Denoising Enhances Visualization of Optical Coherence Tomography Images

Harishwar Reddy K[1], **Anshul Shivhare**[1], **Hemanth Kongara**[1], **Jayesh Saita**[2], **Raghu Prasad**[2], **and Chandra Sekhar Seelamantula**[1]
[1]Department of Electrical Engineering, Indian Institute of Science, Bangalore
[2]Carl Zeiss Pvt. Limited
{harishwarreddy98, shivhareanshul78, hemanthkongara35}@gmail.com,
{jayesh.saita, raghu.prasad}@zeiss.com, css@iisc.ac.in

Abstract

The main aim of this work is to improve the visualization of abnormalities in Optical Coherence Tomography (OCT) images of the human retina. OCT images have substantial noise, which can affect the classification and visualization performance of a neural network. In this work, we show that denoising improves visualization without affecting the classification performance considering the specific pathology of Pigmented Epithelial Detachment (PED). The noise in OCT images may lead to unstable training and poor classification/visualization performance. Hence, the need for image quality enhancement. We consider several image denoising techniques, namely, K-Singular Value Decomposition (K-SVD), Bilateral filter optimized using Poisson Unbiased Risk Estimate (PURE), Poisson Unbiased Risk Estimate - Linear Expansion of Thresholds (PURE-LET), Guided filter, Rolling Guidance filter, Gated Convolutional and Deconvolutional Structure (GCDS), Denoising Convolution Neural Network (DnCNN), and DnCNN with skip connections. We compare the classification and visualization results obtained from the model trained on noisy images with that trained on enhanced images. Several class activation mapping (CAM) based techniques have been developed, for instance, Grad-CAM, Grad-CAM++, Score-CAM, Ablation-CAM, and Self-Matching CAM, which are also the visualization techniques that we employ in this paper. Our results show that denoising improves visualization performance by a factor of $\approx$10 in Jaccard Index, taking the expert segmentation as the ground-truth.

1 Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that is used to image tissues such as the retina of the eye. An OCT image is produced by capturing the reflected light from the retina and the optic nerve. It can be used to diagnose and manage eye diseases such as glaucoma, diabetic retinopathy, Pigmented Epithelial Detachment (PED), Cystoid Macular Edema (CME), Age-related Macular Degeneration (AMD), Retinal Vein Occlusion (RVO) [1]. Over the past few years, there has been an explosion in the medical data that is being generated. Hence, there is a need for automatic screening techniques, and Deep Neural Networks (DNNs) have emerged as a powerful tool for screening. Yet, there are many unanswered questions around them. One of the critical problems is the explainability of DNNs. A deep learning model cannot be deployed for real-life applications based on classification metrics alone. It is important to ascertain that the model's visual explanation is clinically relevant, consistent and correlated with that of the pathologist's analysis and decision-making. In this paper, the focus is on OCT image analysis, specifically for PED, and the objective is to improve the visual explanation of the network without affecting the classification performance of a DNN. Several deep learning based models [2–6] have been developed, which have performed reasonably well for classifying various pathologies in the OCT images. Recent

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

works [7, 8] have used various explainable techniques for verifying the localization performance of deep learning-based classification models. For instance, Choi *et al.* [9] have used Grad-CAM++ to highlight the myopic regions in the image. Most of the neural networks in the aforementioned works are trained on noisy images, which causes the model to fit to noise as well, which may in turn affect the visualization performance. In this work, we show how the visualization performance of a DNN can be improved by mild denoising.

2 Dataset

The dataset consists of OCT B-scan images stacked in the form of cubes obtained using two types of machines made by Carl Zeiss, namely, CIRRUS and PRIMUS. There are in turn two variants of CIRRUS machines, one that captures 200 images per OCT cube and the other 128 images per OCT cube. There are two types of PRIMUS machines, of which one captures 128 images per cube, and the other captures 32 images per cube. The training and test data consist of 2,07,535 and 52,650 images, respectively. Each image in the B-scan from CIRRUS and PRIMUS machines is of size 1024×512 and 1024×200, respectively. The ground truth segmentation data is available for 73 OCT cubes each in turn containing 128 B-scans making up a total of 9344 (73×128) images. Of these 3686 images contain PED pathology. The segmentation ground truth is given by an expert and is used for the computation of Jaccard Index/Intersection over Union (IoU) and Dice score.

3 Methodology

To begin with, we enhance the noisy OCT images using a denoiser. We then train binary classification models using both noisy and enhanced images. Finally, using the ground-truth segmentation data for the PED pathology, we calculate IoU and Dice scores for each classification model.

3.1 Image Denoising

We consider the standard denoising techniques, namely, K-Singular Value Decomposition (K-SVD) [10], Guided filter [11], Rolling Guidance filter [12], Poisson Unbiased Risk Estimate – Linear Expansion of Thresholds (PURE-LET) [13], Bilateral filter optimized using Poisson Unbiased Risk Estimate (PURE) [14] and deep learning based techniques such as Denoising Convolution Neural Network (DnCNN) [15], DnCNN with skip connections (DnCNNS)[16], and Gated Convolutional and Deonvolutional Structure (GCDS) [17]. Our observation has been that the images denoised using the deep learning techniques have fewer artifacts than the others and hence, for subsequent classification and visualization, we employ only the deep learning based techniques.

3.2 Classification

We trained binary classification models using noisy images and the enhanced images obtained from DnCNN, DnCNNS, and GCDS. We used an Inception-ResNet-v2 [18] model pre-trained on the ImageNet dataset [19]. We fine-tuned the pre-trained model using OCT classification training data. We used a learning rate of 10^{-4}, the Adam optimizer [20], and Categorical Cross Entropy (CCE) [21] loss function. We train the classifier over 2,07,535 images with a batch size of 15 for one epoch. In order to mitigate class imbalance, we oversampled the images having PED pathology. Hereafter, we refer to the classification models trained on noisy images, DnCNN, DnCNNS, GCDS enhanced images as the Noisy model, DnCNN model, DnCNNS model, GCDS model, respectively. Table 1 shows the classification results obtained for the four classification models.

3.3 Visualization

For visualization of the PED pathology, we employ standard visualization techniques such as Grad-CAM [22], Grad-CAM++ [23], Score-CAM [24], Ablation-CAM [25], and Self-Matching CAM [26] (SM-CAM). Table 2 shows the IoU and Dice scores obtained for the aforementioned visualization techniques. We calculated the heatmaps using these visualization techniques and obtained the corresponding binary maps using Otsu thresholding. Subsequently, we calculate the IoU and Dice scores by making use of the binary maps and the corresponding masks from the expert annotated segmentation ground-truth.

4 Experimental Results

The classifiers are tested on 52,650 images that are not seen during training. Table 1 shows the classification performance of four models assessed using standard performance measures. We observe

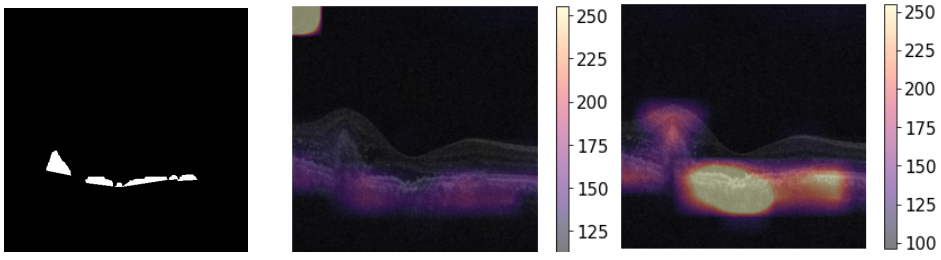
Table 1: Comparison of classification performance for four models on 52,650 test images

Performance measure	Noisy	DnCNN	DnCNNS	GCDS
True Positive	2714	2924	2722	2699
False Positive	2511	2438	1536	2795
True Negative	46434	46507	47409	46150
False Negative	991	781	983	1006
Accuracy	0.9334	0.9389	0.9521	0.9278
Precision	0.5194	0.5453	0.6392	0.4913
Sensitivity	0.7325	0.7892	0.7347	0.7285
Specificity	0.9486	0.9502	0.9686	0.9429
F1score	0.6078	0.6450	0.6837	0.5868

Table 2: A comparison of visualization performance of the four classification models.

	Grad-CAM		Grad-CAM++		Score-CAM		Ablation-CAM		SM-CAM	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Noisy	**0.1100**	**0.1924**	0.0174	0.0311	0.0175	0.0313	0.0176	0.0315	0.0787	0.1390
DnCNN	0.1024	0.1800	0.1014	0.1783	0.1118	0.1944	0.1068	0.1878	0.0920	0.1633
DnCNNS	0.1077	0.1875	0.1042	0.1824	0.1079	0.1878	0.1098	0.1921	**0.1519**	**0.2543**
GCDS	0.1033	0.1819	**0.1187**	**0.2055**	**0.1170**	**0.2033**	**0.1186**	**0.2054**	0.1239	0.2119

that DnCNN and DnCNNS models perform better than the noisy model, whereas GCDS is marginally poorer than the noisy model. Table 2 shows the visualization results obtained for four classification models. We observe that the visualization results have improved for the model trained on denoised images compared to the noisy model with the exception of Grad-CAM. This may be because of the averaging of the gradients that takes place in Grad-CAM, which causes smearing of the heatmap. The percentage improvement in IoU, Dice scores for Grad-CAM++, Score-CAM, Ablation-CAM, and Self-matching CAM is more than the percentage decrement in the case of Grad-CAM. Figure 1 shows the heatmaps obtained using the noisy model and the GCDS model in comparison with the ground truth. We observe that the Grad-CAM++ output obtained using GCDS model is more in agreement with the ground truth than the noisy model. The spurious results in Fig. 1b show that the classification model has fitted noise as well. On the other hand, Fig. 1c, shows that the GCDS model focuses on the relevant regions, and is in better agreement with the ground truth. Thus, the GCDS model improved the visualization. This phenomenon has been observed with other CAM techniques and for other models as well. Additional results are provided in the supporting document.



(a) (b) (c)

Figure 1: GradCAM++ visualization vs. ground-truth: (a) Ground-truth segmentation; (b) Heatmap obtained using the noisy model; (c) Heatmap obtained using the GCDS model.

5 Conclusion

The key objective of this work is to demonstrate that mild image denoising improves visualization of a deep learning classification model. Considering OCT images and PED as an example, we showed, by means of several visualization tools, that there is a significant improvement in quantitative

visualization performance measures compared to the noisy model. Qualitative results also confirm these improvements. However, we would also like to add that there is a significant scope for further improvement in the performance metrics.

6 Acknowledgements and Disclosure of Funding

We would like to thank Dr. Ganesh for his help with data curation and analysis of denoised images. The authors are thankful to Dr. Rajiv Raman Shri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya, Chennai, Tamil Nadu, India for providing the datasets for this study. This research was funded by Carl Zeiss Pvt Limited.

7 Potential negative societal impact

We do not foresee any negative societal impact. An improvement in visualization performance is only likely to increase confidence in adopting the deep learning classification model.

References

[1] James G Fujimoto, Costas Pitris, Stephen A Boppart, and Mark E Brezinski, "Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy," in *Neoplasia*, 2000, pp. 9–25.

[2] G Panozzo, B Parolini, E Gusson, A Mercanti, S Pinackatt, G Bertoldo, and S Pignatto, "Diabetic macular edema: an oct-based classification," in *Seminars in ophthalmology*. Taylor & Francis, 2004, vol. 19, pp. 13–20.

[3] Nabila Eladawi, Mohammed Elmogy, Mohammed Ghazal, Omar Helmy, Ahmed Aboelfetouh, Alaa Riad, Shlomit Schaal, and Ayman El-Baz, "Classification of retinal diseases based on oct images," *Frontiers in Bioscience-Landmark*, vol. 23, no. 2, pp. 247–264, 2018.

[4] Vineeta Das, Samarendra Dandapat, and Prabin Kumar Bora, "Multi-scale deep feature fusion for automated classification of macular pathologies from oct images," *Biomedical signal processing and Control*, vol. 54, pp. 101605, 2019.

[5] Jessica Loo, Cindy X Cai, John Choong, Emily Y Chew, Martin Friedlander, Glenn J Jaffe, and Sina Farsiu, "Deep learning-based classification and segmentation of retinal cavitations on optical coherence tomography images of macular telangiectasia type 2," *British Journal of Ophthalmology*, vol. 106, no. 3, pp. 396–402, 2022.

[6] Yuemei Luo, Qing Xu, Ruibing Jin, Min Wu, and Linbo Liu, "Automatic detection of retinopathy with optical coherence tomography images via a semi-supervised deep learning method," *Biomedical Optics Express*, vol. 12, no. 5, pp. 2684–2702, 2021.

[7] Xiao Ma, Zexuan Ji, Sijie Niu, Theodore Leng, Daniel L. Rubin, and Qiang Chen, "Ms-cam: Multi-scale class activation maps for weakly-supervised segmentation of geographic atrophy lesions in sd-oct images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3443–3455, 2020.

[8] Tasnim Sakib Apon, Mohammad Mahmudul Hasan, Abrar Islam, and Md. Golam Rabiul Alam, "Demystifying deep learning models for retinal oct disease classification using explainable ai," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2021, pp. 1–6.

[9] Kyung Jun Choi, Jung Eun Choi, Hyeon Cheol Roh, Jun Soo Eun, Jong Min Kim, Yong Kyun Shin, Min Chae Kang, Joon Kyo Chung, Chaeyeon Lee, Dongyoung Lee, et al., "Deep learning models for screening of high myopia using optical coherence tomography," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[10] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[11] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European conference on computer vision*. Springer, 2010, pp. 1–14.

[12] Qi Zhang, Xiaoyong Shen, Li Xu, and Jiaya Jia, "Rolling guidance filter," in *European conference on computer vision*. Springer, 2014, pp. 815–830.

[13] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of poisson-corrupted images," in *Signal Process. 90*, 2010, p. 415–427.

[14] Kishan Harini and Seelamantula Chandra, Sekhar, "Optimal parameter selection for bilateral filters using poisson unbiased risk estimate," *IEEE transactions on image processing*, 2012.

[15] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[16] "Dilated residual networks with symmetric skip connection for image denoising," *Neurocomputing*, vol. 345, pp. 67–76, 2019, Deep Learning for Intelligent Sensing, Decision-Making and Control.

[17] B. Anoop, KS. Kalmady, A. Udathu, V. Siddharth, G Girish, AR. Kothari, and J. Rajan, "A cascaded convolutional neural network architecture for despeckling oct images," *Biomedical Signal Processing and Control 66:102463*, 2021.

[18] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[19] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[20] Diederik P. Kingma and Jimmy. Ba, "Adam: A method for stochastic optimization," in *CoRR*, 2014, pp. ,abs/1412.6980.

[21] Zhilu Zhang and Mert Sabuncu, "Generalized cross-entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.

[22] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[23] A. Chattopadhyay, A. Sarkar, P. Howlader, and V.N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, p. 839–847.

[24] H. Wang, Du M. Wang, Z., F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-cam: Score-weighted visual explanations for convolutional neural networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 24–25.

[25] Saurabh Desai and Harish G Ramaswamy., "Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, p. 972–980.

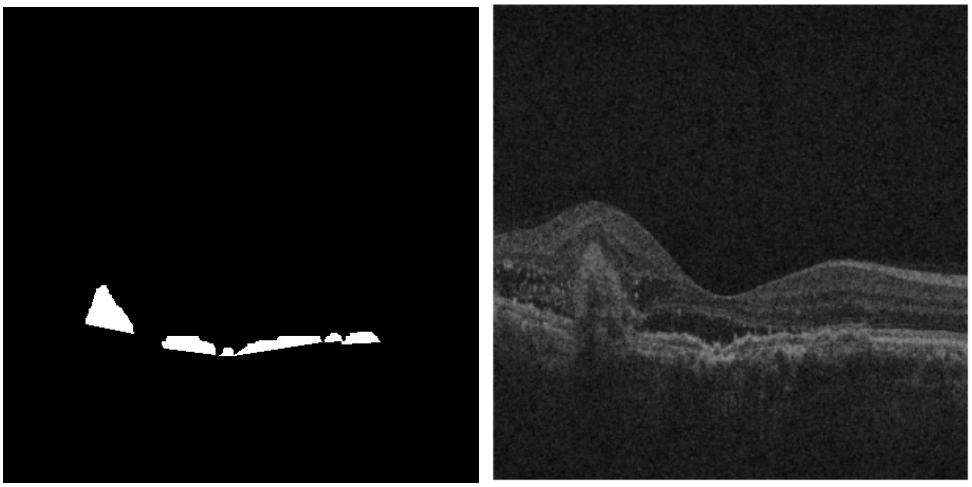
[26] Zhenpeng Feng, Mingzhe Zhu, Ljubiša Stankovic, and Hongbing Ji, "Self-matching CAM: A novel accurate visual explanation of CNNs for SAR image interpretation," *Remote Sens. 13*, p. 1772.

Denoising Enhances Visualization of Optical Coherence Tomography Images (Supplementary Material)

Harishwar Reddy K[1], **Anshul Shivhare**[1], **Hemanth Kongara**[1], **Jayesh Saita**[2], **Raghu Prasad**[2], **and Chandra Sekhar Seelamantula**[1]
[1]Department of Electrical Engineering, Indian Institute of Science, Bangalore
[2]Carl Zeiss Pvt. Limited
{harishwarreddy98, shivhareanshul78, hemanthkongara35}@gmail.com,
{jayesh.saita, raghu.prasad}@zeiss.com, css@iisc.ac.in

1 Qualitative results

In this section, we show an OCT image having PED pathology, and apply Grad-CAM, Grad-CAM++, Score-CAM, Ablation-CAM, Self-Matching CAM using the four classification models. Figure 1 shows the noisy OCT image and its corresponding ground truth. Figures 2 to 5 show the visualizations results obtained from the Noisy model, DnCNN model, DnCNNS model, and GCDS model, respectively.



(a) (b)

Figure 1: OCT image and its ground-truth: (a) Ground-truth segmentation for PED pathology; (b) Noisy OCT image.

We observe from Figure 2, that Grad-CAM++, Score-CAM, Ablation-CAM, Self-Matching CAM are not highlighting the clinically relevant regions in the image. On the other hand, in Figures 3 to 5, we see that the highlighted portion in the heatmaps are more in agreement with the ground-truth. Therefore, we conclude that denoising improves visualization for the OCT images.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

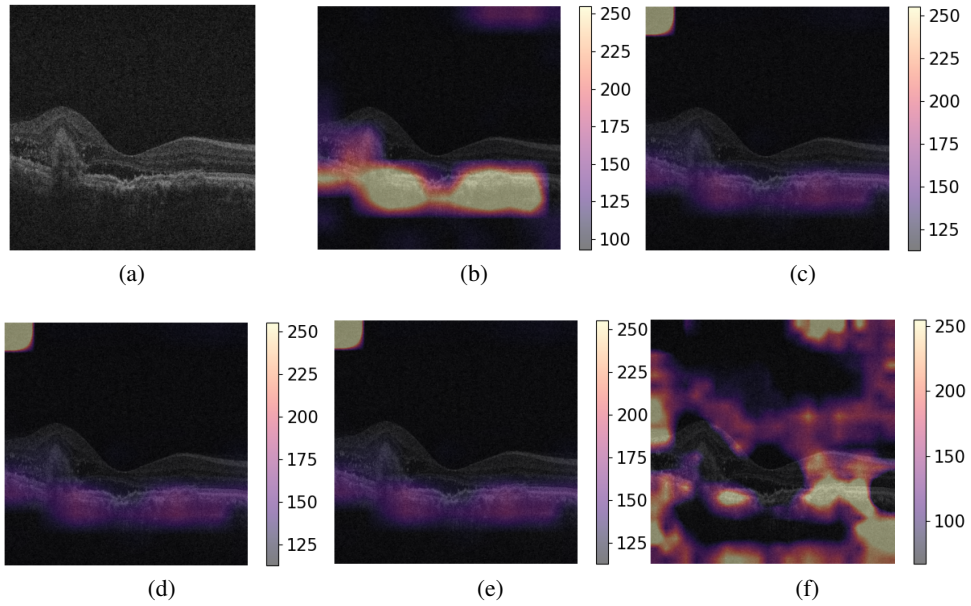


Figure 2: Noisy model visualization results: (a) Noisy OCT image; Heatmap obtained using (b) Grad-CAM; (c) Grad-CAM++; (d) Score-CAM; (e) Ablation-CAM; (f) Self-Matching CAM.

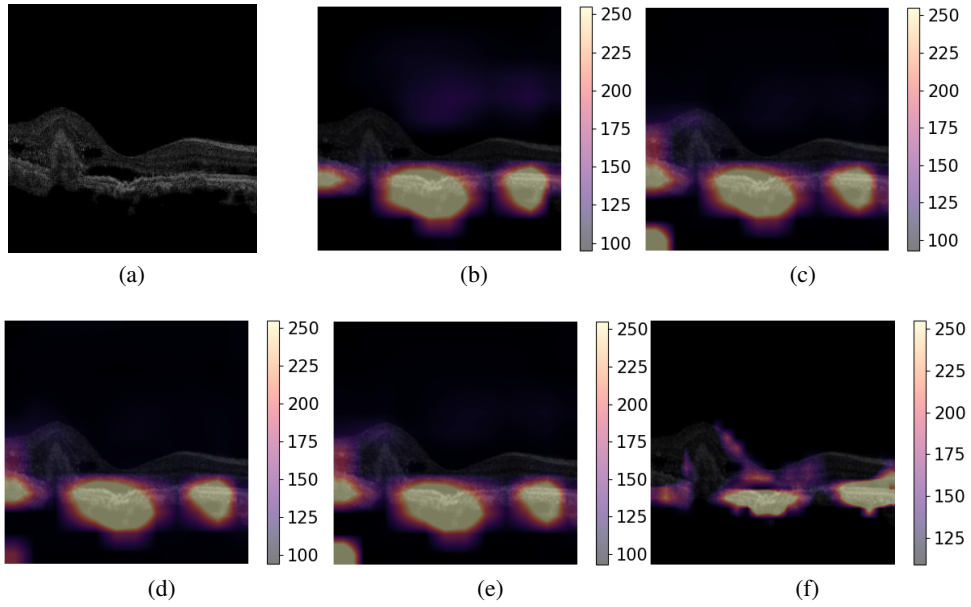


Figure 3: DnCNN model visualization results: (a) Image denoised using DnCNN model; Heatmap obtained using (b) Grad-CAM; (c) Grad-CAM++; (d) Score-CAM; (e) Ablation-CAM; (f) Self-Matching CAM.

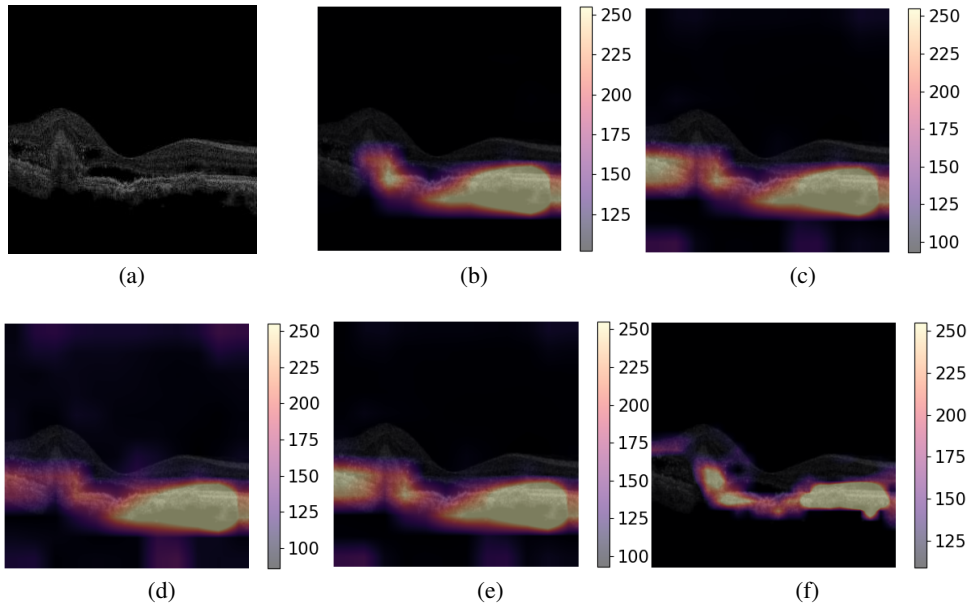


Figure 4: DnCNNS model visualization results: (a) Image denoised using DnCNNS model; Heatmap obtained using (b) Grad-CAM; (c) Grad-CAM++; (d) Score-CAM; (e) Ablation-CAM; (f) Self-Matching CAM.

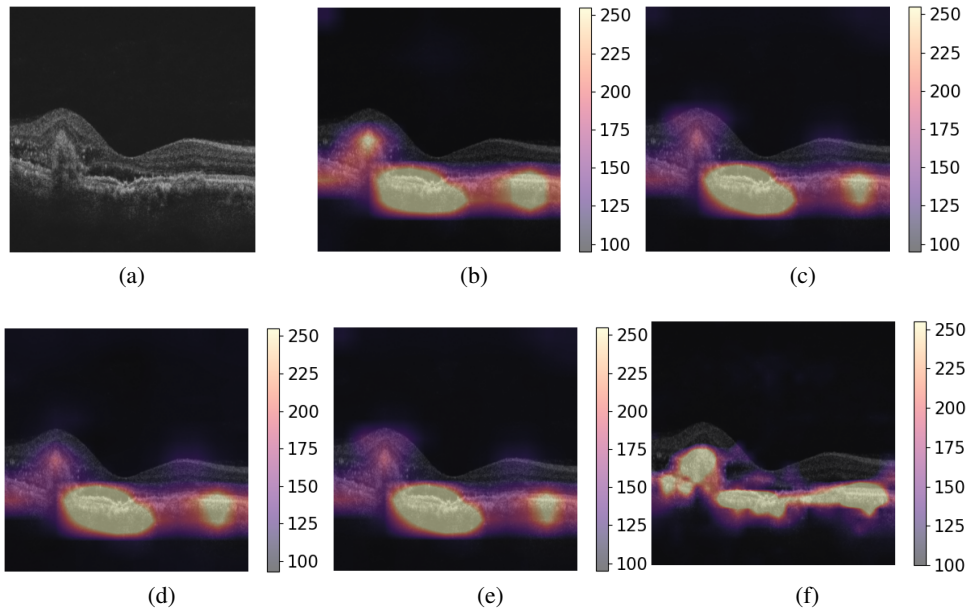


Figure 5: GCDS model visualization results: (a) Image denoised using GCDS model; Heatmap obtained using (b) Grad-CAM; (c) Grad-CAM++; (d) Score-CAM; (e) Ablation-CAM; (f) Self-Matching CAM.