IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 1, 2021, accepted February 14, 2021, date of publication February 24, 2021, date of current version March 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061690

Hybrid Graph Convolutional Network for Semi-Supervised Retinal Image Classification

GUANGHUA ZHANG[1], (Member, IEEE), JING PAN[2], ZHAOXIA ZHANG[3], HENG ZHANG[4], CHANGYUAN XING[5], BIN SUN[3], AND MING LI[6]

[1]Department of Computer Science and Engineering, Taiyuan University, Taiyuan 030000, China
[2]Department of Physics and Chemistry, Taiyuan University, Taiyuan 030000, China
[3]Shanxi Eye Hospital, Taiyuan 030002, China
[4]School of Mathematics and Statistics, Southwest University, Chongqing 400000, China
[5]College of Big Data and Intelligent Engineering, Yangtze Normal University, Chongqing 408100, China
[6]College of Data Science, Taiyuan University of Technology, Taiyuan 030024, China

Corresponding authors: Bin Sun (sunbineye@163.com) and Ming Li (lm13653600949@126.com)

This work was supported in part by the Shanxi Scholarship Council of China under Grant 2020-149, in part by the Research Funds of Health Commission of Shanxi Province under Grant 2018090, in part by the Program of Natural Science Foundation of Chongqing Science and Technology Bureau under Grant cstc2019jcyj-msxmX0873, and in part by the Key Research and Development Program of Shanxi Province under Grant 201903D311009.

ABSTRACT Diabetic Retinopathy (DR) causes a significant health threat to the patient's vision with diabetic disease, which may result in blindness in severe situations. Various automatic DR diagnosis models have been proposed along with the development of deep learning, while there always relies on a large scale annotated data to train the network. However, annotating medical fundus images is cost-expensive and requires well-trained professional doctors to identity the DR grades. To overcome this drawback, this paper focuses on utilizing the easily-obtained unlabeled data with the help of limited annotated data to identify DR grades accurately. Hence we proposes a semi-supervised retinal image classification method by a Hybrid Graph Convolutional Network (HGCN). This HGCN network designs a modularity-based graph learning module and integrates Convolutional Neural Network (CNN) features into the graph representation by graph convolutional network. The synthesized hybrid features are optimized by a semi-supervised classification task which is assisted by a similarity-based pseudo label estimator. Through the proposed HGCN method, the retinal image classification model can be trained efficiently by partially labeled samples and the complicated annotating work is not required for the most retinal images. The experimental results on MESSIDOR dataset demonstrate the favorable performance of HGCN on semi-supervised retinal image classification, and the fully labeled data training also achieves an obvious superiority to the state-of-the-art supervised learning methods.

INDEX TERMS Retinal image classification, semi-supervised, graph convolutional network, modularity-based graph learning.

I. INTRODUCTION

Diabetic Retinopathy (DR) can give rise to evitable blindness for diabetic patients in the whole world. The diabetes has attacked around 210 million humans [29], and at least 10% of them have deteriorated into DR [25], [35]. In future, the number of diabetic patients will be increased to 360 million by 2030 [36], that indicates DR will become a severe health issue in the next decade. In clinical symptom, the main reason of diabetes is the increasing of blood glucose level, which appearing long-term can injury the vessels in retina. The histopathological retinal delicate damage could cause the visual loss or permanent blindness when it is untreated in early stage, which makes most adults in developed areas exposed in the threaten of blindness. That makes the DR become one of the most complicated diseases in diabetic patients [48]. As for preserving the patient's vision, the most effective way is the early diagnosis and prompt treatment. Therefore, an efficient evaluating protocol of distinguishing retinopathy level in visual impairment is a significant requirement to avoid the permanent retinal deterioration.

As a major diagnosing evidence for DR, colorful fundus images are often utilized by ophthalmologists in clinical symptom analysis [8], [38]. Several hand-crafted image

The associate editor coordinating the review of this manuscript and approving it for publication was Kang Li.

features are employed in most early researches on the identification of DR grades, that makes the automatic DR grade classification can be achievable. For instance, Mookiah *et al.* [24] and Mansour [23] conducted DR grade classification by the employed methods, including mathematical morphology characteristics, damaging area tracking, thresholding and transformation model, clustering and matching models, and synthetic methods. Faust *et al.* [6] summarized the DR-related methods about learning DR grade's features from fundus, containing the blood vessel area, exudes, hemorrhages, microaneurysms and texture. Joshi and Karule [14] concluded the studies about detecting exudate from early proposed. Thakur and Juneja [34] introduced related researches on optic disc segmentation and the classification of galucoma. Nevertheless, designing robust hand-crafted features requires well-trained professional knowledge to choose the applicable characteristics for fundus images by exploitation on different scheme and complicated parameter configurations, that brings burdensome challenges in the model generalizing.

Over the years, the size of image datasets and computing power of GPUs are increasingly developed. That accelerates the innovative evolution of deep learning technology, which has demonstrated the prominent achievement in widely-applied technologies, including computer vision, nature language processing and data mining [40], [42]. From the successful applications, deep learning shows obvious superiorities over conventional hand-crafted-feature-based models [3], [43]–[45]. Several Deep learning approaches [12], [19], [39] aim to improve the identification accuracy on the fundus images to diagnose DR, and design various models to support the application in computer-aided diagnosis. For instance, Li *et al.* [19] developed a deep network (OCT_Net) to identity the early-grade in diabetic retinopathy, which is in charge both of extracting robust OCT features and learning discriminative information in retinal layers.

In clinical practice, medical images are plentiful but there is a lack of sufficient annotated data because of the patient's privacy and security considerations, and professional medical image annotator's scarcity and expensive cost [4]. This limitation severely restricts the DL-based models mentioned above in practice, while semi-supervised framework is an effective way to diagnose DR when we have limited annotated data, as shown in Figure 1. Explicitly, semi-supervised learning framework that utilizes little annotated samples with abundant unannotated data offers an efficient mean to address the limitation of insufficient labeled data. Recent works [13], [17] have explored semi-supervised learning approaches with state-of-the-art performance by introducing Graph Convolutional Network (GCN) into semi-supervised image classification task. Many clinical image cases follow the practical application of GCN-based semi-supervised learning where GCN is seldom employed in analyzing retinal fundus images from diabetic patients, and the ability of GCN in semi-supervised retinal image classification would be desirable in practical scenarios.

To address semi-supervised retinal image classification problem in DR diagnosis, this paper builds a Hybrid Graph Convolutional Network (HGCN) as learning from very few labeled images with disease grading annotations, alongside a large set of unlabeled images. The main assumption of HGCN is that the retinal images in same category have stronger inherent discriminative correlation than ones in different category, which can be simulated in node-to-node Graph Structure (GS). HGCN aims to learn hybrid GS-based representations that integrates a graph learning convolutional network into deep learning feature extractor, trained by a modularity-based graph learning module and a hybrid classification module between labeled and unlabeled data. Specifically, HGCN firstly employs several CNN layers to extract retinal image features, and then attaches a graph learning layer with the modularity graph learning loss to establish the topology correlations between retinal image samples according to their similarities. Finally, the graph convolution layers are implemented in the extracted CNN features and modularity-based graph correlations, and generate final retinal image features, which are utilized into the hybrid semi-supervised classification task.

The crucial contributions of HGCN approach for semi-supervised retinal image classification are concluded below:

- We propose a semi-supervised Hybrid Graph Convolutional Network (HGCN) that combines CNN and GCN into a unified architecture to synthesize independent CNN and graph structure features for representing retinal images.
- In HGCN, we design a modularity-based Graph learning module to conduct the graph structure learning with refining the graph construction which can improve the graph learning efficiency in unlabeled data.
- HGCN can exploit more discriminative information of unlabeled data by our proposed hybrid classification module, which provides the training direction for the hybrid graph convolutional network.

To the best of our knowledge, there is a few research works on semi-supervised retinal image classification and this paper is a preliminary research on this topic with employing Graph Convolutional Network in medical images analysis. Experimental results demonstrate that HGCN can effectively solve the semi-supervised retinal image classification in disease gradings.

II. RELATED WORK

We review the related researches of retinal image classification in this section, which are partitioned by three aspects, including retina image classification, unsupervised medical image classification, and a brief introduction of graph convolutional network.

A. DIABETIC RETINOPATHY DIAGNOSIS

The permanent increasing of glucose level in diabetic patients' blood generates a threat of damage for retinal vessel

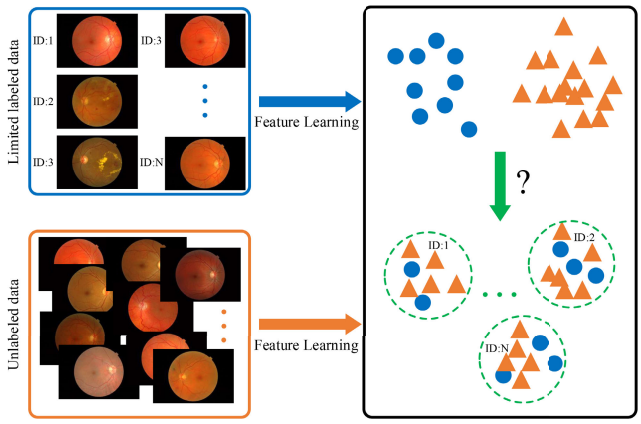

FIGURE 1. Illustration of semi-supervised retinal image classification task.

tissue, when diabetes increases the glucose level in retinal vessels. This phenomenon can break the vessel and result in the blood's divulgation in the retinal fundus, so as to damage the vision of patients. Thus, the diabetic patients must be screened regularly for the eye problems, especially for retinal check by the ophthalmologist. In the early-stage of diabetic retinopathy, the retinal fundus images play an important role in the DR diagnosis, in which morphological representation provides the identification evidence for the DR grades.

The characteristics of retinal abnormalities reveal the DR severities, which can be recognized by deep learning methods. The phases of diabetic retinopathy are with four stages, including micro aneurysms, hemorrhage, neovascularization, and venous leading. The four stages denote the ponderances of retinopathy [37]:

- **Mild non-proliferative retinopathy** (micro aneurysms) is the earliest stage, where only micro aneurysms can occur;
- **Moderate non-proliferative retinopathy** (hemorrhage) is a stage which can be described by losing the blood vessels' ability of blood transportation due to their distortion and swelling with the progress of the disease;
- **Severe non-proliferative retinopathy** (neovascularization) results in deprived blood supply to the retina due to the increased blockage of more blood vessels, hence signaling the retina for the growing of fresh blood vessels;
- **Proliferative diabetic retinopathy** (venous leading) is the advanced stage, where the growth features are secreted by the retina activate diffusion of the newly-generated retinal blood vessels, growing along inside covering of retinal in some vitreous gel, filling the eye.

Due to the significance of the diabetic retinopathy diagnosis, plenty researches pay attention to the automated retinal image classification [9], [12], [20], [39]. For example, Hemanth *et al.* [12] employed deep learning algorithm to integrate image operating by histogram equalization into the deep neural network and utilize the contrast limited adaptive histogram equalization to stimulate the classification of the network; Wu *et al.* [39] aimed at OCT image classification and embedded attention mechanism into deep neural network, which introduces image precessing to strengthen the image representation and the attention module to focus on the crucial area with pathological abnormal characteristics. Luo *et al.* [20] solved the retinal image classification task without labor-expensive image annotations by a self-supervised fuzzy clustering network. Gulshan *et al.* [9] conducted a prospective observational study at 2 eye care centers in India by deep learning algorithm and the results demonstrate that the feasibility of using an automated diabetic retinopathy system to expand screening programs. Though these DL-based methods have achieved impressive performance, they are all under supervised framework requiring sufficient large scale accurately annotated images when training model.

Besides automatic DR grading mentioned above, several DR detection/segmentation works [31]–[33] are proposed recently. Tavakoli *et al.* [31] compared effects of two preprocessing methods, illumination equalization and top-hat transformation on retinal images to detect microaneurysms using combination of matching based approach and deep learning methods either in the normal fundus images or in the presence of DR; Tavakoli and Nazar [32] applied three retinal vessel segmentation methods including Laplacian-of-Gaussian, Canny edge detector, and Matched filter to compare results of microaneurysms detection using combination of unsupervised and supervised learning either in the normal images or in the presence of DR; Tavakoli *et al.* [33] did microaneurysms detection step using combination of Laplacian-of-Gaussian and convolutional neural networks, and the experiments evaluate the accuracy of this work.

B. SEMI-SUPERVISED MEDICAL IMAGE CLASSIFICATION

Semi-supervised learning can exploit discriminative information in unannotated data to accelerate the supervised learning with limited labeled samples, and it can achieve robust models with insufficient annotations [15], which is more expensive in medical image annotation, and has excellent results compared to unsupervised methods [41]. To alleviate the annotation cost, several semi-supervised medical image classification models have been proposed [18], [22], [46], [49]. For example, Zhou *et al.* [49] designed a jointly training model with semi-supervised framework to implement the DR grading and damage segmentation, guided by an attention module. Xie *et al.* [46] designed an adversarial learning mechanism under semi-supervised framework to conduct CT classification, which contains an adversarial auto-encoder R to implement the unsupervised self-expression, an identification network C trained by labeled data, and several trainable transforming layers that are in charge of transferring the image representations learned by R into the identification network C. Madani *et al.* [22] utilized generative adversarial networks to leverage the imbalance of limited labeled samples and massive unannotated data under semi-supervised framework. Kumar *et al.* [18] introduced robust statistical

model to extend the multi-variable Gaussian generator into a scalable kernel Hilbert space under semi-supervised training, which can fully exploit the identical information in the limited annotated data.

According to the successful applications of above semi-supervised learning in medical image classification, this paper aims to address the scarcity of expert-labeled data problem in retinal image classification for diabetes retinopathy diagnosis. Existing semi-supervised image classification models often utilize generative adversarial learning or domain adaptation approaches, while they ignore the influence of the mutual samples (structure relationship between each retinal image samples). This paper will overcome this main drawback by the proposed hybrid graph convolutional network.

C. GRAPH CONVOLUTIONAL NETWORK

We briefly review a semi-supervised GCN framework [17] to illustrate the basic architecture of this paper. Assume $X = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^{n\times p}$ as n p-dimensional feature vectors learned by feature extractor, and denote $G(X, A)$ as the graph feature vectors of X, where the structural correlations in X are built by the pairwise similarities $A \in \mathbb{R}^{n\times n}$. The basic architecture of GCN is often composed by an input layer, a few hidden layers and a final perceptron layer [17]. For a feature vector $X^{(0)} = X$ as the input of GCN, with the correlation graph matrix A, GCN can implement the forward propagation in hidden layers by,

$$X^{(k+1)} = \sigma(D^{-1/2}AD^{-1/2}X^{(k)}W^{(k+1)}) \tag{1}$$

where $k = 0, 1, \cdots, K-1$, and $D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$ denotes a diagonal matrix which $d_i = \sum_{j=1}^{n} A_{ij}$. $W^{(k)} \in \mathbb{R}^{d_k\times d_{k+1}}$, $d_0 = p$ represents the weight matrix in convolutional layer which is required to be optimized. $\sigma(\cdot)$ is the activation mapping, like $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$, and $X^{(k+1)} \in \mathbb{R}^{n\times d_{k+1}}$ is the final graph representation extracted from k-th layer. As in semi-supervised learning framework on classification task, GCN computes the final perceptron output by,

$$Z = \mathrm{softmax}(D^{-1/2}AD^{-1/2}X^{(K-1)}W^{(K)}) \tag{2}$$

where $W^{(K)} \in \mathbb{R}^{d_K\times c}$ and c represents the amount of categories. The output of GCN $Z \in \mathbb{R}^{n\times c}$ stands for the predicted probabilities belonging to each class for the input features X. For a specific row Z_i, it presents the class estimation for i-th feature vector x_i. All the trainable weights in GCN $\{W^{(0)}, W^{(1)}, \cdots, W^{(K)}\}$ are optimized by the cross-entropy loss fixed on the GCN's output,

$$\mathcal{L}_{\text{Semi-GCN}} = -\sum_{i\in L}\sum_{j=1}^{c} Y_{ij} \ln Z_{ij} \tag{3}$$

where L represents the labeled data collection.

This framework [17] is a representative semi-supervised GCN method and many other applications based on this architecture always obtain graphs from domain knowledge or estimated by human which are generally independent of semi-supervised GCN, and ignore their structural relationships between each nodes. This paper designs a modularity-based graph learning module to address this limitation, which detail is presented in the section III.

III. PROPOSED METHOD

A. OVERVIEW

The presented method aims to classify fundus images under semi-supervised framework, to conduct diabetes retinopathy diagnosis more flexible in reality. To achieve this goal, we propose a Hybrid Graph Convolutional Network (HGCN) with three primary modules, including feature extraction module, modularity-based graph learning module, and hybrid graph convolutional module, as shown in Figure 2. First of all, the main procedure to construct graph structure is to extract feature representations from raw retinal images, where we employ Convolutional Neural Network (CNN) as the feature extraction module for retinal images. Then the learned CNN features are fed into our modularity-based graph learning module to establish the node-to-node graph structure which guarantees the adaptation on GCN process. Note that, the modularity is a quality assessing measure for graph clustering in topology analysis, which is introduced to constrain the graph learning process. Next, a hybrid GCN module is designed to learn the synthetic features combining independent CNN and mutual GCN features as the final representation of fundus images. Finally, a similarity-based pseudo label estimator assists the unlabeled features in feeding into the classification loss alongside the existing labeled images. Through this hybrid graph convolutional network, the structural influence inside retinal image samples is learned by the modularity-based graph learning and GCN process, and more discriminative information in unlabeled data is also exploited by the clustering-based pseudo label producer to support 'pseudo-supervised' learning with labeled retinal images.

B. FEATURE EXTRACTION MODULE

The essential step of retinal image classification is to extract robust feature representation from raw retinal data, which includes a few labeled and large number of unlabeled images in semi-supervised manner. We build L stacked convolutional layers to learn CNN features with the inputs of retinal images. Given a retinal image $h_c^{(0)} = x$ as the input of the first convolutional layer with weight parameters $w_c^{(1)}$ and bias of $b_c^{(1)}$, the output of this layer is denoted as $h_c^{(1)} = \sigma(w_c^{(1)}h_c^{(0)} + b_c^{(1)})$ where σ is an activation function. For the l-th convolutional layer in general form, its output is $h_c^{(l)} = \sigma(w_c^{(l)}h_c^{(l-1)} + b_c^{(l)})$. Finally, we adopt the output of the top (L-th) layer as the CNN representation h_c for input retinal image x,

$$h_c = h_c^{(L)} = \sigma(w_c^{(L)}h_c^{(L-1)} + b_c^{(L)}) \tag{4}$$

Through this feature extractor, we can obtain the CNN feature collection $H_c = \{h_c^1, \cdots, h_c^i, \cdots, h_c^N\}$ from the retinal image set $X = \{x^1, \cdots, x^i, \cdots, x^N\}$, where x_i denotes the i-th retinal image in X, and N is the number of images.

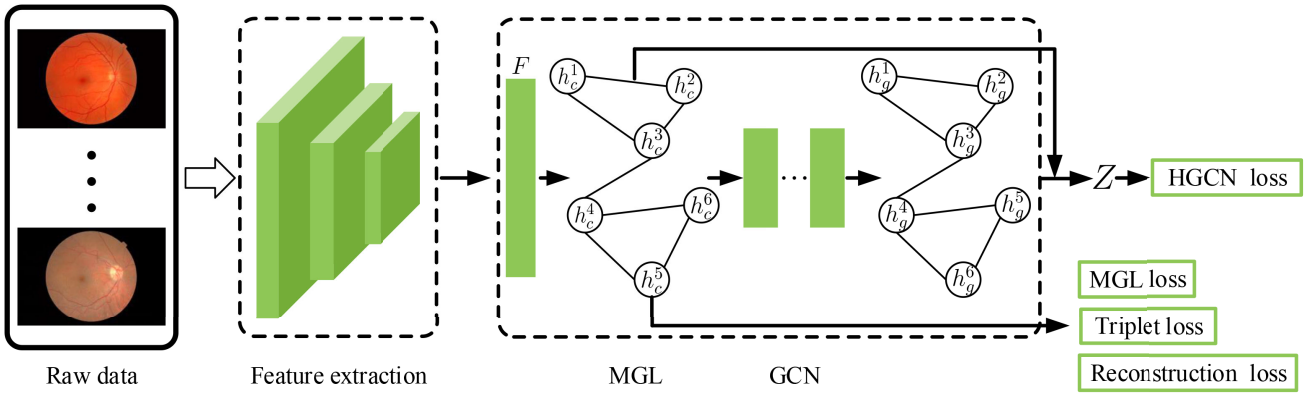

FIGURE 2. The scheme of Hybrid Graph Convolutional Network (HGCN). There contains three modules of feature extraction, modularity-based graph learning, and hybrid graph convolutional network. The DR fundus images are firstly fed into a CNN feature extractor to learn feature representations, and then learn the graph topology correlations among them by MGL. To learn graph representations, we employ GCN layers to exploit the interactive influence and obtain GCN features. Finally, CNN and GCN features are integrated into a hybrid semi-supervised classification network to grade DR fundus images.

To guarantee the representative ability of the learned CNN feature set H_c for retinal images, we attach the triplet loss [27] on the labeled retinal features with accurate annotations,

$$\mathcal{L}_t = \sum_{i=1}^{N} [\|h_c^i - h_c^p\|_2^2 - \|h_c^i - h_c^n\|_2^2 + \alpha] \tag{5}$$

where h_c^p denotes the retinal feature of a positive sample to x_i with same label, h_c^n is of negative sample with a different annotation, and α is a margin parameter to balance the loss. Though the triplet loss, the feature extractor can learn robust feature representations from labeled retinal images.

As for unlabeled retinal data, we employ an image decoder attached on the feature extractor to construct an auto-encoder architecture, which is always effective on unlabeled image feature learning. Specifically, we deploy several deconvolutional (deconv) layers on the learned retinal feature h_c and the last deconv layer outputs a reconstructed retinal image $\hat{x}$. The loss function for these deconv layers is the reconstruction loss,

$$\mathcal{L}_r = \sum_{x \in X} \|x_i - \hat{x}_i\|_2^2 = \|X - \hat{X}\|_F^2 \tag{6}$$

where the deployment of the reconstruction loss guarantees that the feature representation h_c maintains discriminative information from the input retinal image x, so as to reconstruct itself by the decoder.

Note that, the triplet loss constrains the labeled retinal images and the reconstruction loss is attached on both labeled and unlabeled images to facilitate the feature learning not only of the distance characteristic but also of self-representative information.

C. MODULARITY-BASED GRAPH LEARNING MODULE

The crucial step after extracting feature representations is to construct the graph relation $G(H_c, A)$ for retinal images X. A major problem of existing Graph learning is they are generally independent by a specific knowledge or human estimated, which inevitably introduces noises and hard relations. To conquer these blemishes, we propose a novel Modularity-based Graph Learning (MGL) approach, that effectively supports the graph convolution on learning the graph representation $G(H_c, A)$ in a unified framework. Concretely, as shown in Figure 2, the MGL can establish the structural relations among retinal features in H_c by a graph learning layer, inspired by [13]. The following description explains the detail MGL method.

For the learned CNN features $H_c = \{h_c^1, \cdots, h_c^i, \cdots, h_c^N\}$ from retinal images, MGL is in charge of learning a nonnegative mapping $S_{ij} = g(h_c^i, h_c^j)$ that indicates the similarity correlations in feature vectors h_c^i and h_c^j. The mapping $g(h_c^i, h_c^j)$ is achieved by a fully connected layer F, which is vectored though weight parameters $w^g = \{w_1^g, w_2^g, \cdots, w_p^g\}^T \in \mathbb{R}^{p \times 1}$. Mathematically, the graph structure S is calculated by,

$$S_{ij} = g(h_c^i, h_c^j) = \frac{\exp\left(\mathrm{ReLU}\left(w^{gT}\left|h_c^i - h_c^j\right|\right)\right)}{\sum_{j=1}^{N} \exp\left(\mathrm{ReLU}\left(w^{gT}\left|h_c^i - h_c^j\right|\right)\right)} \tag{7}$$

where $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$ is the activation function, and it constrain the nonnegativity of S_{ij}. This softmax transformation assures that the obtained graph S meets the characteristics,

$$\sum_{j=1}^{n} S_{ij} = 1, \quad S_{ij} \geq 0 \tag{8}$$

We update the weights w^g of GCN layer by minimizing a novel Modularity Graph Learning (MGL) loss function,

$$\mathcal{L}_{\mathrm{MGL}} = -\frac{1}{2E} \sum_{S_{ij}>0} \left(\mathbf{S}_{ij} - \frac{k_i k_j}{2E}\right) \mathbb{1}_{[c_i = c_j]} + \gamma \|S\|_F^2 \tag{9}$$

where E represents the total number of the edges in graph S; k_i and k_j denote the degree of node i, j of retinal images, respectively; c_i and c_j are the estimated clustering groups of i-th and j-th; And $\mathbb{1}_{[c_i = c_j]}$ is an indicator function, which is of 1(0) if $c_i = c_j (c_i \neq c_j)$. The remain problem is how to estimate the reasonable annotations of unlabeled retinal

images. Focusing on this problem, we introduce a Similarity-based Pseudo Label Estimator (SPLE), which provides the pseudo labels of unlabeled retinal images. Concretely, given an unlabeled retinal image x_i, SPLE calculates the similarities between x_i and labeled images by introducing Euclidean distance to obtain $[s_1, \cdots, s_{N_l}]$, where N_l is the number of labeled retinal images. After that, the SPLE can estimate the pseudo label c_i for x_i by,

$$c_i = c_{\arg\max([s_1,\cdots,s_{N_l}])} \tag{10}$$

In MGL loss, the term of $\frac{1}{2E}\sum_{i,j\in\mathscr{V}}\left(\mathbf{S}_{ij} - \frac{k_i k_j}{2E}\right)\mathbb{1}_{[c_i=c_j]}$ denotes the modularity $Q \in [0, 1]$ of the learned Graph S, which is a crucial evaluation measurement of the clustering partition ability in unlabeled data. In satisfied graph learning, it is expected that the retinal features in same class are densely connected while ones in different categories are sparse connected in our labeled and unlabeled retinal images. The modularity is often employed in graph clustering algorithm to measure the property of this partition strategy, that has been proved to be effective [2]. To this measurement, the modularity is the best choice in graph clustering view. For the learned graph S, $Q = 0$ indicates that connections in S are randomly connected without any partition, and the structural relation of clusterings shows better division by the densely connected edges along with the increase of Q, where the mixing degree between clusterings becomes smaller.

In addition, to ensure the large distance $\|h_c^i - h_c^j\|_2$ of pairwise retinal samples x_i and x_j remain a weaker relationship S_{ij}, we update the MGL loss by,

$$\mathcal{L}_{\mathrm{MGL}} = -\frac{1}{2E}\sum_{S_{ij}>0}\left(\mathbf{S}_{ij} - \frac{k_i k_j}{2E}\right)\mathbb{1}_{[c_i=c_j]} + \sum_{i,j=1}^{n}\left\|h_c^i - x_c^j\right\|_2^2 S_{ij} + \gamma\|S\|_F^2 \tag{11}$$

The correlation graph S learned by MGL maintains the expected probability characteristic (Eq. 8) that indicates the connected probability between retinal images x_i and x_j according to their similarity. In another word, our modularity-based graph learning (MGL) architecture can constitute the neighborhood relationships in retinal images not only by their similarities in H_c but also considering the modularity characteristics of each node.

D. HYBRID GRAPH CONVOLUTIONAL NETWORK

After modularity-based graph learning, we follow a Hybrid Graph Convolutional Network to learn the graph representations for semi-supervised retinal image classification problem. Figure 2 shows the overview of HGCN architecture. The goal of HGCN is to learn an optimal hybrid representation by integrating CNN and GCN based features, to boost the synthetical performance on semi-supervised retinal image classification.

As shown in Figure 2, HGCN contains one MGL layer, several graph convolutional layers and one final percetron layer. The MGL layer provides an optimal adaptive graph representation S for graph convolutional layer. That is, in graph convolutional layers, it conducts the layer-wise propagation rule based on the adaptive neighbor graph S returned by the modularity-based graph learning layer,

$$h_g^{(k+1)} = \sigma(D_s^{-1/2} S D_s^{-1/2} h_g^{(k)} w_g^{(k+1)}) \tag{12}$$

where $k = 0, 1, \cdots, K-1$, $h_g^{(k+1)}$ indicates the graph representation of $(k+1)$-th hidden layer, and the input of the first GCN layer is $h_g^0 = h_c$; $D_s = \mathrm{diag}(d_1, \cdots, d_i, \cdots, d_N)$ is a diagonal matrix with diagonal component $d_i = \sum_{j=1}^{N} S_{ij}$; $w_g^{(k)} \in \mathbb{R}^{d_k \times d_{k+1}}$ denotes the optimizable weight matrix for each graph convolutional layer; $\sigma(\cdot)$ represents an activation mapping, including $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$, and $h_g^{(k+1)} \in \mathbb{R}^{N \times d_{k+1}}$ indicates the graph representations after activation in the k-th layer. Though the obtained graph relation S conforms to $\sum_j S_{ij} = 1$, $S_{ij} \geq 0$, Eq. 12 can be redefined by,

$$h_g^{(k+1)} = \sigma(S h_g^{(k)} w_g^{(k+1)}) \tag{13}$$

Through the graph convolutional layers, we can obtain the GCN representation of each retinal image x by $h_g = h_g^{(K)} = \sigma(S h_g^{(K-1)} w_g^{(K)})$, and then we integrate this GCN feature h_g with CNN feature h_c into a final hybrid feature representation by,

$$h = concate[h_c, h_g] \tag{14}$$

where *concate* is a concatenating operation on h_c and h_g. Through the hybrid graph convolutional network, we can obtain the final retinal image representations $H = \{h^1, \cdots, h^i, \cdots, h^N\}$

Algorithm 1 Hybrid Graph Convolutional Network

Initialization: The trainable parameters in feature extraction module, trainable parameters of graph learning layers, and trainable parameters in graph convolutional network; $\alpha = 5$, $\gamma = 0.1$, $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = 0.8$.
Input: Annotated retinal images X_l and unlabeled images X_u ($X = X_l \cup X_u$).
1. CNN feature extraction:
While x_i in X **do**
  Train the CNN parameters by Eq.6.
  If $x_i \in X_l$:
    Train the CNN parameters by Eq.5.
Obtain the CNN features $H_c = \{h_c^1, \cdots, h_c^i, \cdots, h_c^N\}$ from images X.
2. Modularity-based graph learning:
While x_i, x_j in X **do**
  Train the parameters in graph learning layers by Eq.11.
3. Graph convolutional network:
  Train the GCN parameters by Eq.16.
Return HGCN.

E. SEMI-SUPERVISED CLASSIFICATION

In semi-supervised retinal image classification task, the overall retinal images are denoted by $X = [X_l, X_u]$, in where $X_l = \{x_l^1, \cdots, x_l^{N_l}\}$ denotes the labeled retinal data with their labels $Y_l = \{y_l^1, \cdots, y_l^{N_l}\}$, and $X_u = \{x_u^1, \cdots, x_u^{N_u}\}$ is the unlabeled retinal data without any label. In this paper, we can obtain the pseudo annotations $\hat{Y}_u = \{\hat{y}_u^1, \cdots, \hat{y}_u^{N_u}\}$ by the Similarity-based Pseudo Label Estimation (SPLE) module for X_u.

Based on the retinal image annotations of Y_l and $\hat{Y}_u$, we attach a percetron layer on the hybrid features H as,

$$Z = \text{softmax}(H) \tag{15}$$

where $Z \in \mathbb{R}^{n \times C}$ and C is the number of retinal image categories. The final output Z represents the label prediction of HGCN network, in which each row Z_i denotes the label prediction for the i-th retinal image. For training the classification model, we employ the cross-entropy loss both on labeled and unlabeled retinal images by,

$$\mathcal{L}_{\text{HGCN}} = -\sum_{x_i \in X_l} Y_i \ln Z_i - \sum_{x_j \in X_u} \hat{Y}_j \ln Z_j \tag{16}$$

The trainable parameters $W = \{w_c^1, \cdots, w_c^L, w_g^1, \cdots, w_g^K\}$ in our proposed HGCN are jointly optimized by,

$$\mathcal{L} = \mathcal{L}_{\text{HGCN}} + \lambda_1 \mathcal{L}_{\text{MGL}} + \lambda_2 \mathcal{L}_t + \lambda_3 \mathcal{L}_r \tag{17}$$

where $\lambda_i \| \{i = 1, 2, 3\}$ denote balance parameters for each loss functions. Besides, the whole optimization of HGCN is conducted by the back-propagation algorithm to learn the optimal weight parameters. The network optimization is summarized in Algorithm 1.

IV. EXPERIMENTS

A. DATASET AND PRE-PROCESSING

To evaluate the effectiveness and benefit of the proposed HGCN on semi-supervised retinal image classification task, we evaluate it on MESSIDOR dataset, which is a public dataset provided by the Messidor research program funded by the French ministry of research and defense [5]. It consists of 1200 retinal images and provides a retinopathy grade for each images from 0 to 3, that was acquired by three ophthalmology departments using colored video 3CCD camera mounted on a Topcon TRC NW6 non-mydriatic retinopathy with a 45° field of view, in different sizes: 1440 × 960, 2240 × 1488, or 2304 × 1536 pixels and were 8 bits per color plane.

We divide the dataset into training (50%) and testing (50%) data without any overlapped images, and the testing data contains equal number of retinal images from different grades. Then, the training images are randomly given annotations (X_l) by a pointed percentage of labeled data (lp) and the other images are partitioned into unlabeled set (X_u). Suppose lp is 20%, then 20% of the training data are employed as X_l and combine the left data x_u to build the model. Different values are set to lp in the experiments ranging from 20% to 100%. All images are resized into 224 × 224 and normalized by the maximum intensity value in each image before feeding into the HGCN network. In addition, the available code will be released at GitHub ('https://github.com/Jieming1022/HybridGCN').

B. IMPLEMENTATION OF HGCN

To achieve the hybrid graph convolutional network, we employ PyTorch framework with 2 NVIDIA Geforce 2080Ti GPUs to implement the model. The CNN feature extraction module is established by a pre-trained ResNet-50 from ImageNet [11], which contains 1 input layer (size of 224×224), 1 convolutional layer, 4 residual blocks and 1 fully connected layer (dimension of 512). In addition, we also employ the image decoder in Cycle GAN [50] to realize the reconstruction loss. Similar to [17], we set the number of graph convolution layers in our HGCN to 2. And the number of units in graph learning layer is set to 70 and it is set to 30 in graph convolutional layer. All the network parameters are initialized using Glorot initialization [7]. We train the HGCN for maximum of 500 epoches using an ADAM optimizer [16] with learning rate 0.001, which will be ×0.1 in each 10 epochs and decayed to 0 in the last 10 epochs. The margin parameter in Eq5 is $\alpha = 5$, parameter in Eq11 is set by $\gamma = 0.1$, and the balance parameters in Eq.17 are set to $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = 0.8$. Keeping consistency with [13], HGCN adopts two graph convolution layers and the number of units in graph learning layer is set to 60, while the number of units in graph convolution layer is 40.

C. MEASUREMENTS AND BASELINES

1) MEASUREMENTS

To reveal the overall effectiveness of HGCN, the proposed method is assessed with respect to accuracy, sensitivity, specificity, and F1-score. The calculation of them is summarized below:

- $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$
- $Sensitivity = \frac{TP}{TP+FN}$
- $Specificity = \frac{TN}{TN+FP}$
- $F1 - score = \frac{2 \times P \times R}{P+R}, P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}$

where TP is true positives, FP is false positives, FN is false negatives, P is precision and R is recall.

2) BASELINES

To shows the comparative performance of our HGCN model, we choose several state-of-the-arts of supervised and semi-supervised learning methods to conduct comparison, including self-training [47], co-training [1], PsoFuzzy [30], VQSSL [26], CLAHE [12], HPSCNN [10], and MAlex [28]. The detail description of them are following. The self-training [47] method is an unsupervised learning algorithm for sense

TABLE 1. The performance of HGCN among different *lp* for MESSIDOR dataset.

lp	Accuracy	Sensitivity	Specificity	F1-score
20%	0.893	0.802	0.909	0.793
40%	0.921	0.844	0.941	0.829
60%	0.948	0.906	0.959	0.892
80%	0.968	0.939	0.981	0.945
100%	**0.979**	**0.958**	**0.989**	**0.952**

TABLE 2. Comparison among the state-of-the-art baselines on MESSIDOR dataset.

Models	Source	Accuracy	Sensitivity	Specificity	F1-score
Self-training (1995) [47]	ACL	0.951	0.894	0.961	0.891
Co-training (1998) [1]	COLT	0.960	0.890	0.960	0.879
PsoFuzzy (2013) [30]	IJCA	0.945	0.910	0.980	-
GCN (2017) [17]	ICLR	0.948	0.919	0.945	0.927
VQSSL (2018) [26]	TAHC	0.975	0.946	0.982	0.942
CLAHE (2019) [12]	NCA	0.970	0.940	0.980	0.940
HPSCNN (2019) [10]	EMBC	0.939	0.888	0.978	-
MAlex(2019) [28]	CEE	0.962	0.924	0.975	-
SFCN(2020) [20]	ACCESS	0.876	-	-	-
HGCN	**Ours**	**0.979**	**0.958**	**0.989**	**0.952**

disambiguation proposed in 1995; Co-training [1] approach provided a PCA-style analysis for the setting of semi-supervised learning from both labeled and unlabeled data; PsoFuzzy [30] involved preprocessing, combination of particle swarm optimization algorithm and fuzzy C-means clustering for the severity grading of diabetic macular edema; VQSSL [26] achieved macula localization, exudate detection, and grading of diabetic macular edema by using a vector quantization technique and formulated using a set of feature vectors; CLAHE [12] introduced the employment of image processing with histogram equalization, and the constrast limited adaptive histogram equalization techniques, then utilized the classifier of convolutional neural network to conduct diagnosis; HPSCNN [10] proposed a hierarchical pruning method based on VGG16-Net which is to modified containing fewer trainable parameters for DR classification; MAlex [28] used convolutional neural network with the application of suitable Pooling, Softmax, and Rectified Linear Activation Unit (ReLU) layers to obtain a high level of accuracy for the DR fundus image classification. Those baselines are under both of semi-supervised and supervised learning frameworks, and we employ the highest performance of each method for fair comparison. To further provide the evidence of the advantages for HGCN, we also deploy GCN [17], which is the basic graph convolution framework for our HGCN, as a technical baseline, and our method achieves the best results when we give total labeled data to GCN.

D. RESULTS

We implement the testing in ten times to compute the average results and report the performance of HGCN in Table 1. We set $lp = [20\%, 40\%, 60\%, 80\%, 100\%]$ to evaluate the effectiveness of HGCN with different numbers of labeled data. From Table 1, it can be seen that our HGCN achieves accuracy of 0.893, sensitivity of 0.802, specificity of 0.909 and F1-score of 0.793 when we adopt 20% of training data as labeled data, and the best performance is achieved by $lp = 100\%$. The obtained performance is increasing along with the scale of labeled data, which demonstrates that our proposed HGCN can solve retinal image classification when given limited labeled images, and it can promote the manifestation of full labeled data, which can be seen in supervised manner.

Beyond that, we compare the best results of HGCN and the state-of-the-art baselines, as mentioned in Section IV-C. The compared results are reported in Table 2, which the comparable best performance are **in bold**. For semi-supervised learning methods, we choose the best performance with appropriate scale of labeled data. For example, VQSSL achieves its best performance of accuracy of 0.975, sensitivity of 0.946, specificity of 0.980 and F1-score of 0.940 when it selects 80% labeled data. From the comparison between GCN, our propose network has a 3.1% improvement of accuracy, which is caused by our graph learning module. It is obvious in Table 2 that our HGCN achieves the best performance among them, that illustrates the model proposed in this paper has clear superiority to the baselines.

The results in Table 2 are used to demonstrate the superiority of our method with 100% labeled annotations, compared to the state-of-the-art. However, the major advantage of this paper is the effectiveness on semi-supervised retinal image classification, which has not addressed in the

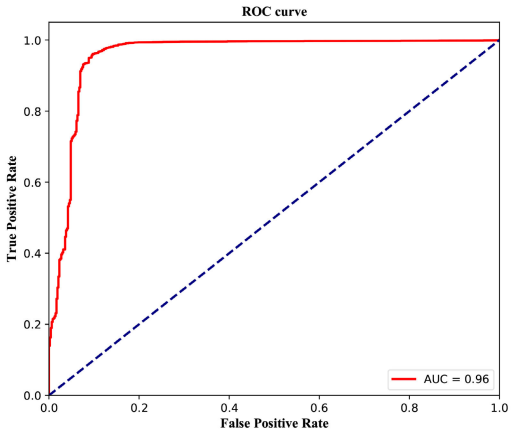

FIGURE 3. ROC curve of DR classification with $lp = 100\%$.

compared methods. The accuracy reaches 89.3% with only 20% annotations, and it exceeds 90% when we utilize 40% annotations. Besides, the performance with 100% annotation is the best among various compared methods. The finding has great significance in practical retinal image classification, because it can save at least 80% professional annotating labors, and achieve effective results.

E. FURTHER DISCUSSION

1) ROC CURVE

To better present the classification ability of the proposed HGCN model, we utilize the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) as evaluation metrics. Taking $lp = 100\%$ as an example to reveal the setting of limited labeled data, we draw the ROC curve in Figure 3, where the AUC is laid out. Figure 3 displays a favorable classification performance, and obtains the AUC value of 0.96, where further demonstrates the effectiveness of HGCN on semi-supervised retinal image classification task.

2) TRAINING CONVERGENCE

Towards the model training, we adopt the accuracy and loss curves along with the training process to imply the training trend on accuracy and model cost. The accuracy and loss curves of HGCN by $lp = 20\%$ is shown in Figure 4, which reflects that the performance achieves satisfactory result by 160-th epoch, and becomes stable. These curves demonstrate the convergence of the model, which explicitly evaluates its stability on semi-supervised retinal image classification. Besides, the total training time of HGCN on MESSIDOR dataset is about three hours when it reaches convergence. That costs 1.125 minutes per epoch. In a word, the training convergence and time reveal the computing efficiency of our hybrid graph convolutional network.

3) T-SNE

During the main training, we monitor separability in feature space generated by the encoder. We generate 2-dimensional embeddings with T-SNE [21] and visualize them in the testing phase for manual control of training performance. Figure 5 shows T-SNE embeddings labeled different colors by ground truth data. From the picture, it can be seen that images with no signs of DR are separable with a large margin from other images that have any sign of DR. Additionally, stages of DR come sequentially in embedding space, which corresponds to semantics in real diagnosis.

TABLE 3. Results of HGCN across different learning rates (*lr* denotes the learning rate).

lr	Accuracy	Sensitivity	Specificity	F1-score
10e-1	0.636	0.623	0.642	0.639
10e-2	0.827	0.809	0.821	0.816
10e-3	**0.979**	**0.958**	**0.989**	**0.952**
10e-4	0.854	0.838	0.847	0.836
10e-5	0.759	0.736	0.757	0.752

TABLE 4. Accuracies of HGCN across different number of units in Graph Learning (GL) and Graph Convolution (GC) layers with $lp = 100\%$.

Units	30	40	50	60	70
GL layer	0.952	0.946	0.960	**0.979**	958
GC layer	0.953	**0.979**	0.962	0.958	0.961

F. PARAMETER ANALYSIS

1) LEARNING RATE *lp* AND UNIT NUMBER

The proposed HGCN in this paper contains many parameters, and this section validates the performance of HGCN when different parameters are employed. We mainly evaluate the learning rate, and the number of unites in graph learning layer and graph convolution layer. Firstly, we utilize different learning rates of the range from 10e-1 to 10e-5 with $lp = 100\%$, and the results is in Table 3, where can be seen that the best results occur with learning rate of 10e-3. Secondly, we evaluate the different number of units in graph learning layer and graph convolution layer with $lp = 100\%$. The results are shown in Table 4, where denotes the best accuracy is generated in the number of units are 60, and 40 in graph learning and graph convolution layers, individually. From these evaluated experiments, the parameter analysis of learning rate, number of units is sufficiently discussed, and our HGCN achieves the best performance when learning rate is 0.001, the number of units is 60 in graph learning layer, and 40 in graph convolution layer.

2) BALANCE PARAMETERS λ_1, λ_2, AND λ_3

Besides, the parameters λ_1, λ_2, and λ_3 in Eq. 17 balance weights of different loss functions $\mathcal{L}_{MGL}$, $\mathcal{L}_t$, and $\mathcal{L}_r$, separately. As mentioned in ***Implementation*** (Section IV-B), the best results are achieved when $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = 0.8$. To justify how to select each balance parameters, we evaluate the accuracy performance when modify each parameter in different values, reported in Table 5. The changes of accuracy in different parameter values elaborate the importances of each term, and the optimal parameters cause the performable effectiveness of our proposed HGCN network on semi-supervised retinal image classification.

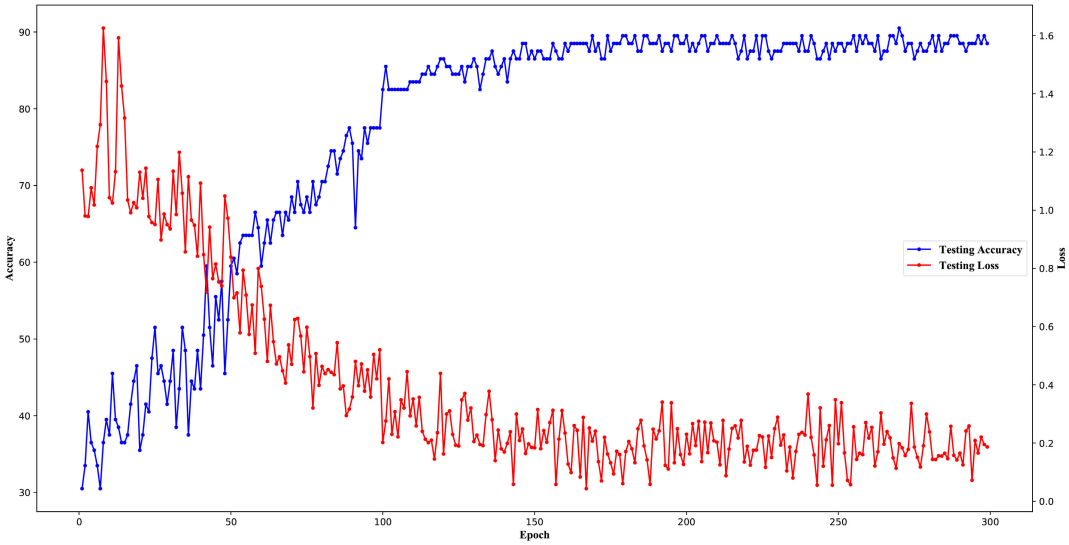

FIGURE 4. Testing loss and accuracy curves along with training process on MESSIDOR dataset.

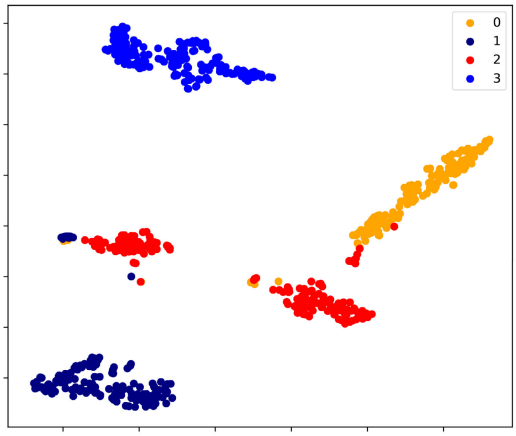

FIGURE 5. Feature embeddings with T-SNE from testing data.

TABLE 5. Accuracy performance from different parameters λ_1, λ_2, and λ_3 ($lp = 100\%$).

Value	λ_1	Value	λ_2	λ_3
0	0.538	0	0.813	0.866
1.8	0.932	0.6	0.906	0.966
2.0	0.979	0.8	0.954	0.979
2.2	0.958	1.0	0.979	0.972
2.4	0.940	1.2	0.948	0.970

To evaluate the effectiveness of these parameters, we set λ_1, λ_2, and λ_3 by 0 alternatively, and obtain 0.538, 0.813, 0.866 accuracies (Table 5). The results elaborate MGL loss $\mathcal{L}_{MGL}$ contributes 0.441 accuracy, triplet loss $\mathcal{L}_t$ and reconstruction loss $\mathcal{L}_r$ also make considerable improvements of 0.166 and 0.113 on accuracy performance. The evaluation of these three balance parameters demonstrates that each module in HGCN have important contribution to semi-supervised retinal image classification.

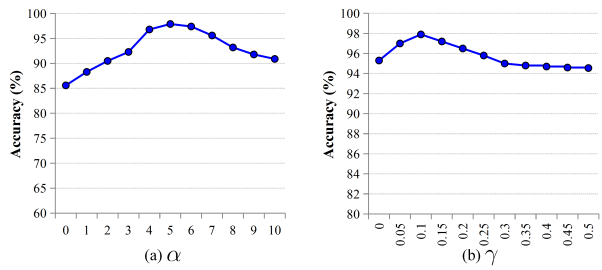

FIGURE 6. Parameter analysis for α (Eq.5) and γ (Eq.11).

3) MARGIN PARAMETER α AND γ

Importantly, the margin parameter $\alpha = 5$ controls the effectiveness of triplet loss, and $\gamma = 0.1$ is in charge of the weight of term $\|S\|_F^2$. Here, we select different values for these two parameters and implement validations to show their effectivenesses, which can be observed in Figure 6. In detail, the margin parameter α in Eq.5 achieves around 12% accuracy improvement which reaches the best performance at $\alpha = 5$, and the term $\|S\|_F^2$ in Eq.11 obtains the best result at $\gamma = 0.1$ with almost 2.6%. This parameter evaluation proves that both of margin parameter of triplet loss and MGL loss make considerable progresses for the semi-supervised retinal image classification task.

V. CONCLUSION

In this paper, we propose a Hybrid Graph Convolutional Network (HGCN) to classify diabetic retinopathy grading with limited labeled data and large amount of unlabeled data (Semi-supervised Learning). The proposed HGCN combines CNN and GCN into a unified framework to learn a hybrid graph representative features, which is conducted by modularity-based graph learning module and semi-supervised hybrid classification module. This network learns the synthetic structural information in semi-supervised

learning, and the experimental results show the effectiveness of HGCN on semi-supervised retinal image classification task, which solves the problem of lacking sufficient labeled data in clinical diabetic retinopathy diagnosis.

REFERENCES

[1] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, 1998, pp. 92–100.

[2] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On modularity clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 172–188, Feb. 2008.

[3] B. Chen, S. Xia, Z. Chen, B. Wang, and G. Wang, "RSMOTE: A self-adaptive robust SMOTE for imbalanced problems with label noise," *Inf. Sci.*, vol. 553, pp. 397–428, Apr. 2021.

[4] V. Cheplygina, M. D. Bruijne, and J. P. W. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, May 2019.

[5] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed database: The Messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

[6] O. Faust, R. Acharya, E. Y. K. Ng, K.-H. Ng, and J. S. Suri, "Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review," *J. Med. Syst.*, vol. 36, no. 1, pp. 145–157, Feb. 2012.

[7] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[8] Early Treatment Diabetic Retinopathy Study Research Group, "Grading diabetic retinopathy from stereoscopic color fundus photographs—An extension of the modified airlie house classification: ETDRS report number 10," *Ophthalmology*, vol. 98, no. 5, pp. 786–806, 1991.

[9] V. Gulshan, R. P. Rajan, K. Widner, D. Wu, P. Wubbels, T. Rhodes, K. Whitehouse, M. Coram, G. Corrado, K. Ramasamy, and R. Raman, "Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India," *JAMA ophthalmol.*, vol. 137, no. 9, pp. 987–993, 2019.

[10] M. Hajabdollahi, R. Esfandiarpoor, K. Najarian, N. Karimi, S. Samavi, and S. M. R. Soroushmehr, "Hierarchical pruning for simplification of convolutional neural networks in diabetic retinopathy classification," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 970–973.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[12] D. J. Hemanth, O. Deperlioglu, and U. Kose, "An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 707–721, Feb. 2020.

[13] B. Jiang, Z. Zhang, D. Lin, J. Tang, and B. Luo, "Semi-supervised learning with graph learning-convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11313–11320.

[14] S. Joshi and P. T. Karule, "A review on exudates detection methods for diabetic retinopathy," *Biomed. Pharmacotherapy*, vol. 97, pp. 1454–1460, Jan. 2018.

[15] M. Kim, D.-G. Lee, and H. Shin, "Semi-supervised learning for hierarchically structured networks," *Pattern Recognit.*, vol. 95, pp. 191–200, Nov. 2019.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[17] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Represent.*, Toulon, France, 2017, pp. 1–13.

[18] N. Kumar, S. Chandran, A. V. Rajwade, and S. P. Awate, "Semi-supervised robust one-class classification in RKHS for abnormality detection in medical images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 544–548.

[19] X. Li, L. Shen, M. Shen, F. Tan, and C. S. Qiu, "Deep learning based early stage diabetic retinopathy detection using optical coherence tomography," *Neurocomputing*, vol. 369, pp. 134–144, Dec. 2019.

[20] Y. Luo, J. Pan, S. Fan, Z. Du, and G. Zhang, "Retinal image classification by self-supervised fuzzy clustering network," *IEEE Access*, vol. 8, pp. 92352–92362, 2020.

[21] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[22] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Washington, DC, USA, Apr. 2018, pp. 1038–1042.

[23] R. F. Mansour, "Evolutionary computing enriched computer-aided diagnosis system for diabetic retinopathy: A survey," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 334–349, 2017.

[24] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Y. K. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2136–2155, 2013.

[25] R. Raman, P. K. Rani, S. R. Rachepalle, P. Gnanamoorthy, S. Uthra, G. Kumaramanickavel, and T. Sharma, "Prevalence of diabetic retinopathy in India: Sankara Nethralaya diabetic retinopathy epidemiology and molecular genetics study report 2," *Ophthalmology*, vol. 116, no. 2, pp. 311–318, 2009.

[26] F. Ren, P. Cao, D. Zhao, and C. Wan, "Diabetic macular edema grading in retinal images using vector quantization and semi-supervised learning," *Technol. Health Care*, vol. 26, pp. 389–397, May 2018.

[27] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823.

[28] T. Shanthi and R. S. Sabeenian, "Modified alexnet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56–64, Jun. 2019.

[29] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," *Diabetes Res. Clin. Pract.*, vol. 87, no. 1, pp. 4–14, Jan. 2010.

[30] K. S. Sreejini and V. K. Govindan, "Severity grading of DME from retina images: A combination of PSO and FCM with Bayes classifier," *Int. J. Comput. Appl.*, vol. 81, no. 16, pp. 11–17, Nov. 2013.

[31] M. Tavakoli, S. Jazani, and M. Nazar, "Automated detection of microaneurysms in color fundus images using deep learning with different preprocessing approaches," *Proc. SPIE*, vol. 11318, Mar. 2020, Art. no. 113180E.

[32] M. Tavakoli and M. Nazar, "Comparison different vessel segmentation methods in automated microaneurysms detection in retinal images using convolutional neural networks," *Proc. SPIE*, vol. 11317, Feb. 2020, Art. no. 113171P.

[33] M. Tavakoli, M. Nazar, and A. Mehdizadeh, "The efficacy of microaneurysms detection with and without vessel segmentation in color retinal images," *Proc. SPIE*, vol. 11314, Mar. 2020, Art. no. 113143Y.

[34] N. Thakur and M. Juneja, "Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma," *Biomed. Signal Process. Control*, vol. 42, pp. 162–189, Apr. 2018.

[35] R. L. Thomas, F. Dunstan, S. D. Luzio, S. R. Chowdury, S. L. Hale, R. V. North, R. Gibbins, and D. R. Owens, "Incidence of diabetic retinopathy in people with type 2 diabetes mellitus attending the diabetic retinopathy screening service for Wales: Retrospective analysis," *Bmj*, vol. 344, p. e874, Feb. 2012.

[36] D. S. W. Ting, G. C. M. Cheung, and T. Y. Wong, "Diabetic retinopathy: Global prevalence, major risk factors, screening practices and public health challenges: A review," *Clin. Experim. Ophthalmol.*, vol. 44, no. 4, pp. 260–277, May 2016.

[37] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," 2020, *arXiv:2003.02261*. [Online]. Available: http://arxiv.org/abs/2003.02261

[38] T. Walter, J. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002.

[39] J. Wu, Y. Zhang, J. Wang, J. Zhao, D. Ding, N. Chen, L. Wang, X. Chen, C. Jiang, X. Zou, and X. Liu, "Attennet: Deep attention based retinal disease classification in OCT images," in *Proc. Int. Conf. Multimedia Modeling*. Cham, Switzerland: Springer, 2020, pp. 565–576.

[40] S. Xia, Y. Liu, X. Ding, G. Wang, H. Yu, and Y. Luo, "Granular ball computing classifiers for efficient, scalable and robust learning," *Inf. Sci.*, vol. 483, pp. 136–152, May 2019.

[41] S. Xia, D. Peng, D. Meng, C. Zhang, G. Wang, E. Giem, W. Wei, and Z. Chen, "A fast adaptive K-means with no bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 13, 2020, doi: 10.1109/TPAMI.2020.3008694.

[42] S. Xia, G. Wang, Z. Chen, Y. Duan, and Q. Liu, "Complete random forest based class noise filtering learning for improving the generalizability of classifiers," *IEEE Comput. Archit. Lett.*, vol. 31, no. 11, pp. 2063–2078, 2019.

[43] S. Xia, Z. Xiong, Y. Luo, L. Dong, and C. Xing, "Relative density based support vector machine," *Neurocomputing*, vol. 149, pp. 1424–1432, Feb. 2015.

[44] S. Xia, Z. Xiong, Y. Luo, L. Dong, and G. Zhang, "Location difference of multiple distances based k-nearest neighbors algorithm," *Knowl.-Based Syst.*, vol. 90, pp. 99–110, Dec. 2015.

[45] S. Xia, Z. Zhang, W. Li, G. Wang, E. Giem, and Z. Chen, "GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification," *IEEE Trans. Knowl. Data Eng.*, early access, May 25, 2020, doi: 10.1109/TKDE.2020.2997039.

[46] Y. Xie, J. Zhang, and Y. Xia, "Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT," *Med. Image Anal.*, vol. 57, pp. 237–248, Oct. 2019.

[47] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proc. 33rd Annu. Meeting Assoc. Comput. Linguistics*, 1995, pp. 189–196.

[48] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S. J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, and S. Haffner, "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes Care*, vol. 35, no. 3, pp. 556–564, 2012.

[49] Y. Zhou, X. He, L. Huang, L. Liu, F. Zhu, S. Cui, and L. Shao, "Collaborative learning of semi-supervised segmentation and classification for medical images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2079–2088.

[50] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

GUANGHUA ZHANG (Member, IEEE) was born in 1986. He received the Ph.D. degree in computer science and technology from Chongqing University, in 2017. He is currently the Director of Shanxi Intelligent Big Data Industry Technology Innovation Research Institute, and a Professor with the Department of Computer Science and Engineering, Taiyuan University. His research interests include machine learning, multispectral image processing, computer vision, and medical images analysis.

JING PAN was born in 1986. She received the Ph.D. degree from Southwest University, in 2015. She is currently a Professor with the Department of Physics and Chemistry, Taiyuan University. She has a wealth of research experience in the synthesis of nanomaterials, battery electrode modified, math, optimization, computer vision, and coding.

ZHAOXIA ZHANG was born in 1976. She received the Ph.D. degree in ophthalmology from Fourth Military Medical University, in 2009. She is currently an Associate Chief Physician with the Shanxi Eye Hospital. Her research interests include the relationship of RGR isoform and mechanism of age related macular degeneration (AMD), the mechanism of autoimmune ophthalmopathy, especially the thyroid associated ophthalmopathy (TAO), and uveitis.

HENG ZHANG received the Ph.D. degree in military logistics from Logistics Engineering University, China, in 2013. His research interests include big data intelligence, machine learning algorithm, and medical informatization.

CHANGYUAN XING received the B.Sc. and M.Sc. degrees from Yanshan University, in 2005 and 2009, respectively, and the Ph.D. degree in computer science from Chongqing University, in 2015. He is currently working with Yangtze Normal University, China. His research interests include pattern recognition and machine learning.

BIN SUN was born in 1964. He received the Ph.D. degree in internal medicine from Xi'an Jiaotong University, in 2018. He is the Dean of the Shanxi Eye Hospital. He is also an Expert of thyroid-associated ophthalmopathy (TAO). His research interest includes the mechanism, diagnosis, and treatment of TAO.

MING LI was the Chairman of the Fifth International Seminar on anti-problem theory and computational analysis, and invited many times to report at international computational methods conferences, such as Taiwan, Australia, and Singapore. He has published more than 70 SCI articles and one English monograph. His research interests include computational mathematics, data science and technology, and grid free computing methods.