Pathology-Aware Deep Network Visualization and Its Application in Glaucoma Image Synthesis

Xiaofei Wang[1], Mai Xu[1,2(✉)], Liu Li[1], Zulin Wang[1], and Zhenyu Guan[1]

[1] School of Electronic and Information Engineering, Beihang University, Beijing, China
MaiXu@buaa.edu.cn

[2] Hangzhou Innovation Institute, Beihang University, Beijing, China

Abstract. The past few years have witnessed the great success of applying deep neural networks (DNNs) in computer-aided diagnosis. However, little attention has been paid to provide pathological evidence in the existing DNNs for medical diagnosis. In fact, feature visualization in DNNs is able to help understanding how the computer make decisions, and thus it shows promise on finding pathological evidence from computer-aided diagnosis. In this paper, we propose a novel pathology-aware visualization approach for DNN-based glaucoma classification, which is used to locate the pathological evidence from fundus images for glaucoma. Besides, we apply the visualization framework to the glaucoma images synthesis task, through which specific pathological areas of synthesized images can be enhanced. Finally, experimental results show that the visualization heat maps can pinpoint different glaucoma pathologies with high accuracy, and that the generated glaucoma images are more pathophysiologically clear in rim loss (RL) and retinal neural fiber layer damage (RNFLD), which is verified by the ophthalmologist.

Keywords: Deep neural networks · Visualization · Glaucoma · Image synthesis

1 Introduction

The recent success of deep neural networks (DNNs) has benefitted medical diagnosis [1], especially for automatically detecting glaucoma in fundus images [2]. In addition to improving the classification accuracy, visualizing the DNNs have also attracted a lot of attention. In automatic medical diagnosis, it is critical to explain the behavior of a machine learning model and to make medical experts know the machine diagnosis. An effective visualization method should be able to point out all possible reasons for making decisions in DNNs, since many diseases have more than one typical pathologies, e.g., rim loss (RL) and retinal neural fiber layer damage (RNFLD) for glaucoma. However, in glaucoma image classification, there are only binary labels (positive or negative) annotated by

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 423–431, 2019.
https://doi.org/10.1007/978-3-030-32239-7_47

medical experts, which limits the existing visualization methods to pinpoint different pathological evidence hidden behind DNNs. To address such a problem, we propose a novel pathology-aware visualization method for glaucoma classification network. Firstly, feature maps of the classification network are selected by their activation values, through which the features with tiny activation values are screened out to reduce the noise effects in the visualization result. Secondly, the features are further screened and divided into two groups by their centroid-centric moment of inertia (CMI) values, consistent with the property of RL and RNFLD in glaucoma fundus images. Finally, the selected features are combined in a gradient-based weight to generate the final visualization heat map. Experimental results show that our pathology-aware visualization method can pinpoint both the RL and RNFLD area in glaucoma fundus images with high accuracy.

In addition, we further use feature visualization to glaucoma image synthesis, such that pathological area can be easily distinguished. Specifically, we propose combining the traditional generative adversarial network (GAN) and our pathology-aware visualization, called pathology-based GAN (Patho-GAN). The key idea of Patho-GAN is to enforce the synthesized images of GAN to have the similar visualization results to the reference image. After training the Patho-GAN, one can obtain the synthesized glaucoma images with pathophysiologically clearer RL and RNFLD, which are verified by the ophthalmologist.

The main contributions of this paper are: (1) we propose a novel pathology-aware visualization method to find different pathological evidence (RL and RNFLD) for glaucoma classification network; (2) we propose Patho-GAN to apply the feature visualization to the glaucoma image synthesis task, thus enhancing RL and RNFLD of the synthesized glaucoma images.

2 Literature Review

Visualization Methods: Recently, many visualization methods have been proposed to explain the reasoning behind the DNNs' decisions [3–7]. Specifically, in natural image domain, [3] proposed the occlusion experiment (Occ), in which they obtain a region-relevance heat map by occluding different portions of the input image and monitoring the output of the classifier. Zhou *et al.* [5] focused on the last convolutional layer and proposed Class Activation Map (CAM) method to combine the feature maps and generate the visualization heat map, while [4] aimed to find the evidence in the input space and proposed the guided backpropagation (GBP) method. Meanwhile, in medical domain, some specially designed visualization methods have also been proposed to better adapt the medical image properties. For instance, [7] modified the Occ by occluding the known pathologies of Alzheimer disease, while [6] focused on the intermediary feature maps and visualized the evidence for skin lesions. Nevertheless, the existing visualization methods are not able to pinpoint different kinds of decision evidence of DNNs with only simply annotated labels (positive or negative). To address such a problem, we propose a pathology-aware visualization method to find and visualize different pathological evidence for glaucoma classification network.

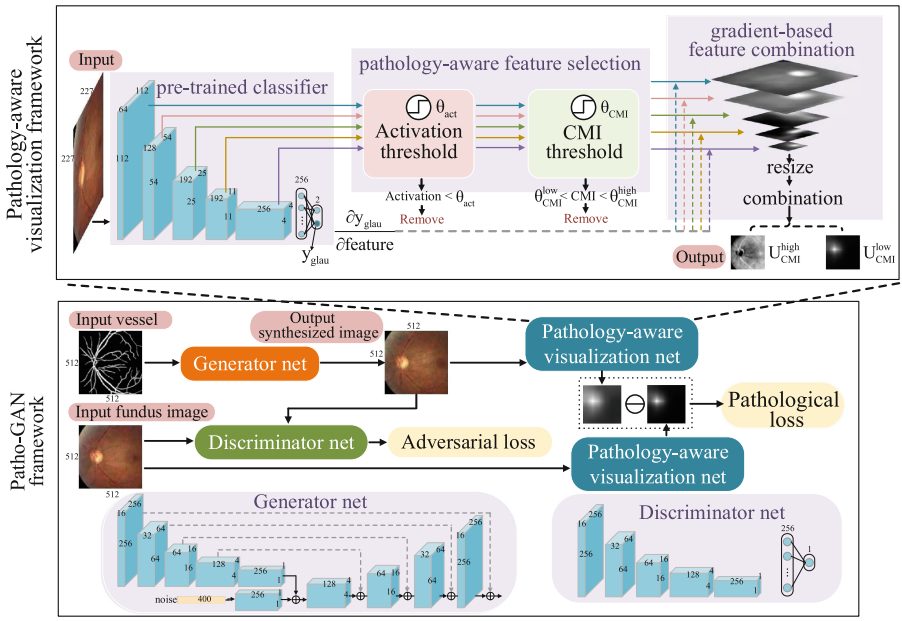

Fig. 1. Framework of pathology-aware visualization (top row) and Patho-GAN (bottom row).

Applications of Feature Visualization: Due to the good performance in finding pathological evidence of DNNs, feature visualization has recently been applied to many tasks, such as weakly supervised image classification [8], disease lesion localization [9], image registration [10] and segmentation [11]. However, different from the existing applications, we apply feature visualization to the medical image synthesis task, particularly for glaucoma in this paper, such that the pathological area can be further enhanced for the synthesized images.

3 Preliminaries

Database: The fundus images used to train the DNNs in this paper were obtained from the Large-scale Attention based Glaucoma (LAG) database [12]. The LAG database contains 5,824 fundus images with 2,392 positive and 3,432 negative glaucoma samples, which are randomly divided into training (4,792 images), validation (200 images) and test (832 images) sets. In a preprocessing step, the images are downscaled to a resolution of 280×280 pixels, and the RGB values are normalized between 0 and 1. We augment the training set by taking random crops of 227×227 pixels, and further augment each crop by rotating (angle sampled uniformly between 0 and 2π).

Ground Truth for Visualization: To evaluate the visualization results, 350 positive glaucoma images were randomly selected from the LAG test set and

further labeled by a professional ophthalmologist online. The ophthalmologist used boxes to locate particular pathologies of glaucoma, including RNFLD and RL. The label results are used as the ground truth of our visualization results.

4 Methodology

4.1 Pathology-Aware Feature Visualization

To visualize the pathological evidence of the glaucoma classification network, we propose a novel pathology-aware feature visualization approach. Our approach consists of two main procedures: pathology-aware feature selection and gradient-based combination. The pathology-aware feature selection includes two subprocess: (1) The features of a pre-trained classification network are selected by their activation values; (2) the features are further screened and divided into two groups by their centroid-centric moment of inertia (CMI) values. Finally, the two groups of features are combined separately in a gradient-based weight and then output the visualization heat maps. The framework is shown in Fig. 1.

Pathology-Aware Feature Selection: Firstly, in order to screen out the "useless" features with tiny activations, we select and retain the active feature maps $\mathbf{A}^{k,l} \in \mathbb{R}^{M^l \times N^l}$ (the k-th feature in the l-th layer, with height M^l and width N^l) in set $\mathbf{U}^l_{\text{act}}$:

$$\mathbf{U}^l_{\text{act}} = \{\mathbf{A}^{k,l} \mid \|\mathbf{A}^{k,l}\|_1 > \theta_{\text{act}} \cdot \max_n \|\mathbf{F}^{n,l}\|_1\}, \tag{1}$$

where $\mathbf{F}^{n,l}$ denotes the n-th feature in layer l and θ_{act} is a threshold of the activation value. Secondly, the features are further selected according to their CMI value, which measures the geometric distortion and brightness distortion of feature maps. Given a feature map $\mathbf{A}^{k,l}$, the CMI value is defined as $J_{k,l}$:

$$J_{k,l} = \sum_{i=1}^{M^l} \sum_{j=1}^{N^l} [(i - i_c^{k,l})^2 + (j - j_c^{k,l})^2] \mathbf{A}^{k,l}_{i,j}, \tag{2}$$

where $(i_c^{k,l}, j_c^{k,l})$ is the centroid coordinate of $\mathbf{A}^{k,l}$. Based on the CMI value, the features are divided into two sets: $\mathbf{U}^{\text{high}}_{\text{CMI}}$ and $\mathbf{U}^{\text{low}}_{\text{CMI}}$, in which the CMI values of the features are higher than $\theta^{\text{high}}_{\text{CMI}}$ and lower than $\theta^{\text{low}}_{\text{CMI}} (< \theta^{\text{high}}_{\text{CMI}})$ respectively.

Gradient-Based Feature Combination: After feature selection, we first compute the gradient of the score for positive glaucoma, y^{glau} (before the softmax), with respect to the selected feature maps $\mathbf{A}^{k,l}$, i.e., $\frac{\partial y^{\text{glau}}}{\partial \mathbf{A}^{k,l}}$. Then, these gradients flowing back are global-average-pooled (GAP) to obtain the feature importance weight $w^{k,l}$:

$$w^{k,l} = \frac{1}{N^l \times M^l} \sum_{i=1}^{M^l} \sum_{j=1}^{N^l} \frac{\partial y^{\text{glau}}}{\partial \mathbf{A}^{k,l}_{ij}}. \tag{3}$$

Weight $w^{k,l}$ captures the "importance" of the map of the k-th feature in the l-th layer for positive glaucoma. Finally, we perform a gradient-weighted combination of forward activation maps to obtain the visualization map $\mathbf{V}$ for an input glaucoma image $\mathbf{I}$:

$$\mathbf{V}(\mathbf{I}) = \sum_{\mathbf{A}^{k,l} \in \mathbf{U}} w^{k,l} \mathbf{A}^{k,l}, \tag{4}$$

where $\mathbf{U}$ is the feature set $\mathbf{U}_{\mathrm{CMI}}^{\mathrm{high}}$ or $\mathbf{U}_{\mathrm{CMI}}^{\mathrm{low}}$.

4.2 Visualization-Based Fundus Image Synthesis

In this section, we design the Patho-GAN to synthesize glaucoma fundus images enhanced in RNFL and RL. The Patho-GAN is achieved by combining a basic GAN and our visualization system. The framework of Patho-GAN is shown in Fig. 1. The overall structure of Patho-GAN consists of three subnets: generator net, discriminator net and pathology-aware visualization net. Given the vessel segmentation image (generated by the method proposed in [13]) and a noise code as input, the generator tries to synthesize images, while the discriminator net tries to tell apart the synthesized images from the real ones. The visualization net enforces the synthesized image to have the similar visualization results to the reference image.

Generator Net: We use the U-Net [14] structure for the generator network. Taking a segmentation image $\mathbf{Y} \in \{0,1\}^{W \times H}$ with a noise code $\mathbf{z} \in \mathbb{R}^Z$ as input, the network outputs a synthesized glaucoma fundus image $\mathbf{X}_{\mathrm{syn}} \in \mathbb{R}^{W \times H \times 3}$, where 3 means the three channels of RGB. Mathematically, the image synthesis process can be expressed as $G_{\boldsymbol{\theta}}(\mathbf{X}_{\mathrm{syn}}) : (\mathbf{Y}, \mathbf{z}) \mapsto \mathbf{X}_{\mathrm{syn}}$, where $\boldsymbol{\theta}$ is the learnable parameters.

Discriminator Net: We can also define a discriminant function $D_{\boldsymbol{\gamma}} : (\mathbf{X}_{\mathrm{real}}) \mapsto p \in [0,1]$. When input with a real image $\mathbf{X}_{\mathrm{real}}$, p should tend to 1; when input with a synthesized image $\mathbf{X}_{\mathrm{syn}}$, p need to tend to 0. We follow the GAN's strategy and solve the following optimization problem that characterizes the interplay between G and D:

$$\max_{\boldsymbol{\theta}} \min_{\boldsymbol{\gamma}} L(G_{\boldsymbol{\theta}}, D_{\boldsymbol{\gamma}}) = \mathbb{E}\left[L_{\mathrm{adv}}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + \lambda L_{\mathrm{patho}}(\boldsymbol{\theta})\right]. \tag{5}$$

In the above equation, $L_{\mathrm{adv}} = \log D_{\boldsymbol{\gamma}}(\mathbf{X}_{\mathrm{real}}) + \log(1 - D_{\boldsymbol{\gamma}}(G_{\boldsymbol{\theta}}(\mathbf{Y}, \mathbf{z})))$ is the adversarial loss and L_{patho} is the pathological loss, with $\lambda > 0$ being a trade-off constant.

Pathology-Aware Visualization Net: To enhance the synthesized glaucoma images in specific pathological area, we enforce the synthesized image to have the similar visualization heat map to the reference input image. In this way, the Patho-GAN can generate glaucoma fundus images with clearer pathologies (e.g., RL and RNFL). The L_{patho} is defined as:

$$L_{\mathrm{patho}} = \|\mathbf{V}(\mathbf{X}_{\mathrm{real}}) - \mathbf{V}(\mathbf{X}_{\mathrm{syn}})\|_1. \tag{6}$$

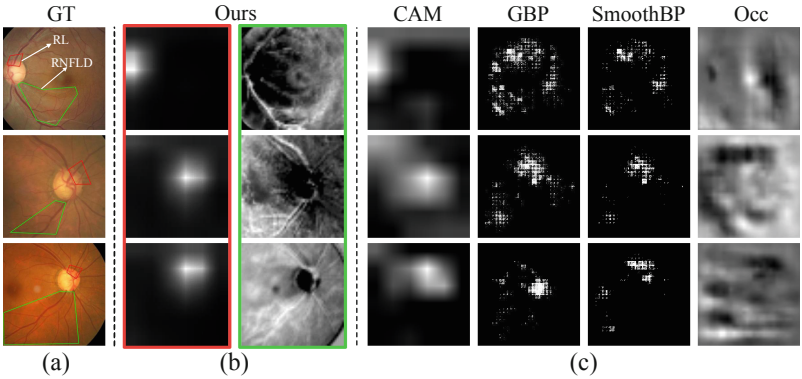

Fig. 2. (a) The ground truth of RNFL and RL. (b) The visualization results of our visualization method. The left column shows visualization results of features in $\mathbf{U}_{\mathrm{CMI}}^{\mathrm{high}}$ and the right one shows results of $\mathbf{U}_{\mathrm{CMI}}^{\mathrm{low}}$. (c) The visualization results of other methods.

5 Experiments and Results

5.1 Setting

In our experiments, the classification network is trained on the training set of LAG database for 500 epochs, with mini-batches of size 8 and the learning rate of 0.001 with Adam optimizer [15]. We have evaluated the performance of the network on the test set of LAG. Our network achieves a classification accuracy of 0.89 and AUC of 0.84. Note that the main scope of this paper is about the visualization method, not optimizing the DNN. In visualization, the selection thresholds θ_{act}, $\theta_{\mathrm{CMI}}^{\mathrm{low}}$ and $\theta_{\mathrm{CMI}}^{\mathrm{high}}$ are set to 0.1, 0.1 and 0.9 respectively.

In training the Patho-GAN, The input images are upscaled from 500 to 512 (pixels), with values scaled to the range of $[0, 1]$. The batch size is set to 1, and random rotation is performed on the input. The training is done using the SGD optimizer and the learning rate is set to 0.0002. We update the generator twice and then update discriminator once. During training, the noise code is sampled element-wise from zero-mean Gaussian with standard deviation of 0.01 and the trade-off constant λ is set to 1000. The training finishes after 240,000 mini-batches, and the inference speed of Patho-GAN is 13 ms per image on a computer with a GPU of Nvidia GeForce GTX 1080.

5.2 Pathology-Aware Visualization Results

In our experiment, we compare the visualization performance between our method and 4 other commonly used methods: CAM [5], GBP [4], SmoothBP [16] and Occ [3]. The visualization results are shown in Fig. 2. As can be seen, our pathology-aware visualization performs better in locating both the RL and RNFLD areas. Besides, CAM creates the heat maps similar to ours but more dispersedly. Occ produces the heat maps highlighting large areas but without

any clear emphasis. Although GBP and SmoothBP can show part of the RL and RNFLD, there exists strong noise that makes the pathological regions not be highlighted.

Table 1. RL and RNFLD localization accuracy (mean ± std %) for our and 4 other methods on the LAG test set.

Accuracy	Ours	CAM	GBP	SmoothBP	Occlusion
RL	$\mathbf{31.8 \pm 0.3}$	21.3 ± 0.2	11.2 ± 0.4	13.3 ± 0.8	4.2 ± 0.2
RNFLD	$\mathbf{41.2 \pm 0.5}$	3.8 ± 0.1	7.6 ± 0.8	6.3 ± 0.7	23.6 ± 0.3

Moreover, to quantitatively evaluate the visualization methods, we calculate the average pathology localization accuracy of these methods. The localization accuracy of a heat map is defined as $\text{Acc} = \frac{\sum_{(i,j)\in \text{GT}} I_{i,j}}{\sum_{(i,j)} I_{i,j}}$, where $I_{i,j}$ indicates the value of pixel (i, j) at the heat map. The localization results are shown in Table 1. Compared to other methods, our pathology-aware visualization reaches a much higher accuracy in both RL and RNFLD localization. Thus, we can conclude that our pathology-aware visualization method can pinpoint different pathological evidence for the glaucoma classification network.

5.3 Synthesized Glaucoma Images of Patho-GAN

After training the Patho-GAN with the pathological constrain from the feature set $\mathbf{U}_{\text{CMI}}^{\text{high}}$ and $\mathbf{U}_{\text{CMI}}^{\text{low}}$, we obtain the synthesized glaucoma fundus images enhanced in RL and RNFLD areas. The training results are shown in Fig. 3. As can be seen in Fig. 3(a), the synthesized images with enhancement in RL can highlight the turning point of blood vessels from the optic cup to the optic disc, and thus the rim is more clear. In Fig. 3(b), the synthesized images with enhancement in RNFLD show a clearly shading boundary than the baseline, making the RNFLD more evident.

To further verify our method, we conducted an evaluation experiment with the assistance of a professional ophthalmologist. In the experiment, the ophthalmologist was asked to evaluate the randomly shuffled fundus images synthesized by our Patho-GAN and the baseline. For each image, he rated two scores (ranged 1–5, higher indicates better): (1) clarity of RL, and (2) clarity of RNFLD. Finally, we collected valid scores on 240 synthesized images and calculated the average scores in RL and RNFLD. In RL, Patho-GAN scores **4.19** while the baseline score is 3.90; in RNFLD, Patho-GAN scores **3.14**, while the baseline obtain the score of 2.05. The p value of T-test is 3.30×10^{-3} and 6.90×10^{-7} for RL and RNFLD clarity, respectively, i.e., our mean score is higher than the baseline. Thus, we can conclude that the Patho-GAN can synthesize glaucoma images with pathophysiologically clearer RL and RNFLD.

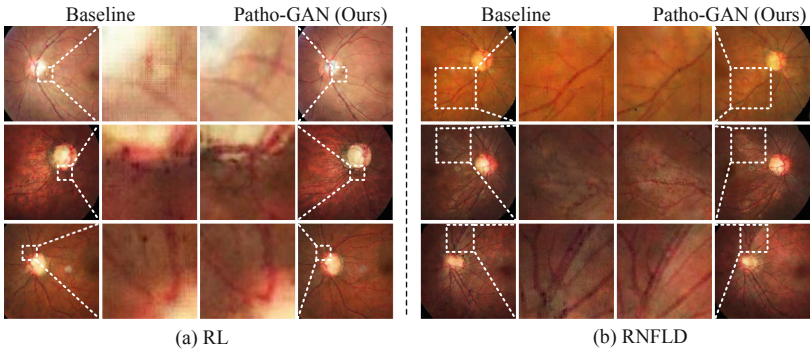

Fig. 3. Details comparison between Patho-GAN and the baseline method (without pathological constrain) in RL (a) and RNFLD (b).

6 Summary

In this paper, we have proposed a novel pathology-aware visualization method that pinpoints different pathological evidence for glaucoma classification network, by pre-selecting the feature maps and combining them in a gradient-based weight. Visualization results showed that our method can well locate the RL and RNFLD in the glaucoma fundus images with high accuracy. Moreover, we proposed the Patho-GAN, which applied our visualization method to the glaucoma image synthesis task. The experimental results showed that our Patho-GAN can synthesize glaucoma images with pathophysiologically clearer RL and RNFLD.

Acknowledgments. This work was supported by the NSFC projects 61876013, 61922009, 61573037, and by BMSTC under Grant Z181100001918035.

References

1. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115 (2017)
2. Chen, X., Xu, Y., Wong, D.W.K., Wong, T.Y., Liu, J.: Glaucoma detection based on deep convolutional neural network. In: EMBC 2015, pp. 715–718. IEEE (2015)
3. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
4. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: the all convolutional net. arXiv preprint arXiv:1412.6806 (2014)
5. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR 2016, pp. 2921–2929 (2016)
6. Van Molle, P., De Strooper, M., Verbelen, T., Vankeirsbilck, B., Simoens, P., Dhoedt, B.: Visualizing convolutional neural networks to improve decision support for skin lesion classification. In: Stoyanov, D., et al. (eds.) MLCN/DLF/IMIMIC

-2018. LNCS, vol. 11038, pp. 115–123. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02628-8_13
7. Rieke, J., Eitel, F., Weygandt, M., Haynes, J.-D., Ritter, K.: Visualizing convolutional networks for MRI-based diagnosis of alzheimer's disease. In: Stoyanov, D., et al. (eds.) MLCN/DLF/IMIMIC -2018. LNCS, vol. 11038, pp. 24–31. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02628-8_3
8. Wang, X., et al.: Weakly supervised learning for whole slide lung cancer image classification. In: MIDL (2018)
9. Gondal, W.M., Köhler, J.M., Grzeszick, R., Fink, G.A., Hirsch, M.: Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images. In: ICIP, pp. 2069–2073. IEEE (2017)
10. Hu, Y., et al.: Label-driven weakly-supervised learning for multimodal deformarle image registration. In: ISBI (2018)
11. Kervadec, H., Dolz, J., Tang, M., Granger, E., Boykov, Y., Ayed, I.B.: Constrained-cnn losses for weakly supervised segmentation. MIA **54**, 88–99 (2019)
12. Li, L., Xu, M., Wang, X., Jiang, L., Liu, H.: Attention based glaucoma detection: a large-scale database with a CNN model. arXiv preprint arXiv:1903.10831 (2019)
13. Xiancheng, W., et al.: Retina blood vessel segmentation using a U-net based convolutional neural network. In: ICDS 2018, Beijing, China, 8–9 June 2018 (2018)
14. 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
15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M.: SmoothGrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825 (2017)