, which are the "soft" labels of classification. Second, the network supports both dermoscopic images and clinical images to evaluate their effects. The different skin disease image modes used in the proposed method have two benefits. On the one hand, they mitigate the sample shortages observed for new categories; on the other hand, the system accepts two modes of input samples, which improves the flexibility of the network. Third, we compare the diagnostic accuracy of the proposed method with that of three levels of certified doctors: (1) a deputy chief doctor of dermatology; (2) an attending dermatologist; and (3) a general practitioner. The results confirm the practical value of the proposed method.

Moreover, for each skin disease category, two different image modes can be simultaneously input into the FSCIL network as the support and query images. The dual-flow network is used to align the sample features of the two modes in the middle of the high-dimensional space. The FSCIL network classifies the input skin disease images, and the output contains the relative probability of each skin disease category. These conclusions are based on fine granularity, which can assist dermatologists in making disease diagnoses and forming treatment plans for the next step. The labelled samples used for training and testing the proposed method are provided by dermatologists with physician qualifications. For every sample, each doctor provides their top-3 diagnosis results, and multiple diagnosis results are collected to generate a diagnosis ranking table. Similarly, during the process of network validation, the FSL network provides a possible category score for each category. The diagnostic scores provided by the dermatologists are compared with the top 3 diagnostic scores generated by the network. In this way, the training process of the network becomes a differentiated process rather than a single predictive output, which is more in line with the actual diagnosis and treatment process. It also provides more references to doctors to effectively reduce misdiagnoses.

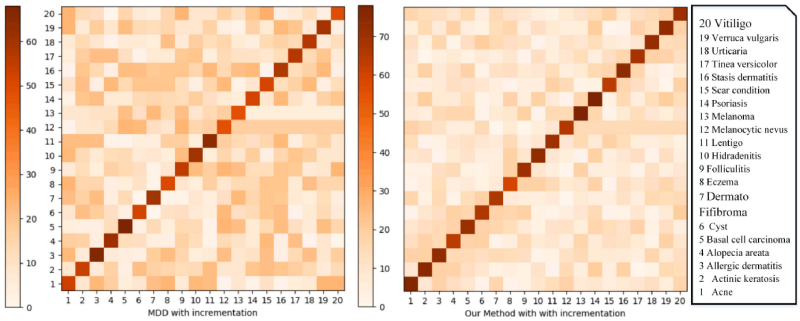

Fig. 6. Confusion matrix in the setting of incremental samples.

IV. Experiment

A. Dataset

At present, no special skin disease dataset is available for training and testing FSCIL approaches. To verify the effectiveness of the proposed method, the cate-ISIC-3^i incremental learning dataset is designed in this paper. The dataset contains 7,508 images of 95 common skin disease categories, including dermoscopic images and clinical images. Both types include deadly skin cancer categories, e.g., melanoma and basal cells, and some common skin disease categories, e.g., atopic dermatitis and seborrheic dermatitis. Two sources of samples are included in the dataset; the first source includes images selected from the current public datasets, including the ISIC-2018, ISIC-2019, ISIC-2020, SD-198, SD-256, PAD-OFED-20, Asan, and Hally datasets. The second source includes the collected and labelled samples provided by the dermatology department of the Henan Traditional Chinese Medicine Hospital. The devices for collecting dermoscopic images are various CIoT devices, such as the Dermat-1000, the Canon EOS 700d, and Glory S60 mobile phones. As shown in Fig. 6, the clinical images and dermoscopic images of the same pigmented nevus have very different appearances, so they can be considered images of two different modes.

According to the training rules of FSL, the images in the 95 categories are divided into basic training sets D_{base} and incremental learning training sets D_{novel}. In particular, D_{base} contains 8 kinds of skin disease images, and each category has many samples; i.e., each category contains approximately 50-400 images. The remaining incremental learning training sets are relatively small, and each category contains only approximately 1-10 dermatological images. Data enhancement methods, including random flipping, rotation, cropping, and colour perturbation, are used to improve the generalization ability of the model.

B. Evaluation Method

The accuracy, sensitivity, and specificity metrics are used to evaluate the effect of the proposed method:

$$Accuracy_i(ACC) = (TN_i + TP_i)/ (TN_i + FP_i + FN_i + TP_i) \quad (23)$$

$$\text{Sensitivity }_i(SEN) = TP_i/(FN_i + TP_i) \quad (24)$$

where TP_i denotes the number of positive cases in category i, FN_i denotes the number of negative cases in category i, TN_i denotes the number of positive cases determined as negative cases, and FN_i denotes the number of negative cases determined to be positive. Accuracy (i.e., efficiency) is expressed as the percentage of the sum of true-positive and true-negative cases out of the total number of subjects. Sensitivity (i.e., the true-positive rate) is the proportion of positive samples that are judged to be positive, i.e., the probability of patients being correctly judged as positive. Specificity refers to the percentage of normal people and patients with unrelated diseases whose diagnosis results are negative. The differential principle and the top-three hit rate (top-3 accuracy) are used to evaluate the effectiveness of the proposed method.

C. Implementation

ResNet-18 is used as the basic network to reduce overfitting since it has fewer parameters than other networks. The network trains for 5,000 epochs on the basic training set, stochastic gradient descent (SGD) with momentum is used as an optimizer, the initial learning rate is set to le-3, and attenuation is performed at a rate of 0.1 in rounds 1000/2000 and 3000/4000.

The proposed network is implemented based on the PyTorch framework. The main parameters of the implementation platform are as follows: the CPU is an i7-8700 CPU @ 3.7 GHz, the memory size is 32 GB; the GPU is an NVIDIA RTX2080Ti with 11 GB of memory, and the batch size is 8.

D. Result

1) Comparison With Current Advanced Methods: To the best of our knowledge, no researcher has studied the automatic classification of skin disease images with an FSCIL method. Therefore, the proposed method FCILOMI is compared with two other methods from relevant research. One is a skin disease classification method with FSL, which is used to verify whether the proposed method can reach the best level achieved by the current algorithm without setting incremental samples, including methods PCN [28], MDD [6], FSL-SL [7]. The other is used to verify whether incremental learning can be effectively implemented to reduce forgetting when incremental data is added, including methods EEIL [13] and TOPTIC [12].

Table I is a comparison of the proposed method with three FSL-based skin disease classification methods without incremental categories. Under the 5-way, 5-shot setting, the

TABLE I
CLASSIFICATION RESULTS WITHOUT INCREMENTAL CATEGORIES (JACCARD (%) STANDARD DEVIATION)

Method	5-Way 1-Shot		5-Way 5-Shot		5-Way 20-Shot	
	ACC	SEN	ACC	SEN	ACC	SEN
PCN[28]	51±0.8	42.6±0.3	62.9±0.5	47.8±1.0	65.2±0.6	50.9±0.7
MDD [6]	64.3±0.4	60.8±0.3	70.8±1.3	65.8±0.6	72.5±1.2	68.5±0.5
FSL-SL[7]	60.2±0.6	59.32±0.8	69.3±1.0	62.2±0.2	65.8±0.4	63.5±0.7
FCILOMI	**68.4**±0.7	**62.2**±0.2	**75.0**±1.5	**68.5**±0.3	**75.3**±0.8	**70.4**±0.6

TABLE II
CLASSIFICATION RESULTS WITH INCREMENTAL CATEGORIES (JACCARD (%) STANDARD DEVIATION)

Method	5-Way 1-Shot		5-Way 5-Shot		5-Way 20-Shot	
	ACC	SEN	ACC	SEN	ACC	SEN
MDD [6]	48.3±0.4	32.6±0.2	54.6±2.0	38.5±0.4	52±0.3	42.2±0.8
EEIL [13]	50.8±0.6	49.6±0.7	52.5±0.6	50.8±0.2	58.0±1.0	55.4±0.6
TOPIC [12]	52.5±0.5	53.8±0.4	62.0±2.3	58.5±0.2	62.5±0.3	61.3±0.6
FCILOMI	**62±0.7**	**60.2±0.2**	**73.2±1.0**	**67.5±0.2**	**70.5±0.4**	**68.6±0.7**

proposed method obtains better classification results than the comparison methods in terms of average accuracy and sensitivity. Specifically, the average accuracy rate is 4.2% higher than that of the MDD method, which performs best among the three alternate methods, and the average sensitivity is 2.7% higher than that of MDD. The main reason for this is that multimodal samples are used and aligned in the high-dimensional space to prevent images of the same category with large appearance differences from causing classification errors. For example, acne induces large appearance changes on the face, arms, and legs, and the modal alignment module effectively alleviates misclassification in this situation.

Table II displays the comparison results of the proposed method and the other two FSL skin disease classification methods, which are all set with incremental categories. In particular, EEIL is a benchmark algorithm for medical image classification that utilizes incremental learning and is the most advanced incremental learning algorithm for medical image classification. However, EEIL does not consider FSL learning scenarios; i.e., it has a poor effect in cases with limited labelled samples. TOPIC is the most advanced FSCIL algorithm thus far and has not been applied in medical scenes.

The top-1 classification results under the 5-way 5-shot setting are all counted, and our method achieves better results than the comparison methods in terms of average accuracy and sensitivity. The average accuracy rate is 18.6% higher than that of the MDD method, and the average sensitivity is 29.0% higher than that of the MDD method without considering incremental learning. The performance of the MDD method without considering incremental learning changes the most. The average accuracy change is 20.2%, and the sensitivity change is 7.6%. The TOPIC method considers incremental categories, providing it with obvious advantages over the MDD method. The average accuracy rate is 7.4% higher than that of the MDD method, and the average sensitivity is 20% higher than that of the MDD method. The TOPIC method considers the case of incremental categories and has significant advantages over the MDD method, and the proposed method achieves the best performance. The average accuracy of the proposed method is 11.2% higher than that of the TOPIC method, and the average sensitivity is 9% higher than that of the TOPIC method. The proposed approach benefits from two aspects. On the one hand, the multimodal problem is considered, which alleviates the classification difficulty caused by the large differences in image classes. On the other hand, the gradient queue function effectively retains the prior knowledge, making the gradient changes gentler, enabling knowledge retention, and achieving the role of knowledge chain propagation.

To further verify the ability of the proposed method when adding incremental samples in a more real scenario, we randomly selected 200 images of 20 categories to form a verification group, and the results are shown in the confusion matrix in Figure 6. The darker the diagonal colour, the higher the classification accuracy of the method. The darker the colour of other parts, the greater the classification error. Figure 6 shows the confusion matrix of the comparison method MDD and the proposed method with the 20-way 5-Shot setting with incremental samples. The results show that when adding incremental samples, the classification accuracy of the comparison method is significantly reduced, the classification is unstable with large errors, and our method obtains more satisfactory results.

2) Comparison With Clinicians: To verify the feasibility of the clinical application of the proposed method, the results of the proposed method are not only compared with the results of the advanced FSL classification methods but also compared with the diagnoses of clinicians. The experiment shows that the performance of the FSCIL method proposed in this paper is satisfactory. First, the proposed method is verified on verification group A, which is a reserved testing set that does not overlap with the training set. Specifically, on validation group A, the average top-1 accuracy is 68%, and the average top-3 accuracy rises to 91%, which can provide valuable suggestions for doctors to make clinical decisions. Second, the

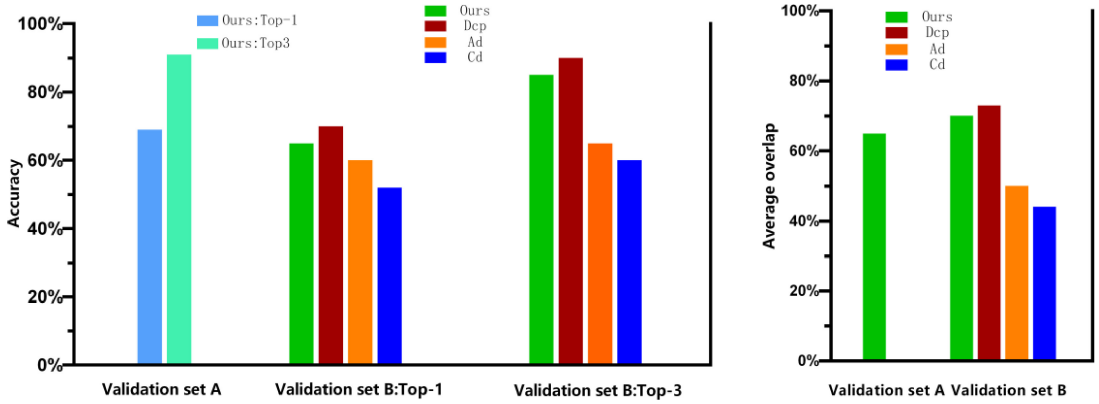

Fig. 7. Performance of our method compared with that of a deputy chief dermatologist (Dcp), attending physician (Ad) and general practitioner (Gp).

proposed method is verified on verification group B, which contains 600 images randomly sampled from validation set A by stratified sampling according to different categories. Validation set B includes the base categories, the test categories, and new categories obtained from dermatology clinics, which focus on rare cases. Fig. 7 shows the comparison between the proposed method and the doctors' diagnoses. Five dermatologists (including one deputy chief dermatologist, two attending physicians, and two community doctors from community hospitals) received different levels of training, and they participated in the verification process. The top-1 accuracy achieved on verification set B by the proposed method is 65%, that of the deputy chief dermatologist is 70%, that of the attending doctors is 60% and that of the general practitioners is 52%. We find that when the number of categories increases, especially when rare cases are included, the diagnostic accuracies of the proposed method and experienced doctors decrease. However, the proposed method does not surpass the deputy chief physician (by 5%), but its accuracy is higher than that of the attending physician and clinician. On validation set B, the top-3 accuracies of the proposed method and the deputy chief doctor both significantly improve. The accuracy of the proposed method is 20% higher than the top-1 accuracy, which explains the importance of differential diagnosis. However, the method exhibits a 5% disadvantage compared to the deputy chief physician. We also found that the diagnostic accuracies of the attending doctors and general practitioners were not significantly improved. The reason for this may be that there are many kinds of skin diseases, and some of them are only slightly different, which makes it difficult to form accurate judgements. The proposed algorithm can distinguish higher-dimensional differences between different diseases. Although it is difficult to distinguish features with high similarity, the model can also narrow them down to a smaller category range.

As shown in Fig. 8, the top-1 accuracies of the proposed method and the clinician are compared on validation set B. The results are stratified according to the accuracy confidence level. The slope of the diagonal is 1, and an arrow pointing to the left indicates that the proposed method performs better than the clinician, and vice versa. The statistical chart shows that the results of the proposed method are superior to the diagnoses of the ordinary doctors in many cases, and

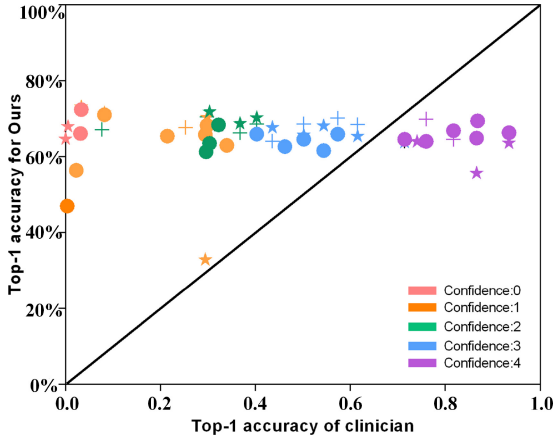

Fig. 8. Accuracy of the proposed method compared with the clinicians' accuracy on verification set B with top-1 accuracy.

the proposed method can be used as a reference for doctors. In particular, since experienced doctors are valuable medical resources, only one deputy director doctor is involved in the comparison in this chapter, and the rest are attending doctors and general practitioners (two each). Although the accuracy has been improved significantly, there is still a significant gap compared to experienced experts, especially for these different skin disease categories with similar characteristics. Therefore, the proposed method can provide auxiliary advice for primary doctors, such as attending doctors and general practitioners.

E. Ablation

Ablation experiments are carried out to verify the novel design of the proposed method, which includes the Q-GEM model, dual-stream aligned network, and DDM. The function of the gradient queue model is to realize FSCIL. The purpose of the dual-stream alignment network is to use dermoscopic images and clinical images to increase the number of images of rare cases so that the algorithm can be generalized to more categories. The results of the ablation experiment are summarized in Table III; all results are obtained under the 5-way, 5-shot setting on validation set B while considering the incremental learning situation. The first line corresponds to the use of the gradient queue function. The top-1 and top-3

TABLE III
RESULTS OF THE ABLATION TEST ON THE CATE-ISIC-3^i DATASET (JACCARD (%) STANDARD DEVIATION)

Q-GEM	SAN	Derm	Clc	Top-1	Top-3
✓	×	✓	×	58±0.4	78±0.2
✓	×	×	✓	56±0.7	75±0.5
✓	×	✓	✓	42±1.2	58±0.7
×	✓	✓	✓	45±0.8	63±1.6
✓	✓	✓	✓	72±0.5	85±0.4

accuracies are satisfactory since the gradient queue function realizes knowledge retention. The second line involves the use of the gradient queue function to perform testing on clinical images; the accuracies are decreased, but good results are also achieved since the clinical images are more affected by light and background, so they have greater differences within the same category. The third line uses dermoscopic images and clinical images simultaneously, and the accuracy rate decreases significantly since the shapes of the two images vary greatly within the same category. In the fourth line, the mode alignment module is added, and the memory-GEM module is removed, which leads to reduced accuracy. The gradient queue function and dual-stream network have great impacts on the resulting classification accuracy. In the last line, all modules are added and verified on the two-mode data, and the best results are obtained.

V. CONCLUSION

Excellent doctors can continuously increase their experience and improve their diagnosis levels through continuous learning. However, CAD systems based on deep learning exhibit catastrophic forgetting when encountering new disease categories, causing them to seriously deviate from the actual needs of clinical diagnosis. Therefore, the FSCIL method is designed in this paper to alleviate the catastrophic forgetting problem in skin disease image classification. First, a model of clinical and dermoscopic images is aligned by an alignment module to train the network on images with more categories. Then, the Q-GEM algorithm based on metalearning is proposed to ensure that the network can quickly generalize to new categories without inducing serious catastrophic forgetting. Finally, the differential classification method is used to replace the single classification method, which effectively reduces the probability of misdiagnosis. The experimental results show that the top-3 skin disease classification results of the proposed method provide effective auxiliary diagnosis opinions for dermatologists.

REFERENCES

[1] S. He, P. E. Grant, and Y. Ou, "Global-local transformer for brain age estimation," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 213–224, Jan. 2022.

[2] Y. Liu et al., "A deep learning system for differential diagnosis of skin diseases," *Nat. Med.*, vol. 26, no. 6, pp. 900–908, 2020.

[3] S. Kong, W. Wang, X. Feng, and X. Jia, "Deep RED unfolding network for image restoration," *IEEE Trans. Image Process.*, vol. 31, no. 5, pp. 31–44, 2022.

[4] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[5] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2482–2493, Jul. 2020.

[6] K. Mahajan, M. Sharma, and L. Vig, "Meta-dermDiagnosis: Few-shot skin disease identification using meta-learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 730–731.

[7] X.-J. Liu, K.-L. Li, H.-Y. Luan, W.-H. Wang, and Z.-Y. Chen, "Few-shot learning for skin lesion image classification," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 4979–4990, 2022.

[8] A. Haque, A. Milstein, and L. Fei-Fei, "Illuminating the dark spaces of healthcare with ambient intelligence," *Nature*, vol. 585, no. 8042, pp. 193–102, 2020.

[9] G. Han, L. Wenyu, and S. Renjun, "Research on methods to overcome catastrophic forgetting in small sample learning," *Comput. Appl. Softw.*, vol. 37, no. 9, pp. 1–7, 2020.

[10] D. Kiranmayi and A. Sharma, "Mobile apps and Internet of Things (IoT): A promising future for Indian fisheries and aquaculture sector," *J. Entomol. Zool. Stud.*, vol. 8, no. 1, pp. 1659–1669, 2020.

[11] D. Carter, J. Kolencik, and J. Cug, "Smart Internet of Things-enabled mobile-based health monitoring systems and medical big data in covid-19 telemedicine," *Amer. J. Med. Res.*, vol. 8, no. 1, pp. 20–29, 2021.

[12] X. Tao, X. Hong, X. Chang, S. Dong, X. Wei, and Y. Gong, "Few-shot class-incremental learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12183–12192.

[13] M. M. Derakhshani et al., "Lifelonger: A benchmark for continual disease classification," 2022, *arXiv:2204.05737*.

[14] Z. Chen and B. Liu, *Lifelong Machine Learning* (Synthesis Lectures on Artificial Intelligence and Machine Learning), vol. 12. Cham, Switzerland: Springer, 2018, pp. 1–207.

[15] D. L. Silver and R. E. Mercer, "The task rehearsal method of life-long learning: Overcoming impoverished data," in *Proc. Conf. Can. Soc. Comput. Stud. Intell.*, 2002, pp. 90–101.

[16] A. Chaudhry, M. A. Ranzato, M. Rohrbach, and M. Elhoseiny, "Efficient lifelong learning with A-GEM," 2018, *arXiv:1812.00420*.

[17] H. Gao, J. Xiao, Y. Yin, T. Liu, and J. Shi, "A mutually supervised graph attention network for few-shot segmentation: The perspective of fully utilizing limited samples," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 14, 2022, doi: 10.1109/TNNLS.2022.3155486.

[18] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," *Proc. Nat. Acad. Sci.*, vol. 114, no. 13, pp. 3521–3526, 2017.

[19] D. Lopez-Paz and M. A. Ranzato, "Gradient episodic memory for continual learning," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6467–6476.

[20] R. M. French, "Pseudo-recurrent connectionist networks: An approach to the 'sensitivity-stability' dilemma," *Connec. Sci.*, vol. 9, no. 4, pp. 353–380, 1997.

[21] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019.

[22] E. Schwartz, L. Karlinsky, R. Feris, R. Giryes, and A. Bronstein, "Baby steps towards few-shot learning with multiple semantics," *Pattern Recognit. Lett.*, vol. 160, pp. 142–147, Aug. 2022.

[23] Y. Lu et al., "Cross-modality person re-identification with shared-specific feature transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 13379–13389.

[24] M. M. Derakhshani et al., "LifeLonger: A benchmark for continual disease classification," in *Proc. MICCAI*, 2018, pp. 314–324.

[25] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," *Dermoscopy Image Analysis*, vol. 10. Boca Raton, FL, USA: CRC Press, 2015, pp. 97–129.
[26] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review," *Biomed. Signal Process. Control*, vol. 39, pp. 237–262, Jan. 2018.
[27] X. Yi, E. Walia, and P. Babyn, "Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by Wasserstein distance for dermoscopy image classification," 2018, *arXiv:1804.03700*.
[28] V. Prabhu, A. Kannan, M. Ravuri, M. Chaplain, D. Sontag, and X. Amatriain, "Few-shot learning for dermatological disease diagnosis," in *Proc. Mach. Learn. Healthcare Conf.*, 2019, pp. 532–552.

Junsheng Xiao received the Ph.D. degree in technology for computer applications from Shanghai University in 2023. He is currently a Lecturer of Computer Science with the Henan University of Chinese Medicine. He has publications in IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *ACM Transactions on Multimedia Computing, Communications, and Applications*, *Computing*, and *Wireless Networks*. His research interests include computer vision, intelligent medical image processing, AI interpretability, and credibility analysis.

Jirui Li received the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications in 2019. She is currently an Associate Professor of Computer Science with the Henan University of Chinese Medicine. Her current research interests mainly include mobile cloud computing, distributed computing, trusted services, and Internet of Things.

Honghao Gao (Senior Member, IEEE) is currently with the School of Computer Engineering and Science, Shanghai University, China. He is also a Professor with the College of Future Industry, Gachon University, South Korea. Prior to that, he was a Research Fellow with the Software Engineering Information Technology Institute, Central Michigan University, USA, and was an Adjunct Professor with Hangzhou Dianzi University, China. He has publications in IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON SERVICES COMPUTING, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, and IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS. His research interests include software intelligence, cloud/edge computing, and AI4Healthcare. He was the 2022 recipient of the Highly Cited Chinese Researchers by Elsevier, and the 2021 recipient of the IEEE Outstanding Paper Award for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.