OPEN-SET OCT IMAGE RECOGNITION WITH SYNTHETIC LEARNING

Yuting Xiao[1] *Shenghua Gao*[1] *Zhengjie Chai*[1.2] *Kang Zhou*[1,2]
Tianyang Zhang[2] *Yitian Zhao*[2] *Jun Cheng*[3] *Jiang Liu*[4]

[1] ShanghaiTech University
[2] Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences
[3] UBTech Research
[4] Southern University of Science and Technology

ABSTRACT

Due to new eye diseases discovered every year, doctors may encounter some rare or unknown diseases. Similarly, in medical image recognition field, many practical medical classification tasks may encounter the case where some testing samples belong to some rare or unknown classes that have never been observed or included in the training set, which is termed as an open-set problem. As rare diseases samples are difficult to be obtained and included in the training set, it is reasonable to design an algorithm that recognizes both known and unknown diseases. Towards this end, this paper leverages a novel generative adversarial network (GAN) based synthetic learning for open-set retinal optical coherence tomography (OCT) image recognition. Specifically, we first train an auto-encoder GAN and a classifier to reconstruct and classify the observed images, respectively. Then a subspace-constrained synthesis loss is introduced to generate images that locate near the boundaries of the subspace of images corresponding to each observed disease, meanwhile, these images cannot be classified by the pre-trained classifier. In other words, these synthesized images are categorized into an unknown class. In this way, we can generate images belonging to the unknown class, and add them into the original dataset to retrain the classifier for the unknown disease discovery.

Index Terms— Open-set, Generative Adversarial Network, Subspace-constrained Synthesis Loss

1. INTRODUCTION

Retinal optical coherence tomography (OCT) image recognition is an important task in the medical imaging field, and the deep learning methods have achieved excellent performance for OCT image recognition [1, 2]. However, to our best knowledge, most of the current researches focus on the closed-set setting where the classes in training and that in testing are the same. In practical clinical applications, encountering eye diseases classes that have never been included [3] in training is a common situation because the rare presence of some diseases leads to the unavailability of those diseases in the training set. Thus it is desirable for an OCT image recognition system to simultaneously recognize images of the observed/known categories and label those images that have never seen before as the unknown category, which is referred to as an open-set[4, 5] problem.

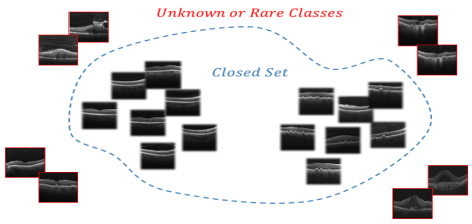

Fig. 1. The closed set (Blue Circle) is the training set. The testing set contains classes in the closed set and unknown or rare classes (Red)

Although the open-set problem is a common problem in both the medical field and computer vision field, very few works have been done in this area. The probability threshold method classifies an image x as the unknown class by predicting its probabilities belonging to known classes $y_i (i = 1, 2, ...N)$. If the maximum probability $\max P(y_i|x)$ is low enough. Recently, Generative Adversarial Network (GAN) [6] has emerged and demonstrated its capability of synthesizing unseen images[7, 8]. Further, in [9], a counterfactual image generation framework is proposed to synthesize images of unseen classes. But in their solution, a latent vector is synthesized by minimizing the pairwise distance between the synthesized latent vector and a latent vector of an image in the training set. Thus the distribution of the synthesized images may be mixed with the distribution of training images, which makes the consequent classification difficult. Further, their solution is designed for general image rather than especially

This work is supported by the ShanghaiTech-UnitedImaging Joint Lab and the ShanghaiTech-Megavii joint lab. And this work is also supported in part by the National Natural Science Foundation of China (NSFC) under Grants No. 61932020.

for OCT images, the two-dimensional or three-dimensional structure of biological tissue can be obtained by detecting the reflection or scattering signals of the incident weak coherent light at different depth layers of biological tissue. The structural information of the retina is relatively fixed, which would further facilitate the image generation.

Motivated by the success of the counterfactual image generation framework[9] for synthesizing images, in this paper, we propose a GAN based OCT image synthesizing framework for open-set classification. Specifically, we first pretrain a GAN for generating real images in the training set, and a classifier for classifying the real images in the training set. Then rather than using the pairwise synthesis loss in [9], we design a subspace-constrained synthesis loss, which makes the synthesized images locates at the boundaries of the training samples, then we add the synthesized images into the training set and retrain the classifier for open-set OCT image recognition.

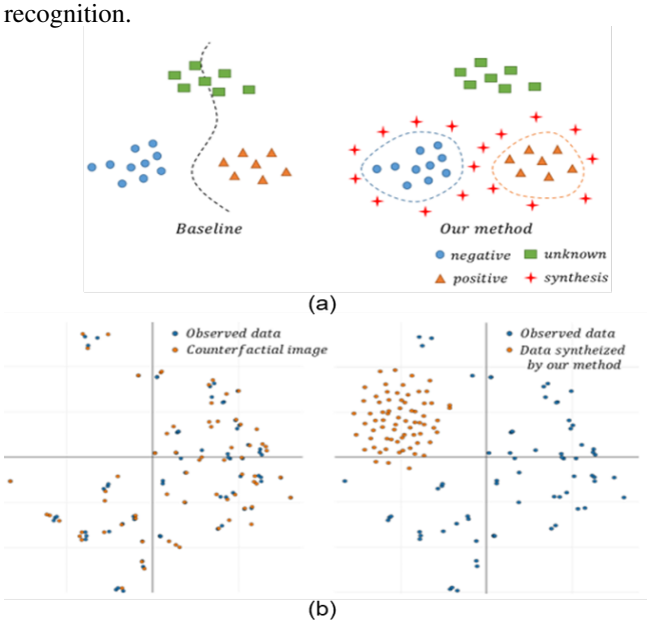

Fig. 2. (a) A typical classification method can only achieve good performance on the classes in the training set (Blue and Orange), but cannot recognize the unknown class (Green) well. The synthesized image (Red) around the known classes distribution enable classifier to recognize the unknown class. (b) Experimental visualization by t-SNE [10]. Synthesized latent vectors are projected onto a 2D space.

We utilize t-SNE in [10] to reduce the dimension of synthesized latent vectors and demonstrate projected points in 2D space in Fig. 2(b). Fig. 2(b) shows the advantage of our method over the counterfactual images method. The synthesized latent vectors of our method are well separated from those in the training set, while they are blended in the counterfactual images method, which demonstrates the effectiveness of our subspace-constrained synthesis loss over that in counterfactual images [9].

We summarize the contribution of this work as follows: (1) a human-mimic open-set OCT image recognition algorithm is proposed which would push the OCT image analysis towards the real application. (2) we propose to leverage a GAN based image synthesizing framework to synthesize images of the unknown class. Meanwhile, a subspace constrained synthesis loss is proposed for synthesizing images that fall out of the distribution of the training data, which consequently improves the classification accuracy; (3) extensive experiments and ablation studies on both the Cell dataset and the BOE dataset validate the effectiveness of our solution for open-set OCT image recognition.

2. METHOLODY

To tackle the open-set OCT image recognition problem, we propose to synthesize the images corresponding to unknown diseases and add the synthesized images corresponding to unknown diseases into the original training set to train a classifier for open-set OCT image recognition. The whole pipeline is shown in Figure Fig. 3. Specifically, we first map an image into a low dimensional latent vector space S with an encoder E. For each sample in S, we can map it back to an OCT image with a GAN model which consists of a generator G and a discriminator D. Then we can pretrain the encoder E and the GAN model. We also pretrain a classifier C to classify all training samples into the observed classes. For OCT images corresponding to unknown diseases, because OCT images have a relatively fixed structure, even with different diseases, they should also correspond to some vectors in the latent vector space S. Therefore we introduce a subspace constrained synthesis loss to learn the latent vectors corresponding to images belong to the unknown class by leveraging the pre-trained generator and the classifier. After that, we map the latent vectors of unknown diseases to the image space with the generator and add the synthesized images into the training set to retrain a classifier. In this way, the new classifier can classify images from the observed classes as well as unknown classes.

2.1. The GAN Model

Compared with the general object image, OCT images are highly structured even with diseases, which facilitates its generation. By following the work for image generation [9], we first propose to encode an input image x into a low dimensional vector $z = E(x)$. Each vector in the low dimensional space can be mapped back to an OCT image with a generator G. To make the generated image looks more real, a discriminator D is introduced to encourage the generated and the real images to be indistinguishable. Further, to avoid the ubiquitous mode collapsing in GAN model, following [11], for an image generated from a latent vector which is mapped from a real image, a pixel-wise similarity loss is imposed on the generator $||x - G(E(x))||_2$. Then we arrive at the following the objective functions for the generator and the discriminator:

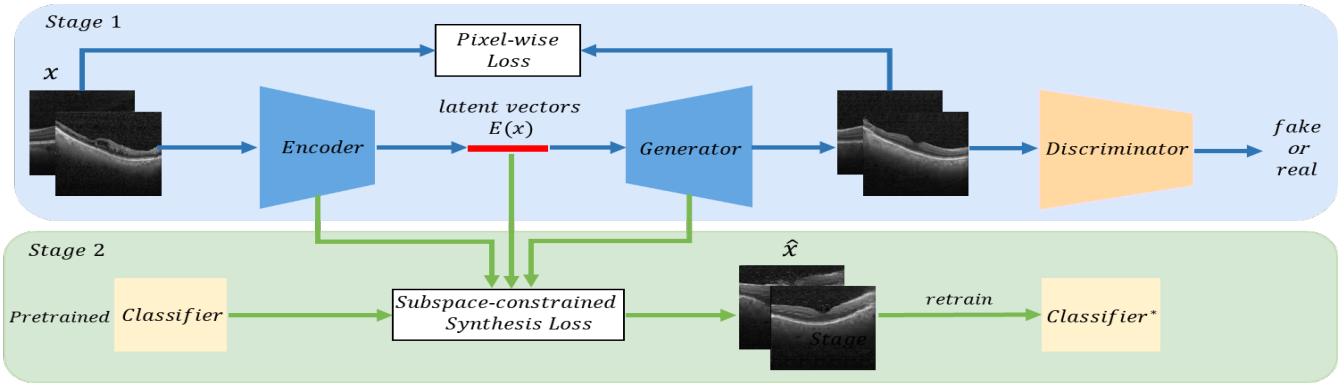

Fig. 3. The overview of the framework for synthesizing OCT images. The process can be divided into two stages:(1) The training of the auto-encoder GAN (Blue). (2) The subspace-constrained synthesis loss optimization for image generation (Green). The output latent vector $\hat{z}$ will be mapped into image space by the generator.

$$\mathcal{L}_{EG}(E,G) = \|x - G(E(x))\|_2 - D(G(E(x))) \quad (1)$$

$$\mathcal{L}_D(D) = D(G(E(x))) - D(x) \quad (2)$$

The encoder, the generator, and the discriminator can be optimized alternatively with the observed training images. In our implementation, the encoder of our network consists of four downsampling layers, As that in [12], and the generator also shares the same structure with the decoder in [12], except that the skip connections between the encoder and decoder in [12] are removed in our implementation. The discriminator D takes the same architecture of that in the patchGAN [13].

2.2. Generating Images from the Unknown Class

On the one hand, even though the relatively fixed structure of OCT images makes the latent vectors share some similarities, for those latent vectors corresponding to unknown diseases, it is desirable that they locate around the boundaries of the subspace formed by the observed training samples, which makes them distinguishable from the observed classes; On the other hand, the images generated from these latent vectors corresponding to unknown diseases should not be classified into the known classes by the pre-trained classifier, which means the probabilities corresponding to all N known categories estimated by the classifier should be low. By taking these properties into consideration, we propose a subspace-constrained synthesis loss for learning the latent vectors of unknown diseases, thus for a latent vector corresponding to an unknown disease, we introduce the following subspace-constrained synthesis loss:

$$\min_{z,\alpha} \|z - Z\alpha\|_2^2 - k\sum_{j=1}^{N} \log(C(G(z))_j) + F(\alpha) \quad (3)$$

Here $Z = [z_1, z_2, \ldots, z_n]$, and $z_i = E(x_i)$. Z corresponds to the latent feature vectors of training samples. Linear coefficients α are constrained by constraint term $F(\alpha)$:

$$F(\alpha) = \sigma\|1 - \mathbf{1}^T\alpha\|_1 + \delta\|\alpha\|_1 \quad (4)$$

where $\sigma\|1 - \mathbf{1}^T\alpha\|_1$ is used to ensure the linear superposition of Z still falling into the distribution of observed classes, and $\|\alpha\|_1$ encourages the coefficients α to be a sparse vector. The coefficients k, σ, δ are the weights of the corresponding terms.

The first term of the objective function measures the distance between the latent vector of a synthesized image z and the linear combination of latent vectors of training images, and the second term makes probabilities $C(G(z))_j$ estimated by the pre-trained classifier low. Both $\log(C(G(z))_j)$ and $\delta||\alpha||_1^1$ push feature vector z towards the boundaries of known classes distribution.

In our implementation, in each batch, we randomly sample some training images and optimize the objective function to obtain the target latent vector $\hat{z}$ by a gradient descent-based optimization method. After we get a latent vector $\hat{z}$, the generator G is used map it to an image $\hat{x} = G(\hat{z})$, which is added into the training set as an image from an unknown class. Then we retrain the classifier on this new training set. In this way, the classifier can recognize images of both the observed training classes and some unknown classes.

3. EXPERIMENT

Our experiments are evaluated with two public OCT datasets: (1) the Cell OCT dataset, which contains OCT images of 4 categories. (2) the BOE dataset, which contains OCT images of 3 categories.

3.1. Datasets

3.1.1. The Cell dataset [1].

There are 4 categories of images in the Cell dataset: Normal, Choroidal NeoVascularisation (CNV), Diabetic Macular

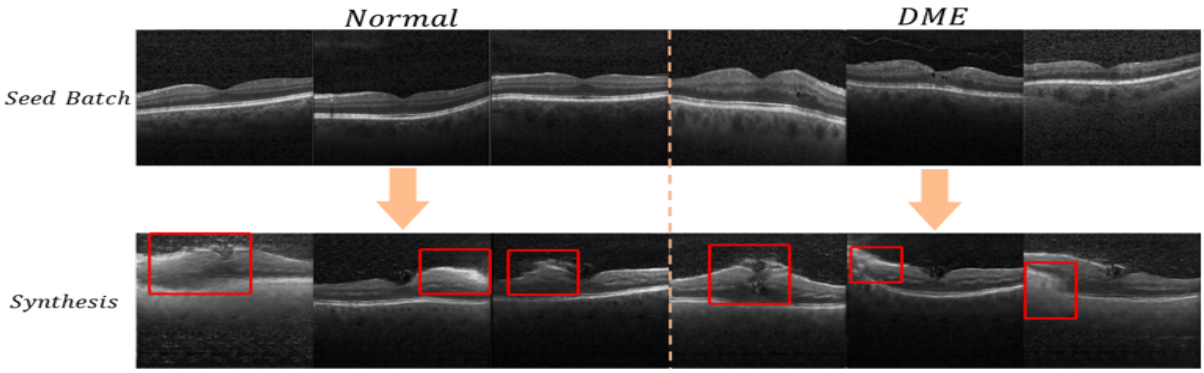

Fig. 4. The unpaired synthesized samples from the Cell dataset. A batch of synthesized images is generated from a batch of images of known class. Different and reasonable lesion areas are generated (Red box)

Table 1. Quantitative experiments accuracy results of different methods

Method	Cell		BOE	
	Normal(%)	DME(%)	Normal(%)	AMD(%)
Baseline	98.40	94.80	94.85	**95.63**
Threshold($\lambda = 0.6$)	98.40	94.80	94.50	94.54
Threshold($\lambda = 0.7$)	98.20	94.60	91.41	93.44
Threshold($\lambda = 0.8$)	98.20	94.20	86.60	89.62
Threshold($\lambda = 0.9$)	98.00	93.80	80.41	80.33
Counterfactual Images	98.80	97.40	72.16	**95.63**
Ours	**99.00**	**100.00**	**96.22**	91.80

Table 2. Quantitative experiments AUC results of different methods

Method	AUC	
	Cell	BOE
Threshold	0.7901	0.5194
Counterfactual Images	0.8192	0.5891
Ours	**0.8578**	**0.6851**

Edema (DME) and Drusen. We use images of 3 categories which are Normal, DME and Drusen to do experiments. We utilize Normal and DME class images as the training set, and 500 images of Normal, DME and Drusen classes as the testing set. The Drusen class is defined as the unknown class.

3.1.2. The BOE dataset [14].

There are 3 categories of images in BOE dataset: Normal, Age-related Macular Degeneration (AMD) and Diabetic Macular Edema (DME). Each class consists of 15 volumes corresponding to 15 different patients. We use 9 volumes of Normal and 9 volumes of AMD categories as the training set. 3 volumes of every category as validation set and 3 volumes of every category as the testing set.

3.2. Results

The evaluation metrics in this paper follow that in [9], which computes the accuracy of closed-set classes and Area Under the ROC Curve (AUC) for a binary classification on the known class and the unknown class. We utilize these metrics because the samples of rare and unknown classes are not the main target for a classifier in the medical image, so the classification accuracy of the closed set should be stable and the ability for detecting unknown class are evaluated by AUC.

Table 1 shows the quantitative closed-set accuracy analysis of the proposed method. It is obvious that our method can achieve better the performance of classification for observed classes than other methods, which is meaningful in practical applications. Fig. 4 shows the qualitative analysis of synthesized images, we can see that the structure information of retinal OCT images is well preserved.

By gradually thresholding the probabilities estimated by the softmax classifier, we can compute the percentage of images being classified as an image of the unknown class but in fact, belonging to some known class (false positive) and the percentage that is classified into an unknown class and also belonging to the unknown class (true positive). Then we can draw the Receiver Operating Characteristic (ROC) curve and compute AUC according to ROC curve. The results are shown in Tabel 2. We can see that our achieves the best performance, which further demonstrates the capability of our approach for the unknown class discovery. It also validates the effectiveness of our method over counterfactual image [9].

4. CONCLUSIONS

This paper represents a GAN based image synthesis approach for open-set OCT image recognition. By generating images from an unknown class with a novel subspace-constrained synthesis loss and retraining the classifier with both training images and synthesized images, our method can simultaneously recognize images from both the observed classes and the unknown class.

5. REFERENCES

[1] Daniel S Kermany and et.al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[2] Yitian Zhao, Lavdie Rada, Ke Chen, Simon P Harding, and Yalin Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE transactions on medical imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.

[3] Thomas G Dietterich, "Steps toward robust artificial intelligence," *AI Magazine*, 2017.

[4] Walter J Scheirer and et.al., "Toward open set recognition," *IEEE TPAMI*, vol. 35, no. 7, pp. 1757–1772, 2013.

[5] Walter J Scheirer and et.al., "Probability models for open set recognition," *IEEE TPAMI*, vol. 36, no. 11, pp. 2317–2324, 2014.

[6] Ian Goodfellow and et.al., "Generative adversarial nets," in *NeurIPS*, 2014.

[7] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan, "Synthetic data augmentation using gan for improved liver lesion classification," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 289–293.

[8] Faisal Mahmood, Richard Chen, and Nicholas J Durr, "Unsupervised reverse domain adaptation for synthetic medical images via adversarial training," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2572–2581, 2018.

[9] Lawrence Neal and et.al., "Open set learning with counterfactual images," in *ECCV*, 2018.

[10] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," 2008.

[11] Tim Salimans and et.al., "Improved techniques for training gans," in *NeurIPS*, 2016, pp. 2234–2242.

[12] Olaf Ronneberger and et.al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[13] Phillip Isola and et.al, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[14] Pratul P Srinivasan and et.al., "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *BOE*, 2014.