2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-8622-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/BIBM62325.2024.10821928

Leveraging Spectrum to Enhance the Accuracy of Retinal Layer Segmentation

Xiang He[1,2], Weiye Song[1], Fuwang Wu[1], Lizhen Cui[2,3] and Yi Wan[1,*]

***Abstract*—In the early development of many fundus diseases, changes in retinal layer thickness occur, and optical coherence tomography (OCT) serves as the gold standard for detecting variations in retinal layer thickness. Consequently, extensive research has been conducted on retinal layer segmentation (RLS) based on OCT images. However, current breakthroughs in RLS primarily focus on innovations in deep learning (DL) models and improvements in training strategies, which constrains the means to enhance the accuracy of RLS. We have discovered a new dimension of spectral information that can be utilized to improve the precision of RLS. Experimental results demonstrate that the accuracy of RLS, measured by the mIoU metric, can be maximized with an improvement of up to 1.57 percentage points, and that different combinations of spectral images vary in their effectiveness at enhancing segmentation outcomes. Our study will introduce novel research avenues in the domain of RLS based on DL.**

***Index Terms*—Retinal layer, Segmentation, Spectral information, TransUnet, Gaussian window**

I. Introduction

The ubiquity of modern electronic devices and detrimental visual habits have precipitated a surge in eye-related ailments, compelling researchers to advance the diagnostic timeline. Notably, the nascent phases of numerous retinal diseases are marked by alterations in retinal layer thicknesses: glaucoma, for instance, often manifests early with a reduction in NFL thickness in tandem with visual field impairment [1], [2]. Furthermore, the progression of certain systemic and neurological conditions can also impact retinal layer thicknesses [3]. As OCT stands as the paramount tool for discerning variations in retinal layer thickness, extensive investigations have been undertaken concerning the segmentation of retinal layers based on OCT imaging [4].

DL has gained significant popularity in recent years, with numerous studies focusing on its application to OCT images. These include fundus disease detection [5], OCT image enhancement [6], and 3D reconstruction of OCT images [7]. Naturally, this also encompasses research on RLS based on DL [8]. Currently, the work in this field is primarily concentrated on three areas: firstly, enhancing DL models [9], [10]; secondly, updating DL training strategies [8], [11]; and thirdly, concentrating on specific diseases [3]. Overall, there is a continuous pursuit of higher accuracy in RLS while incorporating increasingly complex conditions. However, the primary breakthrough point at present is largely within DL methodologies.

Recent research has demonstrated that near-infrared light, when used to penetrate the retinal pigment epithelium, is more effective than visible light [12]. Consequently, there are significant differences in certain layers of the retina observed in OCT images from different central wavelength spectrum. Moreover, according to the principle of multi-spectral fundus cameras, we understand that different wavelengths of light, when shining on the same fundus tissue, produce different imaging effects due to varying backscattering rates [13]. Therefore, we hypothesize that imaging the same retina in the fundus using light with different center wavelengths will not only yield structural information but also additional spectral information. In this context, the objective of this paper is to investigate whether spectral information can enhance the accuracy of existing RLS based on structural information.

The primary contributions of this article are as follows:

1) For the first time in RLS research, we have introduced a novel dimension of information: spectral data. This advancement significantly improves upon the conventional reliance on structural information from OCT images alone, opening new avenues for investigation in this field.
2) We have examined the differences between OCT images with varying center wavelength spectrum and confirmed the existence of spectral information.
3) By inputting images containing spectral information into general retinal layer methods and comparing them to images without spectral information, we have demonstrated that spectral information can enhance the accuracy of RLS.

II. Method

A. Obtaining Spectral Images by Gaussian Windows

The primary component of the OCT imaging system is the Michelson interferometer optical path. After the light emitted by the light source is split, one path of light enters the sample arm and, upon focusing, converges onto the surface of the sample. The other beam enters the reference arm and focuses onto the surface of a planar mirror. The return beams from both arms interfere and enter the detection arm. The returned light, which carries information about the sample, is detected by the detection device. Following subsequent signal processing and analysis, the structural information and optical properties of

This work is supported in part by the Key Technology Research and Development Program of Shandong Province under Grant 2024CXGC010201, 2024JMRH0301 and 2024JMRH0207.

[1]School of Mechanical Engineering, Shandong University, Jinan, China.
[2]Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan, China.
[3]School of Software, Shandong University, Jinan, China.
*: Corresponding author.

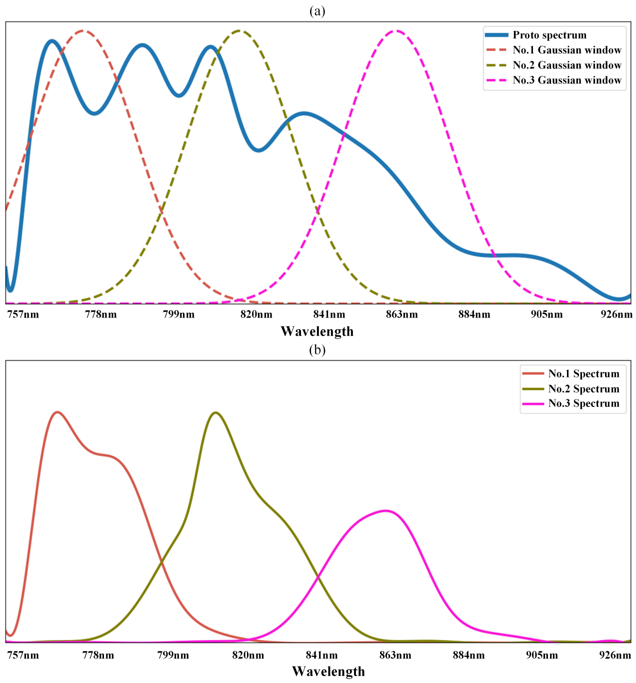

Fig. 1. (a) Add Gaussian windows at different positions in the raw spectrum, (b) three new spectrum obtained by convolving Gaussian windows at different positions with the raw spectrum.

the sample can be reconstructed [4]. Among these, the difference in the central wavelength of the light source spectrum can cause the same retinal tissue to exhibit different optical properties. Based on this, we generate spectrum $S'(x,\mu)$ with different central wavelengths by adding Gaussian windows at different positions on the raw spectrum $S(x)$, as shown in the following formula:

$$G(x,\mu) = e^{\frac{-(x-\mu)^2}{2\times\sigma^2}}, \tag{1}$$

where $G(x,\mu)$ represents the Gaussian function, μ represents a coefficient related to the position of the Gaussian window, σ represents a coefficient related to the shape of Gaussian windows. Then use $G(x,\mu)$ to convolve the raw spectrum $S(x)$ to obtain the sub-band spectrum $S'(x,\mu)$, as shown in Eq. 2.

$$S'(x,\mu) = G(x,\mu) \otimes S(x), \tag{2}$$

In Fig. 1(a), Gaussian windows are incorporated at 778nm, 821nm, and 865nm into the raw spectrum. These are then convolved with the raw spectrum to generate three distinct sub-band spectrum, each with a different central wavelength, as illustrated in Fig. 1(b). Employing these three modified spectrum for concurrent retinal imaging allows the acquisition of OCT images of the retinal layers. These images maintain consistent structural information while incorporating spectral data from varying central wavelengths, as shown in Fig. 2(a), (b) and (c).

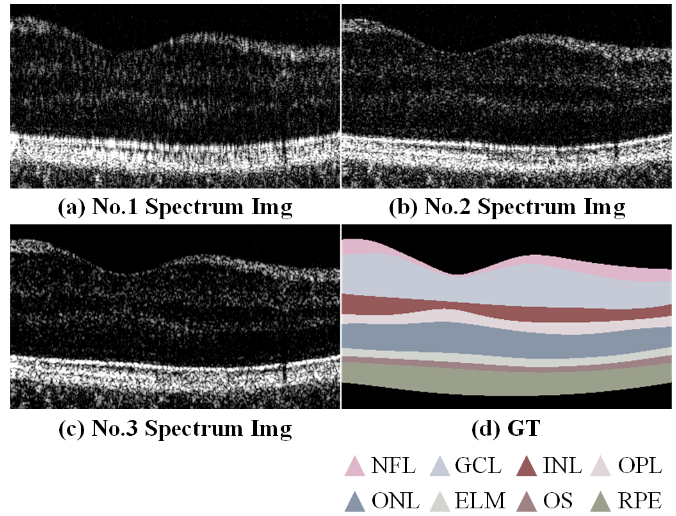

Fig. 2. One sample of the dataset, including three OCT images from different spectrums and one GT annotated image. Additionally, the retinal layers represented by the colors of each category in GT are provided.

B. DL-based Segmentation Methods

In this study, we conduct experiments using three DL based segmentation methods: LightRsSeg [9], Segformer_B0 [14], and TransUnet [15]. LightRsSeg is a relatively new lightweight RLS method, which is improved based on a U-shaped encoder decoder network and combines modules such as TransFormer and MAA. Segformer considers efficiency, accuracy, and robustness simultaneously. It designs a novel position encoding free layered transformer encoder and redesigns the encoder and decoder sections, achieving sota performance on multiple datasets. TransUnet employs a segmentation network grounded on Transformer, merging CNN with Transformer to overcome the constraints of conventional convolutional neural networks in modeling extensive-range dependencies and managing large image dimensions.

III. Experimental Results and Discussion

A. Dataset

The dataset comprises scan data from 100 eyes of 50 healthy volunteers, with each eye undergoing a macular scan and 512 B-scan images captured per eye. From each eye's data, 5 B-scan images are selected, each containing 3 OCT images from different spectrum. Consequently, the dataset totals 500 samples, each comprising 3 OCT images from varying spectrum and one annotated GT image, summing up to 2000 images. The annotation image delineates 9 categories of structures, as shown in Fig 2. This dataset is partitioned into training set and testing set, encompassing 400 cases (1600 images) and 100 cases (400 images), respectively. Each image measures 480 × 400 pixels.

B. Implementation Details

The experiment is implemented based on the PyTorch and trained with the Adam optimizer with the cross-entropy loss.

The initial learning rate is set to 0.001 which is then gradually halved every 40 epochs. Data augmentation is applied, including horizontal flipping with probability P=0.5, random center rotation with probability P=0.5. All experiments are conducted on 4 NVIDIA A100 graphics cards. We evaluate the segmentation performance quantitatively using the Dice similarity coefficient (DSC), mean Intersection over Union (mIoU), and mean pixel accuracy (mPA).

C. Results and Analysis

*1) **Differences of retina between different spectrum:*** Based on our analysis, utilizing light with diverse central wavelengths to image the same retinal region in the fundus yields not only structural information but also spectral information. Consequently, we employ a direct method to contrast the variances across different spectral images, as illustrated in Fig. 3. We aggregate and compute the mean pixel values within each retinal stratum in OCT images No.1, No.2, and No.3, resulting in a singular average pixel value for each retinal layer. By subtracting these mean pixel values between two distinct spectrum, we derive the corresponding differences for each layer, depicted in Fig. 3 (a), (b), and (c). Furthermore, by dividing the mean pixel values of each retinal layer between two separate spectrum, we obtain the respective ratios for each layer, shown in Fig. 3 (d), (e), and (f).

Upon examining Fig. 3, it becomes evident that both direct subtraction and division reveal substantial disparities in the mean pixel values across retinal layers between any two spectral bands. This is particularly apparent in Fig. 3(a), where nearly adjacent layers exhibit significant differences. If such differences did not exist between different spectral images, subtraction would result in all zeros, and division would yield all ones. Consequently, our analysis uncover notable variations in the mean pixel values of each retinal layer across distinct spectral images. Tab. I provides the exact values computed for this image, revealing not only are there differences between retinal layers, but also distinct variations among different spectral combinations. The differences observed between combinations No.1&2 and No.1&3, as well as the inconsistencies found between No.1&2 and No.2&3, underscore the significant differences in the spectral images of diverse retinal layers. We employ the mean pixel value of each layer as a straightforward method to compare differences among various retinal layers in OCT images with differing center wavelengths. While this approach does not encompass all spectral information, it undoubtedly serves as evidence for the presence of spectral information. This implies that OCT images, when captured using different center wavelength spectrum, yield not only structural details but also additional spectral information alongside the structural information.

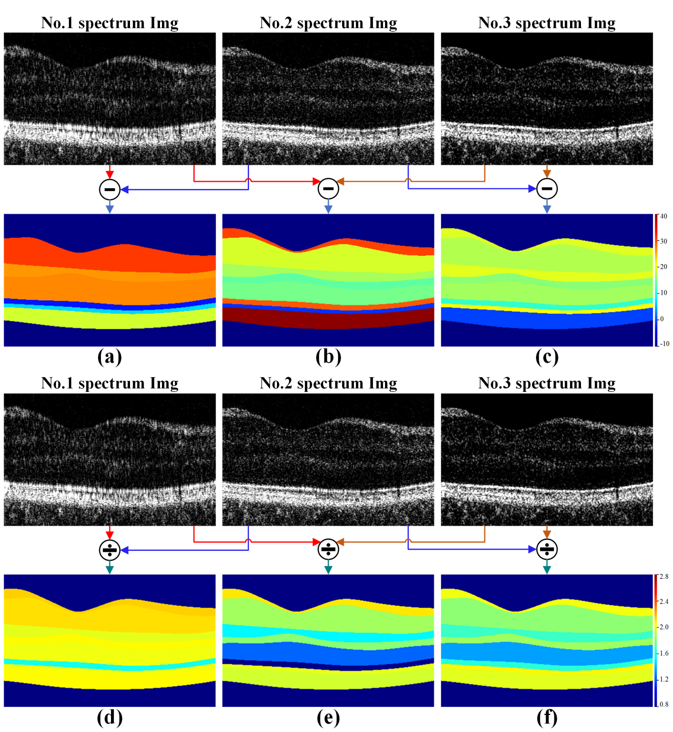

Fig. 3. Qualitative analysis of differences in retinal layers between different spectral images.

TABLE I
QUANTITATIVE ANALYSIS OF DIFFERENCES IN RETINAL LAYERS BETWEEN DIFFERENT SPECTRAL IMAGES.

Layers	Dividing the corresponding retinal layer pixel values			Subtraction of corresponding retinal layer pixel values		
	No.1&2	**No.1&3**	**No.2&3**	**No.1&2**	**No.1&3**	**No.2&3**
NFL	0.92	0.99	1.08	-6.63	-0.86	5.77
GCL	0.96	1.30	1.35	-1.04	6.03	7.07
INL	1.11	1.75	1.57	1.79	7.43	5.64
OPL	1.04	1.36	1.31	1.35	9.61	8.26
ONL	1.08	2.15	1.98	1.27	8.90	7.63
ELM	1.68	2.74	1.63	19.34	30.25	10.91
OS	1.08	1.11	1.03	13.53	18.55	5.02
RPE	1.04	1.16	1.12	6.18	24.23	18.06

*2) **RLS with Spectral Information:*** Based on the conclusion that there is spectral information between retinal layer images with different center wavelength spectrums, we further conduct RLS research with additional spectral information. We conduct three experiments using LightReSeg, Segformer, and TransUnet for segmentation. The first group uses a single spectral image as input, without introducing spectral information. Each method has three settings: No.1, No.2, and No.3; The second group involves inputting two spectral images with different center wavelengths simultaneously. In addition to structural information, the input images also contain spectral information between the two pairwise images. Each segmentation method has three combinations: No.1&2, No.1&3, and No.2&3; The third group adopts the combination of No.1&2&3 to simultaneously input three images into the network. The specific experimental settings are shown in the combination column in Tab. II.

From Tab. II, we can see that the accuracy of RLS varies among different spectral images. In the LightReSeg method, we find that the overall segmentation accuracy of No.2 is higher, while No.3 is lower. In terms of mIoU index, No.2

TABLE II
THREE SEGMENTATION METHODS, SEGFORMER, TRANSUNET, AND LIGHTRESEG, ARE EMPLOYED TO DERIVE SEGMENTATION OUTCOMES ACROSS VARIOUS SPECTRAL IMAGE COMBINATIONS.

Method	**Combinations**	**Dice Score**									**mIoU**	**mPA**
		NFL	**GCL**	**INL**	**OPL**	**ONL**	**ELM**	**OS**	**RPE**	**Ave**		
LightReSeg	No.1	90.24	95.33	90.37	83.17	94.97	90.74	86.26	94.49	90.84	83.23	97.26
	No.2	90.98	95.37	90.42	82.35	94.85	91.37	86.91	94.42	90.99	83.47	97.22
	No.3	90.41	95.60	90.38	83.19	95.46	91.50	84.09	94.36	90.81	83.16	97.57
	No.1&2	90.81	95.44	90.48	83.43	95.15	91.22	86.63	94.71	91.13	83.70	97.39
	No.1&3	90.49	95.63	90.65	83.18	95.32	91.44	86.96	94.85	91.21	83.85	97.59
	No.2&3	90.80	95.30	90.55	83.82	95.08	91.41	87.44	94.67	91.26	83.93	97.32
	No.1&2&3	90.87	95.72	90.79	83.80	95.48	91.75	87.46	94.83	91.47	84.29	97.65
Segformer	No.1	90.24	94.84	89.96	83.42	95.17	90.85	86.40	94.66	90.83	83.21	97.22
	No.2	90.59	95.16	90.61	83.95	95.32	91.65	87.89	94.72	91.36	84.10	97.47
	No.3	90.74	95.55	90.41	82.27	95.15	91.38	86.47	94.46	90.97	83.43	97.51
	No.1&2	90.81	95.84	91.21	84.49	95.68	91.66	86.97	94.97	91.59	84.48	97.73
	No.1&3	90.17	95.68	90.99	84.37	95.64	91.87	87.46	95.00	91.53	84.38	97.68
	No.2&3	90.82	95.54	90.70	83.43	95.62	92.00	87.83	95.09	91.52	84.37	97.70
	No.1&2&3	91.28	95.84	91.15	84.34	95.80	92.01	87.53	95.08	91.76	84.78	97.77
TransUnet	No.1	91.03	95.60	90.18	83.04	95.50	91.51	87.42	95.27	91.35	84.07	97.67
	No.2	91.67	95.85	90.80	84.26	95.64	91.92	87.18	95.03	91.68	84.64	97.75
	No.3	90.41	95.69	90.77	83.37	95.57	91.84	86.17	94.53	91.20	83.82	97.62
	No.1&2	91.89	96.03	91.06	84.49	95.79	92.15	87.69	95.18	91.92	85.04	97.82
	No.1&3	91.71	95.94	91.06	84.14	95.80	92.21	88.37	95.30	91.95	85.10	97.81
	No.2&3	91.68	95.82	90.77	84.38	95.67	92.20	87.85	95.26	91.83	84.91	97.76
	No.1&2&3	91.92	95.91	91.25	84.90	95.83	92.21	87.84	95.28	92.02	85.22	97.83

is 0.24 and 0.31 percentage points higher than No.1 and No.3, respectively; In the Segformer method, we find that the overall segmentation accuracy of No.2 is higher, while No.1 is lower. On the mIoU index, No.2 is 0.89 and 0.67 percentage points higher than No.1 and No.3, respectively. We analyze that there are two main reasons for this difference: Firstly, these OCT images are different because they are generated from different sub-band spectrum. Secondly, there are differences in the feature extraction ability of different DL models, which we infer may be the reason for the differences in RLS between different spectral images.

When we input different spectral images pairwise, we find that there are significant differences in segmentation between different combinations, but from the perspective of evaluation indicators, this difference is significantly reduced. In the TransUnet method shown in Tab. II, the three combinations of No.1&2, No.1&3, and No.2&3 have a difference of only 0.12 percentage points between their maximum and minimum values on the average of each layer in the Dice Score, while the difference between their maximum and minimum values in No.1, No.2, and No.3 is 0.48 percentage points. Similarly, in the Segformer method, they are 0.07 and 0.53 percentage points, respectively. This indicates that after adding spectral information, the difference in segmentation accuracy between pairwise spectral image combinations is smaller than the difference between single spectral images. We infer that the primary reason is that there are already some common factors between the combinations of spectral images, which reduce the differences, such as the presence of a No.1 in both No.1&2 and No.1&3.

When we input three images from the combination of No.1&2&3 into the network simultaneously, we find that the segmentation accuracy of the entire retina layer reaches the highest level. Specifically, on the mIoU metric, the LightReSeg method outperforms No.3 by 1.13 percentage points on the No.1&2&3 combination, the Segformer method outperforms No.1 by 1.57 percentage points on the No.1&2&3 combination, and the TransUnet method outperforms No.3 by 1.4 percentage points on the No.1&2&3 combination.

Tab. II all shows the average trend of three segmentation methods on the Dice Score in three different combinations. We can see that when using a single spectral image, the segmentation accuracy of all three methods is the lowest. When combining pairwise spectral images, the average segmentation accuracy of all three methods is improved. When three different spectral images are simultaneously input into the network, the segmentation accuracy of all three segmentation methods reaches the highest level. We infer that with the increase of spectral information, the segmentation accuracy of various methods will show a trend of improvement, which also means that introducing spectral information will be beneficial for improving the segmentation accuracy of retinal layers.

In addition to quantitative analysis, we also conduct qualitative analysis to provide a more comprehensive evaluation. As shown in Fig. 4, we randomly select two images from the test dataset in each method to observe the effect of different spectral combinations on segmentation ability. In the LightReSeg method, we can see that there are obvious intra-class errors in the predicted images of group No.1 (RPE category results appeared in the NFL category), while in groups No.1&2, we find that this intra-class error disappeared directly, which we believe is the result of introducing spectral

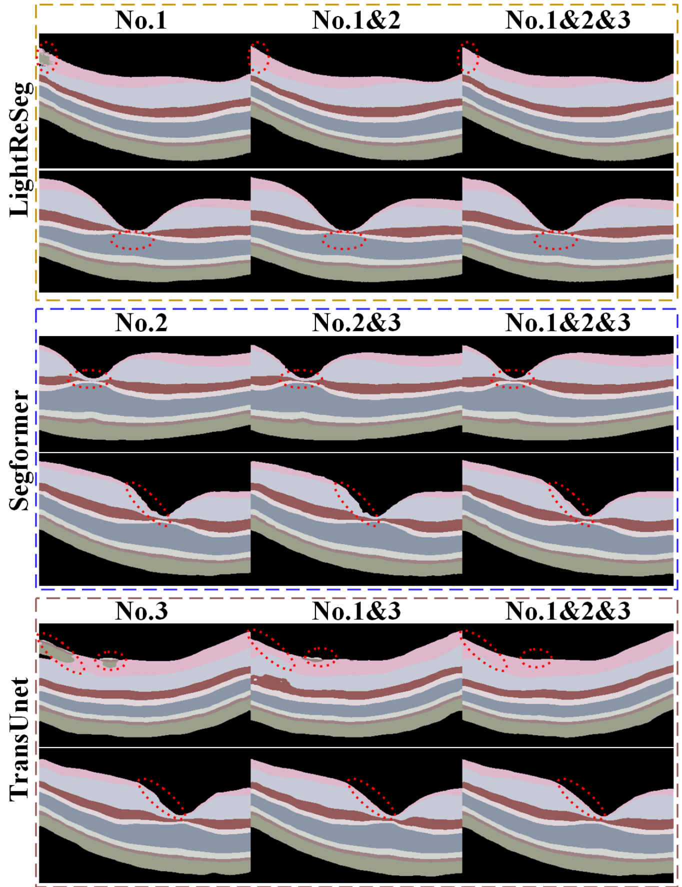

Fig. 4. Qualitative analysis of the segmentation performance of three segmentation methods for each spectral combination.

information. In another predicted image of Group No.1, there is also a clear inter class ambiguity in the macular area. However, as the number of spectral images increased, we observe that the stratification of the macular area achieved the best segmentation results in Groups No.1&2&3. For example, the TransUnet method showed significant intra-class errors in the first predicted image of group No.3, which are significantly reduced in groups No.1&3 and completely disappeared in groups No.1&2&3. This further confirms the significant improvement effect of spectral information on RLS performance.

D. Limitation

In this study, we cannot determine whether an unlimited increase in spectral information will consistently improve the segmentation accuracy of RLS, as our combination only consists of up to three different spectral images. If we continue to increase the number of images input into the network simultaneously, the results may increase, decrease, or even remain unchanged. In addition, this study use the method of directly merging different spectral images into the network, which may not be the best way to utilize spectral information. In the future, spectral information may have a separate carrier like structural information, such as image data or point cloud data.

IV. Conclusion

In this study, we investigate for the first time the differences between spectral images with different central wavelengths and confirm that spectral data helps improve the accuracy of RLS. And we also find that increasing the number of spectral images within a certain range gradually improves the accuracy of RLS. In the future, we will further explore the factors affecting spectral information and conduct disease-based RLS research based on spectral information.

References

[1] Z. Mammo, M. Heisler, C. Balaratnasingam, S. Lee, D.-Y. Yu, P. Mackenzie, S. Schendel, A. Merkur, A. Kirker, D. Albiani *et al.*, “Quantitative optical coherence tomography angiography of radial peripapillary capillaries in glaucoma, glaucoma suspect, and normal eyes,” *American journal of ophthalmology*, vol. 170, pp. 41–49, 2016.

[2] V. Guedes, J. S. Schuman, E. Hertzmark, G. Wollstein, A. Correnti, R. Mancini, D. Lederer, S. Voskanian, L. Velazquez, H. M. Pakter *et al.*, “Optical coherence tomography measurement of macular and nerve fiber layer thickness in normal and glaucomatous human eyes,” *Ophthalmology*, vol. 110, no. 1, pp. 177–189, 2003.

[3] M. Gende, V. Mallen, J. de Moura, B. Cordón, E. Garcia-Martin, C. I. Sánchez, J. Novo, and M. Ortega, “Automatic segmentation of retinal layers in multiple neurodegenerative disorder scenarios,” *IEEE Journal of Biomedical and Health Informatics*, 2023.

[4] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, “Optical coherence tomography,” *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[5] S. Zhou, D. Yu, Y. Cai, Y. Zhang, B. Li, and W. Li, “Tcam-resnet: A convolutional neural network for screening dr and amd based on oct images,” in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 1830–1835.

[6] Y. Xiong, Y. Luo, Q. Tang, Y. Zhang, and B. Du, “Cervical oct image synthesis via noise adaptation diffusion models,” in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 2310–2313.

[7] T. Dan, Z. Fan, Y. Hu, B. Zhang, G. Tao, and H. Cai, “Reconstruction of 3d retina from multi-viewed stereo fundus images via dynamic registration,” in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 980–984.

[8] Y. Tan, W.-D. Shen, M.-Y. Wu, G.-N. Liu, S.-X. Zhao, Y. Chen, K.-F. Yang, and Y.-J. Li, “Retinal layer segmentation in oct images with boundary regression and feature polarization,” *IEEE Transactions on Medical Imaging*, 2023.

[9] X. He, W. Song, Y. Wang, F. Poiesi, J. Yi, M. Desai, Q. Xu, K. Yang, and Y. Wan, “Light-weight retinal layer segmentation with global reasoning,” *IEEE Transactions on Instrumentation and Measurement*, 2024.

[10] Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, T. Wang, S. Qi, J. Mei, and W. Wang, “Deepretina: layer segmentation of retina in oct images using deep learning,” *Translational vision science & technology*, vol. 9, no. 2, pp. 61–61, 2020.

[11] B. Wang, W. Wei, S. Qiu, S. Wang, D. Li, and H. He, “Boundary aware u-net for retinal layers segmentation in optical coherence tomography images,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3029–3040, 2021.

[12] W. Song, L. Zhou, S. Zhang, S. Ness, M. Desai, and J. Yi, “Fiber-based visible and near infrared optical coherence tomography (vnoct) enables quantitative elastic light scattering spectroscopy in human retina,” *Biomedical Optics Express*, vol. 9, no. 7, pp. 3464–3480, 2018.

[13] Z. Huang, Z. Jiang, Y. Hu, D. Zou, Y. Yu, Q. Ren, G. Liu, and Y. Lu, “Retinal choroidal vessel imaging based on multi-wavelength fundus imaging with the guidance of optical coherence tomography,” *Biomedical Optics Express*, vol. 11, no. 9, pp. 5212–5224, 2020.

[14] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, “Segformer: Simple and efficient design for semantic segmentation with transformers,” *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.

[15] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, “Transunet: Transformers make strong encoders for medical image segmentation,” *arXiv preprint arXiv:2102.04306*, 2021.