IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 10, 2021, accepted October 3, 2021, date of publication October 11, 2021, date of current version October 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3119434

Semi-Supervised Auto-Encoder Graph Network for Diabetic Retinopathy Grading

YUJIE LI[1,2], ZHANG SONG[3], SUNKYOUNG KANG[2], SUNGTAE JUNG[2], AND WENPEI KANG[4]

[1] Weifang Key Laboratory of Blockchain on Agricultural Vegetables, Weifang University of Science and Technology, Weifang, Shandong 262700, China
[2] Department of Computer Software Engineering, Wonkwang University, Iksan, Jeonbuk 54538, Republic of Korea
[3] The Affiliated Hospital of Qingdao University, Qingdao 266000, China
[4] Business College of Southwest University, Chongqing 402460, China

Corresponding authors: Sunkyoung Kang (doctor10@wku.ac.kr) and Sungtae Jung (stjung@wku.ac.kr)

This work was supported by Wonkwang University in 2021.

ABSTRACT Diabetic Retinopathy (DR) causes quite a few blindness worldwide, which can be refrained by the timely diagnosis on retinal images. Recently, researches on deep learning-based retinal image classification have accelerated outstanding improvements in DR grading task. However, existing DR grading works are mostly limited to a supervised manner. They require accurately annotated data labeled by professional experts, and the annotating work is very laborious and time-consuming. We propose a Semi-supervised Auto-encoder Graph Network (SAGN) for the challenging DR diagnosis to relax this constraint. Precisely, SAGN consists of three major modules: auto-encoder feature learning, neighbor correlation mining, and graph representation. Firstly, our model learns to extract representations from retinal images and reconstruct them as close to original inputs as possible. Then neighbor correlations among labeled and unlabeled samples are established by their similarities, calculated by the radial basis function. Finally, we operate Graph Convolutional Neural Network (GCN) to grade retinal samples from extracted features and their correlations. To evaluate the performance of SAGN, we conduct sufficient comparative experiments on APTOS 2019 dataset, trained from EyePACS. Results demonstrate that our SAGN model can achieve comparable performance with limited labeled retinal images with the help of large amounts of unlabeled data.

INDEX TERMS Diabetic retinopathy grading, semi-supervised learning, auto-encoder, graph convolutional network.

I. INTRODUCTION

The retinal blood vascular network is the only vascular network of a human body visible to a non-invasive imaging approach. In consequence, automated analysis of retinal vascular structure is the most common way to support examination, diagnosis, and treatment of many diseases [1]–[4], especially for diabetic retinopathy (DR). In practice, ophthalmologists use color and morphological information to diagnose retinal images into DR grades by discriminating between arteries and veins since arteries contain more oxygen and appear brighter than veins and thinner than neighboring veins [5]. These features of retinal vasculature are usually captured by fundus photography due to its lower cost and ease of use, but manual classification of retinal blood vessels is time-consuming and subject to human errors.

In recent years, kinds of researches involved machine learning into automatic DR grading based on retinal images. As an advanced technology in machine learning, deep learning-based automatic retinal image classification methods exhibit outstanding DR grading performance, surpassing traditional machine learning models [6]–[8]. They utilize large amounts of retinal images to train Convolutional Neural Networks (CNNs), supervised by full annotations, which professional DR experts capture. However, the annotation work results in a complicated burden in an actual application that wastes so much professional human resources and brings inevitable noise labels [9], [10].

In order to alleviate the annotating workload for experts, this work introduces a semi-supervised framework to utilize partially labeled retinal data with large-scale unannotated

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

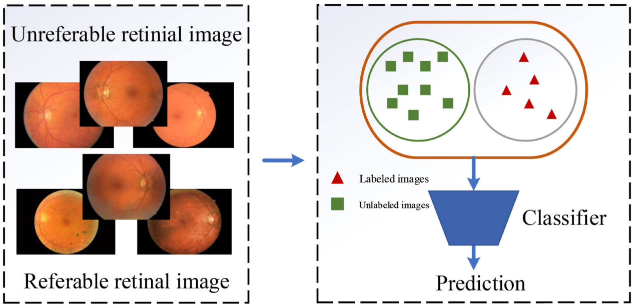

FIGURE 1. The illustration of semi-supervised diabetic retinopathy grading.

images to train a DR grading model, as illustrated in Figure 1. It can be seen in this Figure, as an efficient unsupervised pretraining method [11], [12], the auto-encoder is not limited by label information and can eliminate the noise [13] in the data. We recommend using the autoencoder for network self-training to analyze high-dimensional features further. At the same time, compared with the conventional neural network, the graph neural network utilizes graphs as input and learns to ratiocinate and predict how objects and their relationships evolve. In addition, the graph network can make the network less vulnerable to adversarial attacks because it is a system that represents things as objects instead of pixel patterns and will not be easily disturbed by a bit of noise. Thus, we propose a novel Semi-supervised Auto-encoder Graph Network (SAGN) for training the DR grade predictor using the limited labeled data. SAGN feeds a small quantity of labeled retinal images and plentiful unlabeled data into an auto-encoder to mine the CNN representations by encoder-decoder architecture. Then, it exploits the neighbor correlations among both labeled and unlabeled images according to their similarities. Finally, a convolutional graph network operates graph feature learning with the help of the learned neighbor correlations to output the grades of each input image. To sufficiently train the network, SAGN optimizes the whole network in an end-to-end manner within each batch.

In general, SIGN offers the following contributions:

(1)We propose the Unsupervised Auto-encoder module (UA), which is not restricted by annotation information, to make the network self-training, and it is also considered a powerful feature extractor.

(2)We explore the intrinsic correlation from limited labeled data and massive unlabeled samples at the feature level via Graph Network (GN) module to spread the annotation information to the entire data set.

(3)We conduct comparative experiments on two popular public available DR grading datasets (APTOS 2019 and Kaggle DR), which reveal the superiority of our model on the retinal image classification task.

II. RELATED WORK

This section discusses recently proposed retinal image classification methods based on supervised learning and then introduces many applications of the semi-supervised framework on medical image classification.

A. RETINAL IMAGE CLASSIFICATION

There have been many outstanding works in the application of deep learning in the field of medical imaging [14]–[16]. In recent years, many supervised CNN methods have progressively developed retinal image classification [17]–[20] with the evolution of deep learning. For example, Marin *et al.* [18] detected retinal exudates by applying digital image processing algorithms to the retinal image to obtain a set of candidate regions, which are validated utilizing feature extraction and supervised classification techniques. Xu *et al.* [19] proposed an improved supervised artery and vein classification method in retinal images, which uses intra-image regularization and inter-subject normalization to reduce the differences in feature space. Playout *et al.* [17] employed a novel approach for training a convolutional multitask architecture of retinal with supervised learning and reinforcing it with weakly supervised learning. Similarly, Sreeja and Kumar [20] presented a supervised machine learning algorithm based on retinal hemorrhage detection and classification with the help of splat level and GLCM features extracted from the splats.

However, all of these approaches require a large number of tagged retinal datasets to supervise the training procedure, which requires a lot of time and effort for manual annotation. By contrast, this paper proposes a novel semi-supervised retinal image classification model to conduct automatic DR grading only requiring a small number of annotated retinal images, which can largely save professional manpower and time.

B. SEMI-SUPERVISED LEARNING IN MEDICAL IMAGE ANALYSIS

Because the annotating work in medical image analysis is more expensive and scarce than traditional computer vision tasks (e.g., face, person, dog recognition), Semi-Supervised Learning (SSL) approaches play an important role in automatic medical image recognition alleviate the professional labeling work. At the same time, unlabeled data is much more in practice. Some unsupervised and semi-supervised methods have made breakthroughs in medical graphics analysis [21], [22]. Inspired by these medical image methods, the researchers applied their ideas to retinal analysis. To leverage unlabeled data, Bakalo *et al.* [23] proposed a deep learning architecture based on SSL for multi-class classification and localization of abnormalities in medical imaging illustrated through experiments on mammograms, which enables detection of abnormalities at full mammogram resolution for both weakly and semi-supervised settings; Han *et al.* [24] exploited a weak and semi-supervised deep learning framework to segment prostate cancer in TRUS images, alleviating the time-consuming work of radiologists to draw the boundary of the lesions and training the neural network on the data that do not have complete annotation. Menon *et al.* [25] presented a semi-supervised algorithm for lung cancer screening in which a 3D Convolutional Neural Network (CNN) is using the expectation-maximization meta-algorithm.

Inspired by the successful application of semi-supervised learning in medical image analysis, this paper introduces a novel SSL framework to solve the cumbersome labeling work in the retinal image classification task.

III. METHOD

This paper proposes a semi-supervised retinal image classification method with auto-encoder feature learning, neighbor correlation mining, and graph representation modules. In this task, we define input retinal images as $I = I^l \cup I^u$ with $N = N_l + N_u$ images, where $I^l = \left\{i_1^l, i_2^l, \cdots, i_{N_l}^l,\right\}$ are labeled images with the correlated ground-truth class labels $\boldsymbol{y}^l = \left\{y_1^l, y_2^l, \cdots, y_{N_l}^l\right\}$, and $\boldsymbol{I}^u = \left\{i_1^u, i_2^u, \cdots, i_{N_u}^u\right\}$ represent the large scale of unannotated retinal images without any annotations.

A. AUTO-ENCODER FEATURE LEARNING

In our SAGN model, we firstly design a CNN-based encoder-decoder to exploit the feature learning capability for each retinal image in I. Aiming to discover robust representations of retinal images, we utilize an encoder F to extract appropriate CNN feature embeddings for labeled and unlabeled images. Besides, we also integrate a decoder D to re-construct the images from CNN feature embeddings, which makes the feature vectors contain meaningful representations for retinal images, further developing the feature learning efficiency of the auto-encoder.

Mathematically, the encoder F can transform each retinal image into a low-dimensional feature space, such as a labeled retinal image i_j^l and an unlabeled sample i_k^u, which can be compacted into feature vectors $F(i_j^l; W_f)$, and $F(i_k^u; W_f)$. To optimize the auto-encoder architecture, we introduce the decoding loss attached to the decoder D, following,

$$L_{dec} = \frac{1}{N_l}\sum_{j=1}^{N_l}\left\|\mathcal{D}\left(F\left(\boldsymbol{i}_j^l; W_f\right); W_d\right) - \boldsymbol{i}_j^l\right\|_2^2 + \frac{1}{N_u}\sum_{k=1}^{N_u}\left\|\mathcal{D}\left(F\left(\boldsymbol{i}_k^u; W_f\right); W_d\right) - \boldsymbol{i}_k^u\right\|_2^2 \quad (1)$$

where W_f, W_d are trainable parameters in encoder and decoder, respectively.

Through the optimization of decoding loss, the auto-encoder can make the feature embeddings expressing meaningful information for themselves, and the remaining task is to distill class information from raw data. Here, we introduce a classifier C to predict the category for each retinal image $i_j \in I$, by mapping the feature embedding $F(i_j; W_f)$ into a class space $C(F(i_j; W_f); W_c)$, where W_c is the learnable parameters in the classifier. In our semi-supervised retinal image classification framework, the CNN Cross-Entropy (CCE) loss is minimized to train the classifier C and encoder F jointly,

$$L_{cce} = -\sum_{j=1}^{N^l} y_j^l \log C(F(i_j; W_f); W_c)) \quad (2)$$

where the CE loss is only calculated on the labeled retinal images I^l, because their existing corresponding ground-truth labels $\boldsymbol{y}^l$ can supervise the network.

Due to the limited number of labeled retinal images, the classifier C cannot reach a desirable performance only with the encoder. In our auto-encoder feature learning module, the utilization of the decoder D and its decoding loss L_{dec} can reinforce the representation capability of the auto-encoder.

B. NEIGHBOR CORRELATION MINING

As we all know, the labeled and unlabeled samples in I follows a uniform distribution rather than individual objects. We believe that intrinsic correlations must remain among each sample in I after CNN feature embedding. A simple rule is that the feature embeddings from the same category are more similar than ones from different DR classes. According to the similarity between different retinal images, we can establish similarity-based correlations among both the neighboring massive unannotated samples and labeled images, which is very useful for training the classifier C. Though fully annotating sufficient retinal images is unbearable in real applications, exploiting the massive unannotated images and constructing similarity-based correlations underlying various categories of fundus images can further mine meaningful information from limited labeled and large amounts of unlabeled retinal images.

Given the CNN feature embedding $F(i_j)$ from retinal images I, the Radial Basis Function (RBF) [26] is introduced to calculate the similarity $s(i_j, i_k)$ between retinal images i_j and i_k,

$$s\left(i_j, i_k\right) = \exp\left(-\frac{d\left(F\left(i_j\right), F\left(i_k\right)\right)}{2\sigma^2}\right) \quad (3)$$

where $d(\cdot, \cdot)$ represent Euclidean distance and σ is a scale factor. This term keeps the similarity ranging from 0 to 1 and $s(i_j, i_j) = 1$, which will be smaller when the distance of $F(i_j)$ and $F(i_k)$ is increasing.

According to this similarity calculation, we can build the correlation graph G by calculating similarities between each pair of retinal image features. Particularly, each node in graph G denotes a retinal image, and the edge between two nodes represents the similarity between these two image features, which are from both labeled and unlabeled images. Assume an adjacent matrix $A \in \mathbb{R}^{N\times N}$ to represent G, and $A(j, k) = s(i_j, i_k)$ if $s(i_j, i_k) > \tau$ else $A(j, k) = 0$. Furthermore, $s(i_j, i_j) = 1$ ensures the graph A is self-connected, and two similar images are connected, and the edge between them is large. As we all know, the connected similar images can provide much more information to update each other, while the disconnected image features should be optimized individually to avoid misleading. Through the similarity-based graph

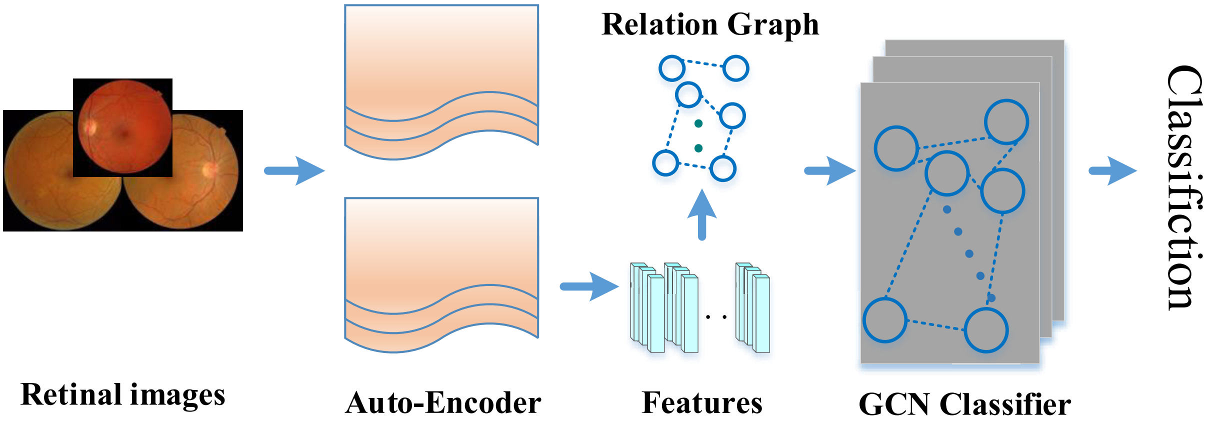

FIGURE 2. The scheme of our proposed semi-supervised auto-encoder graph network on diabetic retinopathy grading task.

establishment, the correlations among labeled and unlabeled retinal images can be mined by G, which provides essential cues for further feature learning.

C. GRAPH REPRESENTATION MODULE

This paper utilizes Graph Convolutional Network(GCN) to explore the feature-level correlations among retinal image features, labeled or unlabeled ones. It composes of M graph convolutional layers, attached with two fully connected layers. Besides, ReLU is integrated after the graph convolutional layer, and a PReLU is on the first fully connected layer.

Specifically, the graph convolution calculation for the m-th layer ($1 \leq m \leq M$) is mathematically formulated by,

$$X^m = \mathrm{ReLU}(\hat{A}X^{m-1}W^m) \tag{4}$$

where X^{m-1} and X^m represents the input and output of this layer, respectively; $X^0 = \{F(i_1), F(i_2), \cdots, F(i_{N_l+N_u})\}$ is the collection of learned CNN features by encoder F; $\hat{A} = \Lambda^{-\frac{1}{2}}A\Lambda^{-\frac{1}{2}}$, where Λ is the diagonal matrix of A; and W^m is the weight of m-th graph convolution layer.

Through the M stacked graph convolutions, the correlations among retinal images can be explored by graph representation, and we integrate softmax on the final perceptron layer as,

$$Z = X^M = \mathrm{softmax}(\hat{A}X^{M-1}W^M) \tag{5}$$

where $W^M \in \mathbb{R}^{d_M \times N_c}$ (N_c is the number of DR grades). The final output $Z \in \mathbb{R}^{(N_l+N_u)\times N_c}$ represents the predictions for each retinal images in which each row Z_j represents the predicted DR grades for j-th image in I. Finally, the optimization of weight parameters $\{W^1, W^2, \cdots, W^M\}$ in GCN is conducted by minimizing the semi-supervised cross-entropy (SCE) loss according to,

$$L_{sce} = -\sum_{i_j^l \in I^l} y_j^l \log Z_j \tag{6}$$

where I^l represents the labeled retinal images, and this loss function replaces the CNN cross-entropy loss in Eq. 2.

Depend on the neighbored samples' correlation in G. The convolutional graph network can distill the discriminative information from the limited labeled images to further mine knowledge from massive unlabeled retinal samples. Thus the supervision knowledge from the small number of labeled retinal images can guide the graph representations for unlabeled samples. Intrinsically, the annotations can propagate along with the connections in G, involving the weights among different nodes. As a result, the optimization of our network facilitates the classifier to grade the retinal images with the help of neighbor correlations G, which provide essential cues to make predictions more accurate and robust.

D. OPTIMIZATION

The auto-encoder and graph convolutional network ensure the end-to-end training manner in our semi-supervised retinal image classification task, simultaneously learning the graph representations of retinal images and output the predicted category for each feature. As illustrated in Figure 2, we firstly feed the limited labeled and massive unlabeled retinal images into the CNN encoder F to generate features $F([I^l, I^u])$, then build the neighbor correlations by RBF similarity (Eq. 3). Finally, conduct graph convolutions on the CNN features $F([I^l, I^u])$ with neighbor correlation graph G to output predicted class annotations. In our training stage, the learned CNN features $F([I^l, I^u])$ are re-constructed into original images for labeled and unlabeled samples. In particular, a more detailed figure of the network architecture is shown in Figure 3.

To train the whole network, we jointly optimize the decoder loss L_{dec} (Eq.1), and semi-supervised cross-entropy loss L_{sce} (Eq 6) jointly into a final loss function,

$$L = (1-\alpha)L_{dec} + \alpha L_{sce} \tag{7}$$

where $\alpha \in [0, 1]$ is the hyper-parameter to balance the weights of L_{dec} and L_{sce}. Besides, the mini-batch training algorithm is presented in Algorithm 1.

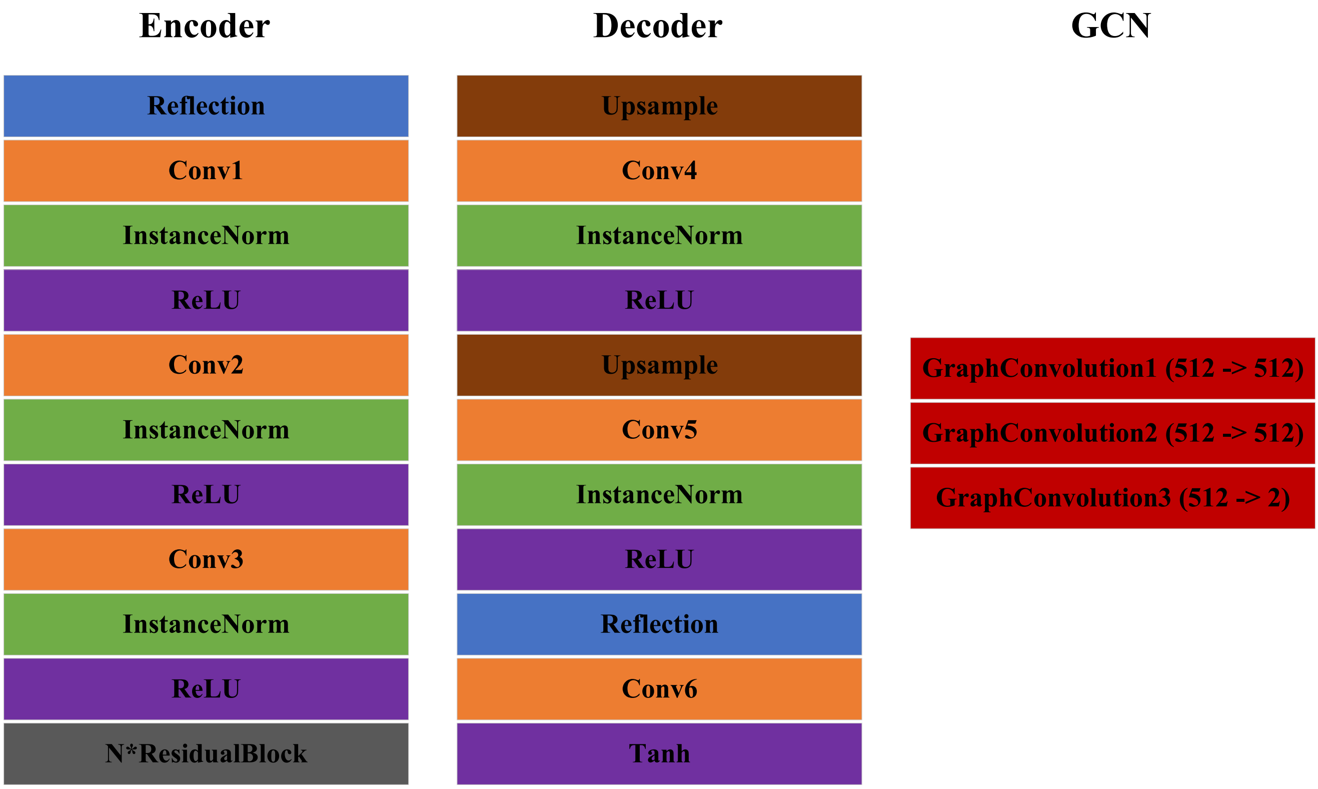

FIGURE 3. The network structure of our proposed model.

IV. EXPERIMENTAL RESULTS

To evaluate our SAGN network, this paper executes adequate implementation on popular diabetic retinopathy grading datasets, including APTOS 2019 [27] and EyePACS [28]. This section firstly introduces datasets and experimental details, and then reports the performance compared with state-of-the-art methods. Besides, the discussion of the main modules is also analyzed in this part.

A. EXPERIMENTAL DATASETS

EyePACS [28] collects 88,702 annotated colorful fundus images from different patients. These images are captured by different fundus cameras in multiple primary care sites throughout California and elsewhere, and the resolutions are resized to 512 × 512 pixels, categorized into five DR grades, including No, Mild, Moderate, Severe, and Proliferative DRs. The distribution is also summarized in Table 1. This dataset is employed as the training set, which provides partial annotations for SAGN, and this paper utilizes the APTOS 2019 as the testing set to report the classification performance on the semi-supervised DR grading task. In detail, **APTOS 2019** [27] is proposed in the APTOS 2019 diabetic retinopathy classification contest, which is organized by the Asia Pacific Tele-Ophthalmology Society. It comprises of 3,662 retinal images with available annotations, which are captured from multi-clinics with different imaging conditions under fundus photography at Aravind Eye Hospital in India. The distribution of this dataset is highly imbalanced, as summarized in Table 1. In our experiments, we deploy the EyePACS to train our SAGN model and test the model on APTOS 2019.

TABLE 1. The class distributions of EyePACS and APTOS 2019 datasets.

Label	EyePACS	APTOS 2019
No DR	65,343	1,805
Mild DR	6,205	370
Moderate DR	13,153	999
Severe DR	2,087	193
Proliferative DR	1,914	295

B. IMPLEMENTATION DETAILS

The whole network is implemented by the PyTorch framework on Ubuntu 18.04 with 2 Nvidia 3070 8G GPUs. The average time for each image to pass through the network is 0.03 seconds, and training stops when the loss function is smooth. After many verifications, we found that the model converged in about 40 epochs. The entire training process took 11.7 hours. To alleviate the influence of useless regions of the fundus images, we first remove the black regions for each image by cropping operation. Then, each retinal image is resized into 512 × 512 pixels before feeding into the network, and each image is augmented by randomly horizontal and

Algorithm 1 Training of the Semi-Supervised Auto-Encoder Graph Network

```
Input: Retinal image dataset $I = I^l \cup I^u$, and corresponding grade annotations $y^l$ of annotated samples.
repeat
    Choose random labeled and unlabeled samples from $I^l$ and $I^u$ separately to constitute the training batch $B$;
    Feed the chosen images $B$ into the CNN encoder $F$ and obtain the features $F(B)$;
    Establish the neighbor correlations among images by the similarity based graph $G$ following Eq 3;
    Feed the correlation graph $G$ and CNN features $F(B)$ into the GCN, and output the predicted category $Z_j$;
    Compute the semi-supervised cross-entropy loss $L_{sce}$ by Eq 6;
    Feed the learned CNN features $F(B)$ into the decoder $D$ and compute the reconstruction loss $L_{dec}$ via Eq.(2);
    Compute the final loss function $L = (1 - \alpha)L_{dec} + \alpha L_{sce}$;
    Optimize the network parameters of CNN encoder, decoder, and GCN according to back-propagation algorithm;
until Convergence;
Output: The optimized CNN encoder and GCN.
```

vertical rotation. As for the model training, SAGN is updated by Adam optimizer, and we set the learning rate and maximum epochs as $1e$-5, and 190, separately. The batch size is set to 32, where the ratio of labeled and unlabeled data in a batch is 1:1, and the ratio of the total quantity is 1:4. In detail, we utilize ResNet-50 [29] as encoder F, which removes the last fully connected layer, and the decoder D follows the architecture [30]. Besides, we adopt three convolutional graph layers to conduct GCN on the learned CNN features and output predictions. For parameter settings, the scale factor σ and threshold τ in graph building are 0.01 and $1e$-5, while α is set by 0.6. In this work, we consider the DR grading task as a binary classification (DR/No Dr) to validate the performance of SAGN.

C. EVALUATION METRICS

To quantitatively reveal the performance, this paper measures the model performance by three metrics, including Accuracy, Sensitivity, and Specificity, which are calculated by following equations,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{10}$$

where TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative.

TABLE 2. Performance of semi-supervised DR grading with different numbers of annotated training images (%).

Number	Accuracy	Sensitivity	Specificity
1K	74.6	68.2	71.5
10K	85.0	77.6	76.2
30K	94.4	84.0	82.2

Moreover, we also visualize the DR grading performance by t-Stochastic Neighbor Embedding (t-SNE), Receiver Operating Characteristic (ROC) curve. In detail, t-SNE is an effective tool for visualizing high-dimensional data by transforming each feature vector into a two-dimensional space, where nearby points model similar objects and dissimilar objects are modeled by distant points with high probability; The ROC is a graph illustrating the property of classification network at different probability thresholds on True Positive Rate (TPR) and False Positive Rate (FPR), calculated by

$$\text{TPR} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{FPR} = \frac{FN}{FP + TN} \tag{12}$$

where the Area Under ROC Curves (AUC) are also employed to evaluate the performance, indicating the classification capability of a classifier on DR grading.

D. PERFORMANCE OF SEMI-SUPERVISED DR GRADING

1) CLASSIFICATION WITH DIFFERENT NUMBER OF ANNOTATED IMAGES

In this paper, the proposed Semi-supervised Graph Network utilizes limited annotated retinal images and large amounts of unlabeled samples to train a discriminative DR grading model. To evaluate the effectiveness of SAGN, we select different numbers of annotated images to conduct experiments, including 1K, 10K, and 30K. As summarized in Table 2, Our SAGN can obtain 74.6% accuracy, 68.2% sensitivity, and 71.5% specificity when utilizes 1K annotated images, and it reaches an accuracy of more than 80% with 10K annotations. Besides, we also implement SAGN with 30K annotated retinal images, which realizes comparable results of 94.4% accuracy, 84.0% sensitivity, and 82.2% specificity. The results indicate that the performance is gradually increasing along with more annotated images.

2) RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE

ROC analysis can reveal the diagnostic performance for a classification model, displaying its graphical capability in the DR grading task. From Figure 4, the ROC curve of SAGN with 30K annotations has a potential increase of TPR with larger FPR, and it achieves the area under the ROC curve (AUC) of 0.90. The visualization of the ROC curve can fully demonstrate the effectiveness of the SAGN classifier on semi-supervised DR grading tasks with limited annotated retinal images and massive unlabeled data.

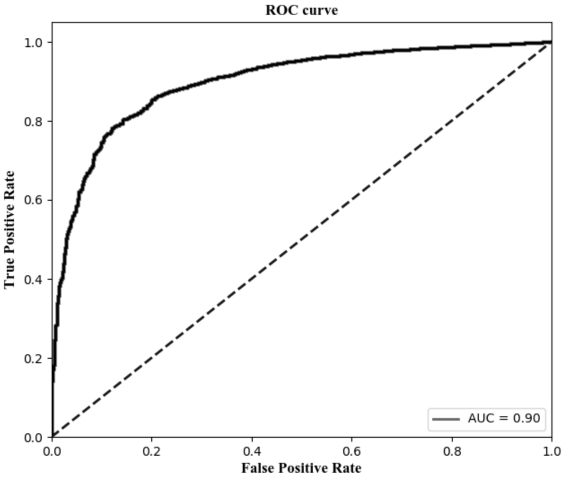

FIGURE 4. ROC curve of the proposed model on APTOS dataset with 30K annotations for DR grading.

3) T-STOCHASTIC NEIGHBOR EMBEDDING (T-SNE) VISUALIZATION

The t-SNE visualization can transform the learned high-dimensional feature representations into a low-dimensional space to reveal the feature learning ability of the classifier, which tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding the high-dimensional data. We visualize the feature representations from the GCN layer before the classifier. From Figure 5, the data points can be clearly divided into two groups (DR/no DR) with limited confused samples. These two groups represent the predicted classes belonging to DR and No DR, which states that the GCN feature representations contain enough discriminative information learned from raw images, benefit from the feature mining from the labeled data, and the graph learning from the unlabeled retinal samples.

E. COMPARISON WITH SUPERVISED MODELS

To further demonstrate the advanced DR grading performance, we compare our method with three completely supervised methods, including SE-ResNeXt50 [31], EfficientNet [32], and EnsembleNet [33], which are proposed recently with same training and testing data. In detail, SE-ResNeXt50 [31] designed a squeeze-and-excitation (SE) block adaptively recalibrating channel-wise feature responses by explicitly modeling interdependencies between channels, which boost the representational power of a network. EfficientNet [32] is an advanced neural architecture uniformly scaling all dimensions of depth/width/resolution using a highly effective compound coefficient. Different from the formers, EnsembleNet [32] is an ensemble network specially designed for the DR grading, composing a multi-task learning strategy with classification, regression, and ordinal regression for DR diagnostic classification. We also compared with three recent baseline models Resnet-50, Vgg-16, and Inception-V3.

We summarize the compared results in Table 3, and it can be observed that SAGN surpasses four supervised models (SE-ResNeXt50, EfficientNet, Vgg-16, and Inception-V3) and achieves 94.4% accuracy, 84.0% sensitivity, and 82.2% specificity with 30K annotated retinal images. Compared to EnsembleNet, SAGN only keeps a small distance, such as drop 4.2% accuracy. It is worth mentioning that, benefiting from the strong correlation between the samples mined by the graph neural network, SAGN can better identify suspected cases and submit them to experts for further screening, thus avoiding the possibility of missed diagnosis. In contrast, traditional supervision methods are limited to the sufficiency of annotation, but they often require large amounts of labeled data with cost-expensive and time-consuming human power. As for our SAGN, it perform weaker sensitivity and specificity, with limited distance to supervised models. However, SAGN only requires a small number of annotated samples under semi-supervised framework to save considerable annotating manpower. To sum up, our method has potential effectiveness on semi-supervised DR grading, and it is even superior to some supervised models.

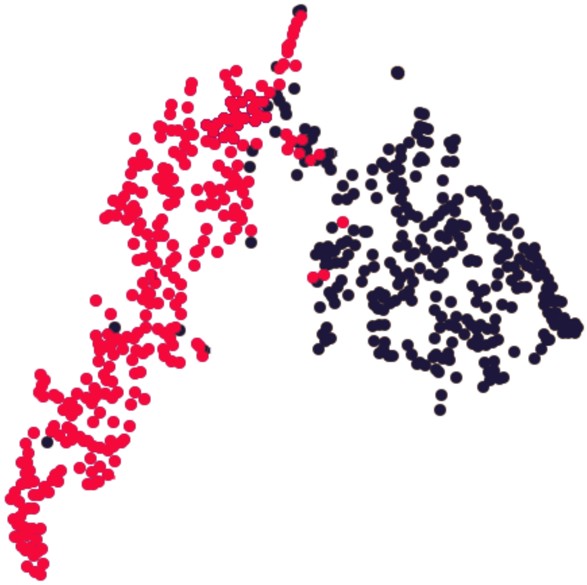

FIGURE 5. Feature embeddings with T-SNE for some retinal images under binary classification.

TABLE 3. Comparison with supervised models (%).

Model	accuracy	Sensitivity	Specificity
SE-ResNeXt50 [31]	92.4	87.1	98.2
EfficientNet [32]	90.7	80.7	97.7
EnsembleNet [33]	98.6	99.1	99.1
Resnet-50	94.4	96.9	91.05
Vgg-16	92.36	92.58	92.01
Inception-V3	92.22	93.95	89.77
SAGN (Our)	94.4	84.0	82.2

F. PARAMETER ANALYSIS

In this section, we also evaluate the influence of hyperparameters in SAGN.

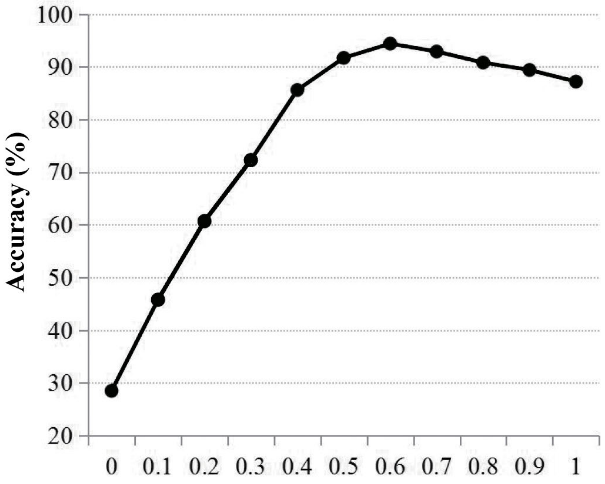

FIGURE 6. DR grading performance with different value of parameter α.

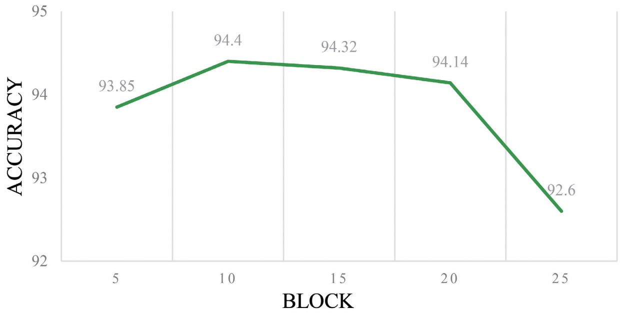

FIGURE 7. DR grading performance with different numbers of block.

1) INFLUENCE OF BALANCE PARAMETER α

We first analyze the influence of the balance parameter α (Eq.7). Specifically, the DR grading performance is discussed when the balance parameter α changes in $[0 : 0.1 : 1]$. AS illustrated in Figure 6, our SAGN obtains accuracy 28.5% when $\alpha = 0$, which denotes removing the semi-supervised cross-entropy loss L_{sce}, and only optimizing the network by decoding loss L_{dec}. This proves that the semi-supervised cross-entropy loss contributes a considerable improvement (65.9%) on DR grading accuracy. When we set $\alpha = 1$, that means removing the decoding loss L_{dec}, and it drops accuracy of 7.2%. That elaborates the decoding loss contributes 7.2% improvement in accuracy.

2) INFLUENCE OF THE NUMBER OF RESIDUAL BLOCK

We then discuss the influence of the number of residual blocks on the model performance. As shown in Figure 7, the model's performance improvements as the number of blocks gradually increase, which means that with the increase of effective parameters, the robustness of the model has been improved. However, when the number of blocks is greater than 10, the model's performance begins to decrease, which means that as the depth of the network increases, the redundant parameters increase, and the complexity of the model increases, resulting in a decrease in model performance.

V. CONCLUSION

In order to solve complicated annotating work in diabetic retinopathy grading tasks, this paper proposes a semi-supervised auto-encoder graph network to extract robust feature representations from limited labeled retinal images and sufficient unlabeled data. In detail, it firstly learns CNN features by an encoder-decoder CNN architecture, trained from both labeled and unlabeled retinal images, and then exploits the neighbor correlations based on CNN features across labeled and unlabeled images. Finally, the graph representation module utilizes the CNN features and their correlations to predict the DR grades. With the help of sufficient unlabeled images, SAGN can achieve performable grading accuracy with fewer labeled retinal images. The extensive experiments also demonstrate excellent performance on the semi-supervised DR grading task.

REFERENCES

[1] H. Tsujinaka, J. Fu, J. Shen, Y. Yu, Z. Hafiz, J. Kays, D. McKenzie, D. Cardona, D. Culp, W. Peterson, B. C. Gilger, C. S. Crean, J.-Z. Zhang, Y. Kanan, W. Yu, J. L. Cleland, M. Yang, J. Hanes, and P. A. Campochiaro, "Sustained treatment of retinal vascular diseases with self-aggregating sunitinib microparticles," *Nature Commun.*, vol. 11, no. 1, pp. 1–13, Dec. 2020.

[2] M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. Van Ginneken, J. C. Folk, and M. D. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 30, no. 11, pp. 1941–1950, Nov. 2011.

[3] S. G. Vázquez, B. Cancela, N. Barreira, M. G. Penedo, M. Rodríguez-Blanco, M. P. Seijo, G. C. de Tuero, M. A. Barceló, and M. Saez, "Improving retinal artery and vein classification by means of a minimal path approach," *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 919–930, Jul. 2013.

[4] T. Na, J. Xie, Y. Zhao, Y. Zhao, Y. Liu, Y. Wang, and J. Liu, "Retinal vascular segmentation using superpixel-based line operator and its application to vascular topology estimation," *Med. Phys.*, vol. 45, no. 7, pp. 3132–3146, Jul. 2018.

[5] V. S. Joshi, J. M. Reinhardt, M. K. Garvin, and M. D. Abramoff, "Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks," *PLoS ONE*, vol. 9, no. 2, Feb. 2014, Art. no. e88061.

[6] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1958–1967, Mar. 2021.

[7] Q. Yan, B. Chen, Y. Hu, J. Cheng, Y. Gong, J. Yang, J. Liu, and Y. Zhao, "Speckle reduction of OCT via super resolution reconstruction and its application on retinal layer segmentation," *Artif. Intell. Med.*, vol. 106, Jun. 2020, Art. no. 101871.

[8] Y. Chai, H. Liu, and J. Xu, "A new convolutional neural network model for peripapillary atrophy area segmentation from retinal fundus images," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105890.

[9] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, and R. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[10] J. Krause, V. Gulshan, E. Rahimy, P. Karth, K. Widner, G. S. Corrado, L. Peng, and D. R. Webster, "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *Ophthalmology*, vol. 125, no. 8, pp. 1264–1272, Aug. 2018.

[11] L. Pasa and A. Sperduti, "Pre-training of recurrent neural networks via linear autoencoders," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 3572–3580.

[12] X. Lu, S. Matsuda, C. Hori, and H. Kashioka, "Speech restoration based on deep learning autoencoder with layer-wised pretraining," in *Proc. 13th Annu. Conf. Int. Speech Commun. Assoc.*, 2012, pp. 1504–1507.

[13] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 1–38, 2010.

[14] C. You, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, Y. Zhang, W. Cong, and G. Wang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, vol. 6, pp. 41839–41855, 2018.

[15] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, and S. Ju, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2019.

[16] C. You, L. Yang, Y. Zhang, and G. Wang, "Low-dose CT via deep CNN with skip connection and network-in-network," *Proc. SPIE*, vol. 11113, Sep. 2019, Art. no. 111131W.

[17] C. Playout, R. Duval, and F. Cheriet, "A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2434–2444, Oct. 2019.

[18] D. Marin, M. E. Gegundez-Arias, B. Ponte, F. Alvarez, J. Garrido, C. Ortega, M. J. Vasallo, and J. M. Bravo, "An exudate detection method for diagnosis risk of diabetic macular edema in retinal images using feature-based and supervised classification," *Med. Biol. Eng. Comput.*, vol. 56, no. 8, pp. 1379–1390, Aug. 2018.

[19] X. Xu, W. Ding, M. D. Abrámoff, and R. Cao, "An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image," *Comput. Methods Programs Biomed.*, vol. 141, pp. 3–9, Apr. 2017.

[20] K. A. Sreeja and S. S. Kumar, "Automated detection of retinal hemorrhage based on supervised classifiers," *Indonesian J. Electr. Eng. Informat.*, vol. 8, no. 1, pp. 140–148, Mar. 2020.

[21] C. You, J. Yang, J. Chapiro, and J. S. Duncan, "Unsupervised wasserstein distance guided domain adaptation for 3D multi-domain liver segmentation," in *Interpretable and Annotation-Efficient Learning for Medical Image Computing*. Cham, Switzerland: Springer, 2020, pp. 155–163.

[22] C. You, R. Zhao, L. Staib, and J. S. Duncan, "Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation," 2021, *arXiv:2105.07059*. [Online]. Available: https://arxiv.org/abs/2105.07059

[23] R. Bakalo, J. Goldberger, and R. Ben-Ari, "Weakly and semi supervised detection in medical imaging via deep dual branch net," *Neurocomputing*, vol. 421, pp. 15–25, Jan. 2021.

[24] S. Han, S. I. Hwang, and H. J. Lee, "A weak and semi-supervised segmentation method for prostate cancer in trus images," *J. Digit. Imag.*, vol. 33, pp. 1–8, Feb. 2020.

[25] S. Menon, D. Chapman, P. Nguyen, Y. Yesha, M. Morris, and B. Saboury, "Deep expectation-maximization for semi-supervised lung cancer screening," 2020, *arXiv:2010.01173*. [Online]. Available: https://arxiv.org/abs/2010.01173

[26] A. C. Good and W. G. Richards, "Rapid evaluation of shape similarity using Gaussian functions," *J. Chem. Inf. Comput. Sci.*, vol. 33, no. 1, pp. 112–116, Jan. 1993.

[27] Aptos. (2019). *Blindness Detection*. [Online]. Available: https://www.kaggle.com/c/aptos2019-blindness-detection/

[28] M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0217541.

[29] 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.

[30] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[31] J. Hu, L. Shen, and G. Sun, "Squeeze- and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[32] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[33] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," in *Proc. 9th Int. Conf. Pattern Recognit. Appl. Methods (ICPRAM)*, M. D. Marsico, G. S. di Baja, and A. L. N. Fred, Eds. Valletta, Malta: SCITEPRESS, Feb. 2020, pp. 501–509.

YUJIE LI received the B.S. degree from the School of Computer, Ludong University, Yantai, China, in 2008, and the B.S. and M.S. degrees from the Department of Computer Engineering, Wonkwang University, Iksan, South Korea, in 2008 and 2010, respectively. She is currently pursuing the Ph.D. degree with the Department of Computer Software Engineering, Wonkwang University. She is also a Lecturer with the College of Computer Science and Engineering, Weifang University of Science and Technology, Weifang, China. Her current research interests include image processing, deep learning, and pattern recognition.

ZHANG SONG received the bachelor's degree in clinical medicine and the master's degree in pediatrics from Qingdao University, in 2014 and 2017, respectively. She is currently a Pediatric Physician with The Affiliated Hospital of Qingdao University, Qingdao, China. Her main research interests include blood system diseases and tumors.

SUNKYOUNG KANG received the bachelor's and Ph.D. degrees from the Department of Computer Engineering, Wonkwang University, Iksan, South Korea, in 2000 and 2010, respectively. From 2010 to 2017, she was a Research Director of Good Information Technologies Company Ltd. Since 2017, she has been a Professor with the Department of Computer Software Engineering, Wonkwang University. Her current research interests include image processing and big data.

SUNGTAE JUNG received the M.S. and Ph.D. degrees from the Department of Computer Engineering, Seoul National University, South Korea, in 1989 and 1994, respectively. Since 1995, he has been a Professor with the Department of Computer Software Engineering, Wonkwang University. His current research interests include image processing, machine learning, and computer graphics.

WENPEI KANG is currently pursuing the bachelor's degree with Southwest University. His main research interests include artificial intelligence and image recognition.