Semi-supervised Adversarial Learning for Diabetic Retinopathy Screening

Sijie Liu, Jingmin Xin(✉), Jiayi Wu, and Peiwen Shi

Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China
jxin@mail.xjtu.edu.cn

Abstract. It is well known that in medical image analysis, only a small number of high-quality labeled images can be often obtained from a large number of medical images due to the requirement of expert knowledge and intensive labor work. Therefore, we propose a novel semi-supervised adversarial learning framework (SSALF) for diabetic retinopathy (DR) screening of color fundus images. Specifically, our proposed framework consists of two subnetworks, an extended network and a discriminator. The extended network is obtained by extending a common classification network with a generator used for unsupervised image reconstruction. Thus, the extended network can utilize some labeled and lots of unlabeled fundus images. Then the discriminator is attached to the generator of the extended network to judge whether a reconstructed image is real or fake, introducing adversarial learning into the whole framework. Our framework achieves promising utility and generalization on the datasets of EyePACS and Messidor in a semi-supervised setting: we use some labeled and lots of unlabeled fundus images to train our framework. And we also investigate the effects of image reconstruction and adversarial learning on our framework by implementing ablation experiments.

1 Introduction

In many countries, diabetic retinopathy (DR) is the most common cause of blindness in adults. Fortunately, early diagnosis and timely treatment can effectively prevent the occurrence of blindness. With the development of color fundus photography, experienced ophthalmologists can observe various DR lesions in fundus images, rate the severity of DR, and decide corresponding treatments. To reduce the burden of ophthalmologists, various automatic DR screening methods [1] have been proposed. Recently, deep learning has become a leading methodology for medical image analysis and also has achieved promising performance [4] in DR screening. As we all know, a superior deep neural network usually involves large numbers of medical images with corresponding high-quality annotations. However, the process of obtaining these annotations not only is time-consuming, but also requires large amounts of expert knowledge. Hence, it is a challenging task to use a small number of labeled fundus images to achieve superior performance of DR screening.

H. Fu et al. (Eds.): OMIA 2019, LNCS 11855, pp. 60–68, 2019.
https://doi.org/10.1007/978-3-030-32956-3_8

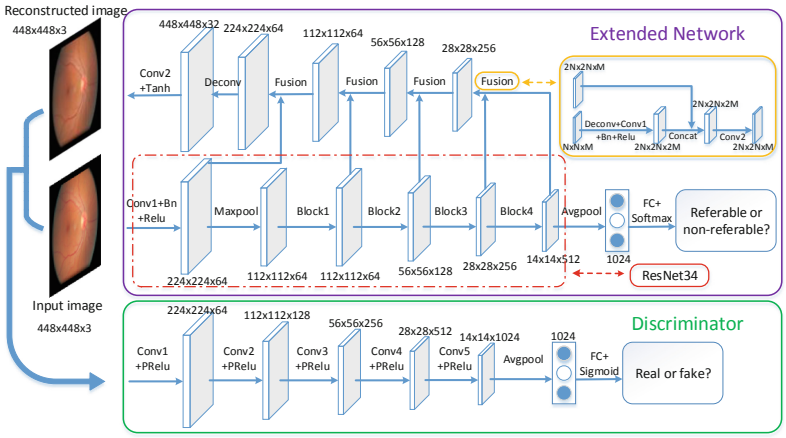

Fig. 1. Schematic of our semi-supervised adversarial learning framework (SSALF) for DR screening. Components of the extended network and the discriminator are in the purple and green solid boxes. ResNet34 and 'Fusion' are also showed in the red and yellow dotted boxes, where '**Block**' represents multiple stacked residual modules. (Color figure online)

Many works have been conducted to address the relative tasks [10,12,14,18,19]. To reduce the number of images that need to be labeled, Yang et al. [18] exploited deep active learning to select the most effective medical images to be labeled. To utilize some labeled and lots of unlabeled images, Ladder network [12] and SWWAE [19] were proposed to simultaneously minimize the sum of the classification term and the reconstruction term in a semi-supervised setting. As the generative adversarial network (GAN) [3] becomes a research hotspot in semi-supervised and unsupervised learning, many researchers [6,10,14] proposed the GAN-based classification networks, such as ImprovedGAN [14]. In these networks, researchers unified a discriminator of GAN and a classifier into a single network. The new discriminator could predict N+1 classes, where N means categories of medical images, and 1 means whether a medical image is real or fake. Thus, during the training phase, these networks were trained with labeled medical images to predict N classes, and were trained with unlabeled and generated medical images to judge whether a medical image is real or fake. Lecouat et al. [8] proposed a patch-based semi-supervised classification approach to recognize abnormal fundus images, which was based on ImprovedGAN. TripleGAN [2] proposed a tripartite adversarial model, including three separated networks: a classifier, a generator and a discriminator. In fact, TripleGAN divided a discriminator of ImprovedGAN into two parts, the discriminator of TripleGAN and the classifier of TripleGAN, while adding an adversarial mechanism between the two parts. In these mentioned GAN-based methods, all the relationships between a generator and a classifier/a discriminator were cascading.

However, it is also worthy of further study for DR screening to unify a classifier and a generator of GAN into a single network. Since the common GAN generator [10,11,14] takes noise as input but the common classification network a image, it is necessary to ensure they have the same input. Thus, a GAN-based image reconstruction network [9] (a variant of GAN), where the generator reconstructs input images as much as possible while the discriminator strives to distinguish between input images and reconstructed images, attracts our attention. Naturally and intuitively, we attempt to extend a common classification network by combining a GAN-based image reconstruction network.

Therefore, we propose a novel semi-supervised adversarial learning framework (SSALF) for diabetic retinopathy (DR) screening of color fundus images. Our proposed framework consists of two subnetwork, an extended network and a discriminator. In order to utilize some labeled and lots of unlabeled fundus images, we extend a common classification network for classification and reconstruction by U-net's "fusion" [13], and call it the extended network. Thus, the extended network comprises two components, a classifier for supervised classification and a generator for unsupervised image reconstruction. Then like a common GAN, we attach the discriminator to the generator of the extended network to judge whether a reconstructed image is real or fake, which introduces adversarial learning into the whole framework. In summary, our contribution are as follows: (i) We propose a novel semi-supervised adversarial learning framework for DR screening, extending a common classification network by combining a GAN-based image reconstruction network. (ii) We also propose an appropriate training strategy to effectively and efficiently train our framework. (iii) Our framework achieves promising utility and generalization on the datasets of EyePACS and Messidor in a semi-supervised setting: we use some labeled and lots of unlabeled fundus images to train our framework. We also investigate the effects of image reconstruction and adversarial learning on our framework by implementing ablation experiments.

2 Methods

2.1 Common Networks for Classification

With the rapid development of deep convolution neural networks (DCNNs), some common classification networks [5,7,15,16] were proposed successively. Nowadays, many works [17] modified these common networks according to specific medical image analysis tasks, and achieved promising results. Thus, considering convergence speed and memory overhead, we exploit ResNet34 [5] (denoted as C) as the base model, and use the binary cross entropy loss as a loss function:

$$\mathcal{L}_{CLS}^{sup}(C) = -(y \log C(\mathcal{X}) + (1-y) \log(1 - C(\mathcal{X})), \tag{1}$$

where y is a class label of an input image $\mathcal{X}$ from a supervised subset.

2.2 Semi-supervised Adversarial Learning Framework

Unlike the aforementioned GAN-based methods [2,6,8,10,14], our framework focuses on unifying a classifier and a generator of GAN into a single network, instead of dividing them into two cascading models. Our framework consists of two subnetworks, i.e., (1) an extended network for classification and reconstruction, and (2) a discriminator for introducing adversarial learning. For the extended network, we use ResNet34 as a backbone architecture (which is served as a supervised classifier (donated as C)) and extend it with an unsupervised generator (donated as G). Specifically, we use ResNet34 as our classifier, but the fully connected layer of ResNet34 is modified to output two-dimension values. Then we use the U-net's "fusion" (See Fig. 1) and deconvolution to upsample from the last convolutional layer of our classifier until the output size is consistent with the input image, which constructs our generator and makes it can combine low-level and high-level features to reconstruct input images. In our point of view, the generator not only acts as a regularizer [19], but also forces the classifier to focus on abstract invariant features on the higher level [12] by utilizing the "fusion". For the discriminator (donated as D), following the rules in [11], we design a simple seven-layer DCNN. We aim to add a regularization item, which can learn the distribution of real fundus images, to the extended network by introducing adversarial learning. Our framework schematic is depicted in Fig. 1, and is theoretically easy to deploy to the other aforementioned common networks.

Our pipeline is that the extended network outputs results of classification and reconstruction simultaneously, and that then the discriminator determines whether a fundus image is reconstructed by the extended network or not. Therefore, the classification and the GAN-based image reconstruction are unified into a single framework.

The total loss of the extended network is formulated as:

$$\mathcal{L}_{EXT}^{semi}(C,G) = \mathcal{L}_{CLS}^{sup}(C) + \mu(\mathcal{L}_{MSE}^{unsup}(G) + \lambda\mathcal{L}_{AD\_G}^{unsup}(G)), \tag{2}$$

where

$$\mathcal{L}_{MSE}^{unsup}(G) = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}(I_{x,y} - G(I_{x,y}))^2, \tag{3}$$

$$\mathcal{L}_{AD\_G}^{unsup}(G) = -\log D(G(\mathcal{X})), \tag{4}$$

where $\mathcal{L}_{MSE}^{unsup}(G)$ and $\mathcal{L}_{AD\_G}^{unsup}(G)$ indicate the image reconstruction loss of the generator and the adversarial loss of the generator in an unsupervised subset, respectively. $\mathcal{L}_{CLS}^{sup}(C)$ represents the binary cross entropy loss of the classifier in a supervised subset. μ and λ refer to the weighting coefficients of the unsupervised loss and the adversarial loss, respectively. And W and H denote the size of a input image $\mathcal{X}$ while $I_{x,y}$ indicates the image pixel value.

The adversarial loss of the discriminator is formulated as:

$$\mathcal{L}_{AD\_D}^{unsup}(D) = -(\log D(\mathcal{X}) + \log(1 - D(G(\mathcal{X})))). \tag{5}$$

2.3 Appropriate Training Strategy

Training a deep neural network is also non-trivial. Therefore, we propose the following steps to effectively and efficiently train our framework.

1. We use weights of ResNet34 pre-trained on the ILSVRC to initialize the extended network [17], and train it only for image reconstruction.
2. We use the weights trained in step 1 to reinitialize the extended network.
3. We fix the extended network, and then train the discriminator once with minimization of $\mathcal{L}_{AD\_D}^{unsup}(D)$.
4. We fix the discriminator, and then train the extended network once with minimization of $\mathcal{L}_{EXT}^{semi}(C, G)$.
5. We iterate the step 3 and 4 until the extended network converges.

Among the above steps, step 1 has been demonstrated to be quite effective in [17], and step 2 is extremely crucial for semi-supervised DR screening, which will be demonstrated in Sect. 3.2. Pytorch[1] is adopted to implement our proposed framework. Scaling radius, random crop, random translation, random rotation and random flip are applied to preprocess and augment our dataset. Besides, all the fundus images are resized to $448 \times 448 \times 3$. Our framework is trained on a Nvidia GTX 1080Ti of 11 GB memory with a batch size of 16. The Nesterov SGD algorithm with an initial learning rate of 1e−3, a momentum of 0.9 and a weight decay of 5e−4 is used to optimize the extended network and the discriminator during the training. μ is set as 50 initially and will be reduced later in order to keep the ratio of losses between the classifier and the generator more than 4:1. And λ is set as 3e-4.

3 Experiments

3.1 Dataset Description

Our framework is evaluated on two publicly available datasets: the dataset of 'Kaggle Diabetic Retinopathy Detection' (EyePACS)[2] and the Messidor dataset[3].

The EyePACS dataset contains 35,126 training images with graded labels and 53,576 test images without graded labels. The presence of the diabetic retinopathy in each image has been graded by a clinician into one of the five stages: no DR, mild, moderate, severe, and proliferative DR. Here we only focus on the non-referable DR stage (including the no DR stage and the mild stage) and the referable DR stage (including the moderate stage, the severe stage, and proliferative DR stage). We divide the training images into three subsets: kaggle-train (the first 21,076 images), kaggle-val (the middle 7026 images), and kaggle-test (the last 7026 images). In our semi-supervised setting, we randomly select 500,

[1] https://github.com/pytorch/pytorch.
[2] https://www.kaggle.com/c/diabetic-retinopathy-detection/data.
[3] http://www.adcis.net/en/Download-Third-Party/Messidor.html.

Table 1. AUC of different methods on the EyePACS dataset.

Method	500	1000	2000	3000
ResNet34	0.773	0.833	0.869	0.886
ImprovedGAN	**0.806**	**0.858**	0.876	0.890
SSALF (ours)	0.800	0.854	**0.883**	**0.900**

Table 2. Ablation experiments on the EyePACS dataset.

Method	500		1000		2000		3000	
	AUC	SSIM	AUC	SSIM	AUC	SSIM	AUC	SSIM
ResNet34	0.773	-	0.833	-	0.869	-	0.886	-
ResNet34+Rec*	0.751	0.748	0.828	0.841	0.871	0.892	0.887	0.905
ResNet34+Rec	0.791	0.891	0.847	0.892	0.879	0.929	0.895	**0.939**
SSALF (ours)	**0.800**	**0.893**	**0.854**	**0.910**	**0.883**	**0.939**	**0.900**	0.932

1000, 2000 and 3000 images from the kaggle-train as supervised subsets respectively. Meanwhile, we only use the entire kaggle-train as a unsupervised subset. These subsets are balanced by oversampling (random crop).

The Messidor dataset contains 1,200 color fundus images. Different from the EyePACS dataset, the Messidor dataset divides all the images into four stages. Similarly, we can obtain 699 non-referable fundus images and 501 referable fundus images from this dataset. Here we use the whole Messidor dataset as an independent dataset for test.

3.2 Experiment Results

We perform semi-supervised experiments on the datasets of EyePACS and Messidor. The area under the receiver operating curve (AUC) and the structural similarity index (SSIM) are used to quantify the performance of the classification and the image reconstruction, respectively.

EyePACS: To evaluate the performance of our proposed framework, we compare our framework with ResNet34 and ImprovedGAN [14], as shown in Table 1. To make a fair comparison, we adopt ResNet34 as the discriminator of ImprovedGAN. For the generator of ImprovedGAN, we use 200-dimension vectors as input and add several deconvolutional layers to the original version [14] in order to generate $448 \times 448 \times 3$ fundus images. It is observed in Table 1 that our SSALF trained with 500 or 1000 labeld fundus images can achieve comparable AUCs with ImprovedGAN while with the increase of labeld fundus images, our SSALF can achieve more improvements than ImprovedGAN. Furthermore, in the case of 2000 or 3000 labeled fundus images, ImprovedGAN only achieves a little improvement compared to ResNet34 while our SSALF doesn't show significant gain reduction.

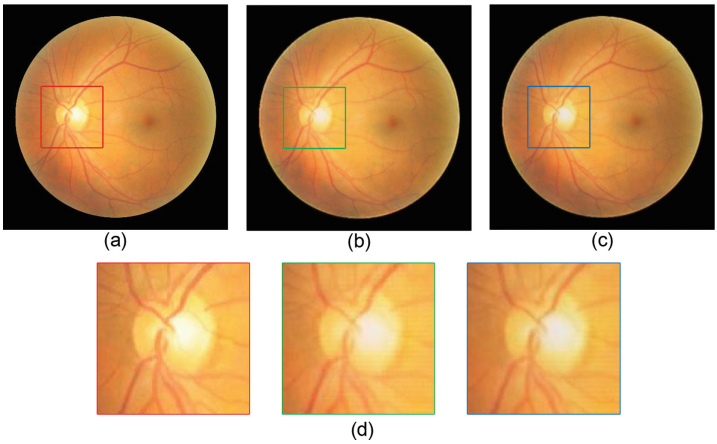

Fig. 2. Fake images generated from (a) an input image of the kaggle-test through (b) ResNet34+Rec and (c) our SSALF, where (d) optic disks are selected for comparison.

Table 3. AUC of different methods on the Messidor dataset.

Method	500	1000	2000	3000
ResNet34	0.817	0.892	0.923	0.932
ImprovedGAN	**0.907**	**0.922**	0.934	0.941
SSALF (ours)	0.877	0.910	**0.936**	**0.945**

To investigate the effects of the image reconstruction and the adversarial learning respectively, we conduct several ablation experiments, as shown in Table 2. ResNet34+Rec (ResNet34+Rec*) indicates the extended network with (without) the initialization in the aforementioned step 2. We can find in Table 2 that ResNet34+Rec* can only achieve the comparable results with ResNet34. (1) This shows that without good initialization in the aforementioned step 2, the generator can't provide good regularization for the classifier during the training. Particularly noting, the adversarial learning has no relationship with improving SSIM, which is also pointed out in [9]. Figure 2 displays the fake images generated from the kaggle-test by using ResNet34+Rec and our SSALF. A closer look reveals our SSALF produces thinner but clearer texture, especially texture in the optic disk. In Table 2 we can also find that with different numbers of labeled fundus images, ResNet34+Rec can achieve better AUC than ResNet34 while our SSALF achieves the best AUC. (2) This shows combining the image reconstruction can indeed improve the performance of DR screening dramatically, while introducing the adversarial learning can further enhance the performance.

Messidor: In order to demonstrate the generalization ability of our framework, we also evaluate it on the Messidor dataset, but only use this dataset for testing.

Results are shown in Table 3. It is observed that all the results from the Messidor dataset keep the same trend with those from the EyePACS dataset, and that our SSALF can even achieve an AUC of 0.945. This shows that our framework has good generalization ability.

4 Conclusions

In this paper, we propose a novel semi-supervised adversarial learning framework for diabetic retinopathy screening of color fundus images, and an appropriate training strategy. Experiment results on the datasets of EyePACS and Messidor show that our framework can achieve comparable or better utility and generalization than ImprovedGAN. Our ablation experiments show that combining the image reconstruction can indeed improve the performance dramatically, while introducing the adversarial learning can further enhance the performance.

Acknowledgements. This work was supported in part by the Programme of Introducing Talents of Discipline to University: B13043, and the National Key Research and Development Program of China under grant 2017YFA0700800.

References

1. Abràmoff, M.D., et al.: Automated early detection of diabetic retinopathy. Ophthalmology **117**(6), 1147–1154 (2010)
2. Chongxuan, L., Xu, T., Zhu, J., Zhang, B.: Triple generative adversarial nets. In: NIPS, pp. 4088–4098 (2017)
3. Goodfellow, I., et al.: Generative adversarial nets. In: NIPS. pp. 2672–2680 (2014)
4. Gulshan, V., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA **316**(22), 2402–2410 (2016)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
6. Hu, B., et al.: Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. IEEE J. Biomed. Health Inform. **23**(3), 1316–1328 (2018)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
8. Lecouat, B., et al.: Semi-supervised deep learning for abnormality classification in retinal images. In: Machine Learning for Health (ML4H) Workshop at NeurIPS (2018)
9. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR, pp. 4681–4690 (2017)
10. Madani, A., Moradi, M., Karargyris, A., Syeda-Mahmood, T.: Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. In: ISBI 2018, pp. 1038–1042. IEEE (2018)
11. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)

12. Rasmus, A., Berglund, M., Honkala, M., Valpola, H., Raiko, T.: Semi-supervised learning with ladder networks. In: NIPS, pp. 3546–3554 (2015)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
14. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANS. In: NIPS, pp. 2234–2242 (2016)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
16. Szegedy, C., et al.: Going deeper with convolutions. In: CVPR, pp. 1–9 (2015)
17. Vo, H.H., Verma, A.: New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space. In: 2016 IEEE International Symposium on Multimedia (ISM), pp. 209–215. IEEE (2016)
18. Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z.: Suggestive Annotation: a deep active learning framework for biomedical image segmentation. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 399–407. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_46
19. Zhao, J., Mathieu, M., Goroshin, R., Lecun, Y.: Stacked what-where auto-encoders. arXiv preprint arXiv:1506.02351 (2015)