Hindawi
Journal of Healthcare Engineering
Volume 2022, Article ID 2722608, 9 pages
https://doi.org/10.1155/2022/2722608

Research Article

Multiview Volume and Temporal Difference Network for Angle-Closure Glaucoma Screening from AS-OCT Videos

Luoying Hao,[1] **Yan Hu**,[1] **Risa Higashita**,[2] **James J. Q. Yu**,[1] **Ce Zheng**,[3] **and Jiang Liu**[1,4,5]

[1]*Department of Computer Science and Engineering, Southern University of Science and Technology, 518055 Shenzhen, China*
[2]*Tomey Corporation, 451-0051 Nagoya, Japan*
[3]*Department of Ophthalmology, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China*
[4]*School of Ophthalmology & Optometry, School of Biomedical Engineering, Wenzhou Medical University, Zhejiang, China*
[5]*Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, 518055 Shenzhen, China*

Correspondence should be addressed to Yan Hu; huy3@sustech.edu.cn

Received 16 December 2021; Accepted 15 March 2022; Published 7 April 2022

Academic Editor: Weihua Yang

Background. Precise and comprehensive characterizations from anterior segment optical coherence tomography (AS-OCT) are of great importance in facilitating the diagnosis of angle-closure glaucoma. Existing automated analysis methods focus on analyzing structural properties identified from the single AS-OCT image, which is limited to comprehensively representing the status of the anterior chamber angle (ACA). Dynamic iris changes are evidenced as a risk factor in primary angle-closure glaucoma. *Method*. In this work, we focus on detecting the ACA status from AS-OCT videos, which are captured in a dark-bright-dark changing environment. We first propose a multiview volume and temporal difference network (MT-net). Our method integrates the spatial structural information from multiple views of AS-OCT videos and utilizes temporal dynamics of iris regions simultaneously based on image difference. Moreover, to reduce the video jitter caused by eye movement, we employ preprocessing to align the corneal part between video frames. The regions of interest (ROIs) in appearance and dynamics are also automatically detected to intensify the related informative features. *Results*. In this work, we employ two AS-OCT video datasets captured by two different devices to evaluate the performance, which includes a total of 342 AS-OCT videos. For the Casia dataset, the classification accuracy for our MT-net is 0.866 with a sensitivity of 0.857 and a specificity of 0.875, which achieves superior performance compared with the results of the algorithms based on AS-OCT images with an obvious gap. For the Zeiss AS-OCT video dataset, our method also gets better performance against the methods based on AS-OCT images with a classification accuracy of 0.833, a sensitivity of 0.860, and a specificity of 0.800. *Conclusions*. The AS-OCT videos captured under changing environments can be a comprehended means for angle-closure classification. The effectiveness of our proposed MT-net is proved by two datasets from different manufacturers

1. Introduction

Glaucoma is an eye disease with extremely complex etiology, ranking second among the four major blinding eye diseases. By 2040, it is estimated that 112 million people in the world will be affected by this disease [1, 2]. Globally, glaucoma (40–80 years old) is estimated to increase to 66–80 million people worldwide by 2020, and 11 million of these patients will eventually become blind [1]. With the ageing of the population, the number of glaucoma patients is increasing year by year. In China, primary angle-closure glaucoma (PACG) is more prevalent. But fortunately, it is preventable after early treatment of anterior chamber angle (ACA), such as laser peripheral iridotomy (LPI). Therefore, early screening and treatment are critical. Recently, anterior segment optical coherence tomography (AS-OCT) is widely accepted by ophthalmologists in glaucoma examination because of its efficient and noncontact imaging anterior chamber with depth information [3].

The shallow anterior chamber is an important risk factor for PACG [4–6], so ophthalmologists often judge the open

or closure status of ACA from AS-OCT. Some computer-aided angle-closure classification algorithms based on ACA are proposed to reduce the doctors' burdens based on machine learning [7–10] or convolutional neural network (CNN) [4, 11, 12]. Most of the present algorithms give out the classification results based on several statically captured AS-OCT images. However, static anatomical factors alone cannot fully explain the relatively high prevalence of PACG and dynamic changes of the anterior chamber structure are more convincing for the diagnosis [13]. For example, as shown in Figure 1, we randomly selected two video samples with PACG and normal ACA. Figure 1(a) is a PACG video sample with the angle status in dark (3rd frame) and bright conditions (55th frame), while Figure 1(b) shows a normal sample with the angle status in dark (4th frame) and bright conditions (34th frame). The video frames under light conditions in Figure 1 are compared when the pupil contracts to the maximum. For the two samples, it is noted that the ACA status is almost closed in dark environments, but after light illumination, it is changed to open. It will lead to inconsistent results for the same sample if only based on a single image.

Thus, it is difficult to distinguish the patients' types only by statically captured AS-OCT images, and most of the present angle-closure classification methods, only based on the angle status of a certain state, have certain limitations [14–16]. But it is correctly classified by the iris motion state (such as the iris motion information as shown in Figure 1, which also can better reflect the complete angle state of the eyes at different times). There is some research explaining this phenomenon. The iris is spongy and compressible in the eyes of healthy and PACG subjects, but it is incompressible in the eyes of PACG and suspected angle-closure [17]. Moreover, the movement features of angle-closure eyes and angle-opening eyes are researched, and the angle-closure group has a slower iris contraction speed in the reflection of light, which is faster after receiving effective treatment [18]. Iris elastic acceleration and pupil block acceleration are correlated with PACG [19]. Therefore, in this article, the angle-closure detection is based on the AS-OCT videos, which are captured in the dark-bright-dark changing environments. As far as we know, there is no research on angle-closure detection concerning the movement of iris based on AS-OCT videos.

In this article, a deep learning-based framework is proposed for angle-closure detection that makes use of AS-OCT videos. The contributions are summarized as follows: (1) we first propose to detect the chamber angle status based on AS-OCT videos in changing environments, which are proven to be more complete representation of the patients' anterior chamber. (2) We propose a multiview volume and temporal difference network (MT-net) for ACA status detection, which integrates the spatial structural information from multiple views of AS-OCT videos and simultaneously utilizes temporal dynamics based on image difference. (3) We propose an automated AS-OCT video alignment algorithm based on the corneal part in video frames, to reduce the impacts of video jitter. Regions of interest (ROIs) in 3D appearance and dynamics are also detected based on the position of the scleral spur (SS) and image difference to enlarge the informative features. (4) We carry out comparison and ablation study experiments to demonstrate the effectiveness of our proposed algorithm by seven evaluation metrics based on two AS-OCT video datasets.

2. The Proposed Method

Figure 2 illustrates the framework of our proposed MT-net (short of multiview volume and temporal difference network). First, the AS-OCT video jitter is removed by the automated image registration method, and the ACA is located by extracting the position of the SS, while motion information is obtained by image difference. Then, the proposed MT-net is introduced that multiple views of ACA volumes are fed to extract spatial features, while the motion feature is input to study temporal information of iris dynamic. Finally, the prediction scores based on spatial and temporal information are integrated to further enhance the performance of angle-closure detection.

2.1. AS-OCT Video Alignment and ROIs Extraction

2.1.1. AS-OCT Video Alignment. Due to the impacts of involuntary eye movement and improper placement of the optical axis of the eye, misalignment exists between adjacent video frames. As shown in Figure 3, the corneal in the 1st and 38th frame cannot overlap, which may lead to the resulting video frame sequence being unreliable [20].

Assume a video contains N frames, and the frames are denoted by $f_i (i \in [1, N])$. To ensure the consistency of the placement of the anterior chamber structure in the video frames, we transform the frames $f_i (i \in [2, N])$ into the coordinate system of frame f_1 and crop the transformed frames to be the same dimension as f_1. First, the multiscale face point features p^f and corner-like features p^c are extracted from the frames. Rotation, translation, and scale are considered the main changes between video frames; thus, the affine transformation parameters θ are estimated based on the similarity metrics, and an iterative optimization process is further used to refine the transformation, defined as follows:

$$\mathscr{E}\left(\theta; \zeta_f, \zeta_c\right) = \operatorname{argmin}\left(\sum_{\left(p_i^f, q_i^f\right) \in \zeta_f} \omega_f \rho\left(d_f\left(p_i^f, q_i^f, \theta\right)\right) + \sum_{\left(p_i^c, q_i^c\right) \in \zeta_c} \omega_c \rho\left(d_c\left(p_i^c, q_i^c, \theta\right)\right) \right), \tag{1}$$

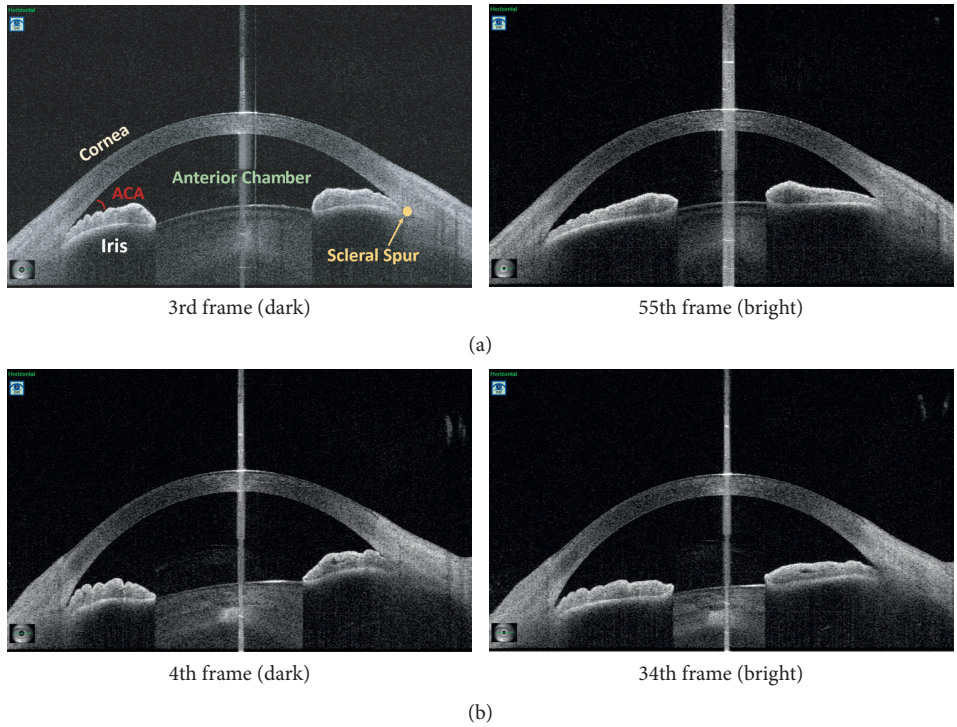

Figure 1: The example of (a) angle-closure and (b) open angle (normal) video.

where $d_f(\cdot)$ is the normal distances of pair face points, $d_c(\cdot)$ is the Euclidean distances of pair corner points, and $\rho(\cdot)$ is the Beaton–Tukey [21]. The face point feature matching sets and corner feature matching sets are denoted as $(p_i^f, q_i^f) \in \zeta_f$ and $(p_i^c, q_i^c) \in \zeta_c$, respectively. The ω_f and ω_c are the distance-based robust weight factors.

Besides, to speed up the alignment procedure, median filter and frame resize are adopted before alignment. As shown in Figure 3(b), the corneal is overlapped between frames after alignment.

2.1.2. ROIs Extraction. The ACA and iris region are the ROIs during ophthalmologists diagnosing PACG [22]. In this study, ROI extraction includes ACA extraction and image difference, which can reinforce ACA spatial and iris temporal representation.

(1) ACA Extraction. Locating local regions can retain more useful information at the last feature map of the backbone network [11, 23–25]. The SS is the key point of the ACA; thus, we obtain the ACA for angle status detection by SS localization in the article. We propose to use a UNet-like architecture based on nested and dense skip connections (UNet++) [26] to get accurate SS localization. Then, the ACAs are cropped directly from aligned videos and resized to one fixed resolution. In this way, the network can focus on visual contents by cropped bounding boxes. Moreover, the scenes of frame inputs are enlarged to capture more useful visual content.

(2) Image Difference. To better extract long-term temporal information, a motion representation is carried on to obtain iris motion first. For motion modelling, the optical flow has been used extensively as a motion representation [27, 28]. However, the extraction of optical flow is expensive in both time and space, which is often calculated in advance and then stored in hard drives. Motivated by this, efforts have been made to find good alternatives. Researchers [29, 30] found that the difference between adjacent frames, namely, image difference, can be useful instead of optical flow.

In this study, image difference, also known as the Eulerian motion, is used to represent the motion of images [31]. Instead of calculating the motion between consecutive frames in a video, this article puts the focus on the iris change compared to the first frame. As shown in Figure 2, the image difference of two images is defined as $V = I_t - I_1$, where I_t is a frame within the scope of $[2, N]$, while I_1 is the first frame in a video. Image differences can capture the short-term motion information to effectively facilitate to model longer-range temporal relations in videos.

2.2. MT-net Framework. The proposed MT-net framework is composed of two subnetworks, multiview volume subnetwork for spatial information (as shown in Figure 2(a)) and temporal difference subnetwork (as shown in Figure 2(b)) for temporal information.

2.2.1. Multiview Volume Subnetwork. The ACAs contain spatial information in the video frame sequence. In this work, the ACAs are composed as a volume with size $H \times W \times T$ as Figure 2(a)(a1), which provides context

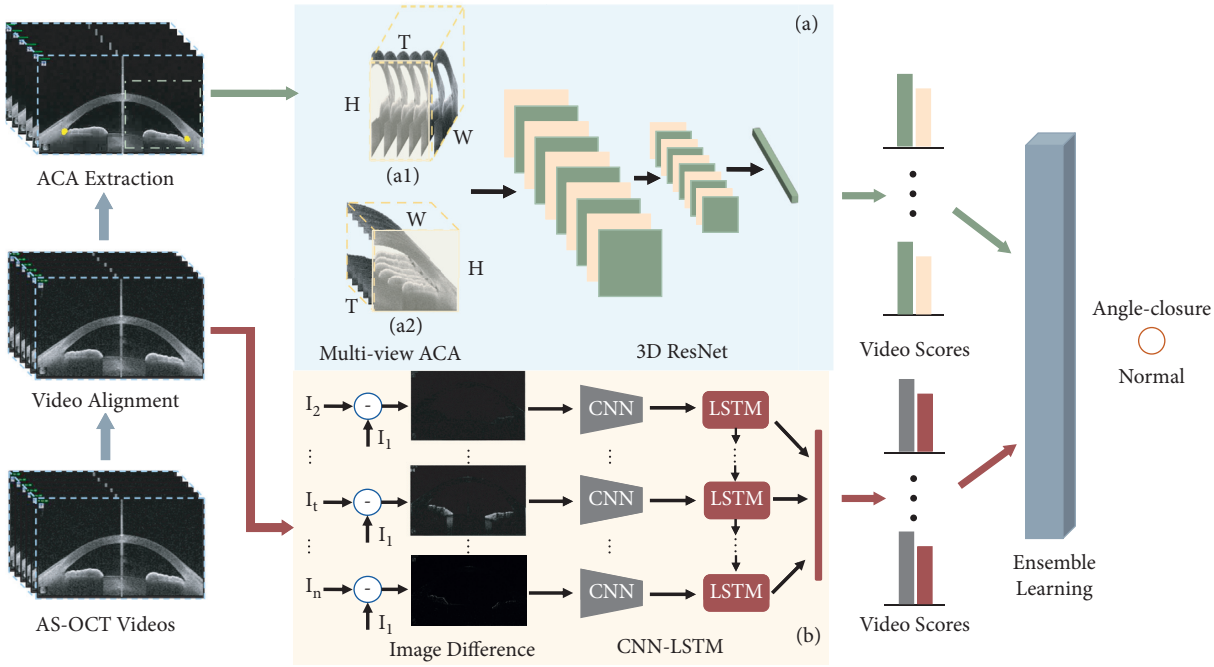

Figure 2: The pipeline of our MT-net architecture. The input AS-OCT videos are aligned to reduce the video jitter. Then, the ACA extraction and image difference are carried on for the two subnetworks: (a) multiview volume network and (b) temporal difference network. Finally, a soft voting-based ensemble model is adopted to incorporate the two subnetworks to output the final classification results.

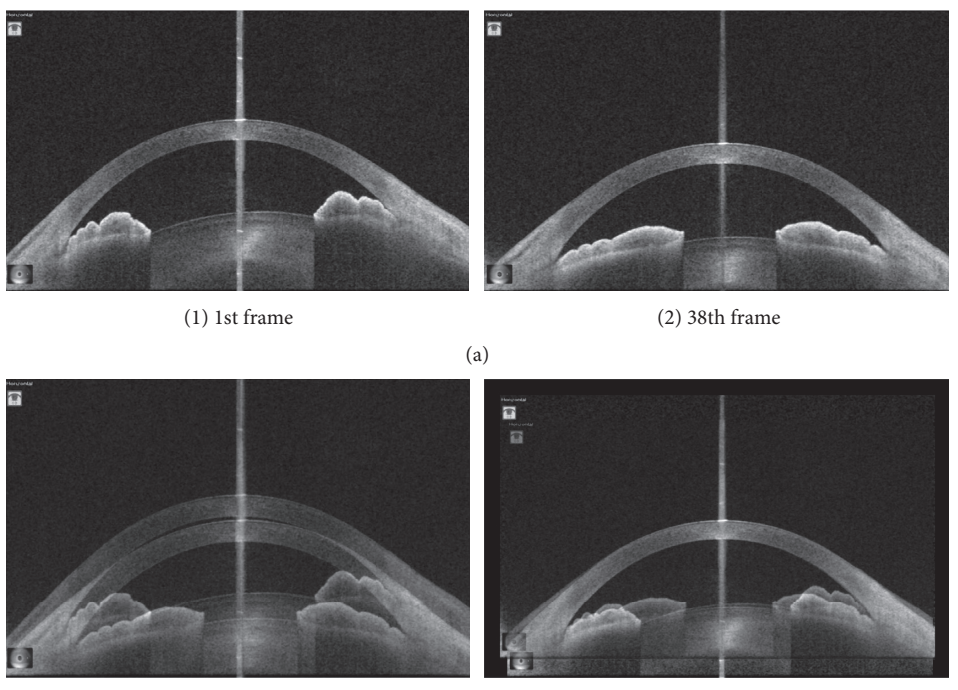

Figure 3: The example of an angle-closure video alignment. (a) Origin angle-closure video. (b) Alignment effect.

information of ACAs in the time dimension T. During the volume analysis, we find that when the volume is rotated with size $H \times T \times W$ as Figure 2(a)(a2), it reveals the fluctuation characteristics of ACAs. Thus, to adopt more useful information for angle status classification, we propose a multiview volume subnetwork by integrating the above different-view volumes. The 3D ResNet is adopted as the backbone since it makes full use of the 3D context

information and is easier to optimize with high accuracy from considerably increased depth [32]. The sizes of convolutional kernels in 3D ResNets are $3 \times 3 \times 3$, and both the temporal and spatial stride are 2. The 16-frame ACA clips are input into the subnetwork with the size of $3 \times 16 \times 224 \times 224$. Since the small scale of the medical image dataset is the main reason for low classification accuracy, fine-tuning pretrained mode on large-scale datasets becomes an effective way [33]. We also fine-tune the pretrained 3D ResNets model on Kinetics [34]. Also, identity connections and zero paddings for the shortcuts of the ResNet block are utilized to avoid the increasing number of parameters [35].

2.2.2. Temporal Difference Subnetwork. As the feature of iris dynamic movement under the dark-light-dark environment is helpful for the angle-closure state classification, temporal information of the AS-OCT videos is adopted in the article. To reduce the computation complexity of the subnetwork, we propose to apply a ResNet model to extract features of image difference. Then, the extracted features are input into the long short-term memory (LSTM) layer with batch normalization [36], which encodes the states and models the long-term dependencies between the feature map along the time axis. Finally, a fully connected layer on the top of LSTM output is adopted for multiway classification [34].

2.2.3. Angle-Closure Detection. Temporal information plays an important role in understanding the iris motion, while ACA volume provides anatomical features of the anterior segment at different times. We take into account two kinds of context information in our model: scene volume context and temporal changing information over the entire span of videos. Finally, we adopt the model ensemble, specifically the soft voting ensemble method [37], to integrate multifaceted contents and obtain a more comprehensive and accurate classification result. The soft voting ensemble method is a soft variant of a voting scheme that takes into account the class probabilities of each algorithm and combines these decisions through the averaging process, instead of hard voting through on-off decisions [37]. In this article, we independently train each subnetwork, get the probability distribution of the test set (As shown in Figure 2), and finally synthesize the performance of different classifiers of each subject to get the final classification results.

3. Experimental Results

3.1. Clinical AS-OCT Video Dataset. Our AS-OCT video datasets are collected by two devices: Swept-source OCT [38] (Casia Swept-source-1000 OCT, Tomey, Nagoya, Japan) and Visante OCT [39] (Visante OCT, Model 1000, software version 2.1; Carl Zeiss Meditec). We collect the AS-OCT videos of normal people and patients with PACG under a dark-light-dark environment. Subjects are recruited from the outpatient and inpatient departments of the Singapore National Eye Centre (SNEC) and joint Shantou International Eye Centre of Shantou University and the Chinese University of Hong Kong, which include patients and volunteers aged over 40 years. In particular, the recording of the AS-OCT videos is started one minute after dark adaption using a standard protocol, and the light intensity is approximately 20 lux. The iris and anterior chamber changes between the dark and light environments are recorded. A single ophthalmologist performs all AS-OCT testing for data consistence. For each video, the ground-truth label of normal or angle-closure is determined from the majority diagnosis of senior ophthalmologists.

For the Casia dataset, it includes 148 videos, including 68 videos of normal eyes and 80 videos of eyes with PACG. The resolution of video frames is 1644×1000. The Zeiss dataset consists of 194 videos, including 116 videos of normal eyes and 78 videos of eyes with PACG. The resolution of video frames is 600×300. For the two datasets, Table 1 lists the maximum, minimum, and median of video frames. We equally and randomly divide 30 videos as the testing set, while the remaining videos are divided into the validation set and training set. The size of all input video frames for the deep learning network is fixed at 224×224.

3.2. Implementation Details. The proposed architecture is implemented using the publicly available PyTorch Library. In the training phase, for the multiview volume subnetwork, we utilize stochastic gradient descent to optimize the model (200 epochs), with a gradually decreasing learning rate starting from 0.1, a momentum of 0.9, and a batch size of 128. For the temporal difference subnetwork, we employ an Adam optimizer to optimize the model (180 epochs), with a learning rate of 0.0001, a momentum of 0.01, and a batch size of 128. For all the processes of training and testing, we conduct them on one NVIDIA TITAN V GPU.

3.3. Experimental Criterion and Baseline. To measure the performance of our network, we employ seven evaluation criteria: balanced accuracy (B-Acc), precision (Pre), recall, F1 score, sensitivity (Sen), specificity (Spe), and Kappa analysis. Kappa analysis and F1 score are used to reflect the trade-offs between Sen and Spe.

As shown in Table 2, we use the basic subnetwork backbones of 3D CNN and CNN-LSTM to conduct training and testing on our private Casia dataset. For a small-scale medical image dataset, different proportions of validation set and training set affect the anterior chamber status classification. We conduct experiments for the two subnetworks with the proportion of validation set and training set to 5%, 10%, and 20%, and the results are shown in Table 2.

For 3D CNN, it can be seen from Table 2 that 3D ResNet18 has the highest B-Acc and F1 score of the three dataset splits. In the training process, the relatively shallow network is easier to converge than the deeper network. For the experiment of CNN-LSTM, the ResNets are fine-tuned from initialization with the pretrained deep model. As shown in Table 2, based on the same testing set, the B-Acc and F1 scores of this network are basically higher than that of 3D CNN. The possible reason is that CNN-LSTM models the global movement of the iris better, which also further proves

Table 1: The maximum, minimum, and median of video frames for the two datasets.

	Maximum	Minimum	Median
Casia dataset	121	21	53
Zeiss dataset	135	20	48

Table 2: Performance of different subnetworks on the private Casia video dataset.

3D CNN (B-Acc/F1 score)					
Splits	18-Layer	34-Layer	50-Layer	101-Layer	152-Layer
5%	0.638/ 0.632	0.464/ 0.282	0.625/ 0.627	0.562/ 0.430	0.558/ 0.463
10%	**0.692/ 0.695**	0.518/ 0.463	0.612/ 0.589	0.594/ 0.487	0.562/ 0.430
20%	0.589/ 0.589	0.482/ 0.437	0.509/ 0.492	0.589/ 0.514	0.562/ 0.430
CNN-LSTM (B-Acc/F1 score)					
Splits	18-Layer	34-Layer	50-Layer	101-Layer	152-Layer
5%	0.531/ 0.367	0.643/ 0.614	0.777/ 0.763	0.607/ 0.562	0.719/ 0.678
10%	0.500/ 0.371	0.679/ 0.662	**0.781/ 0.789**	0.714/ 0.707	0.656/ 0.589
20%	0.500/ 0.297	0.754/ 0.757	0.710/ 0.695	0.714/ 0.707	0.571/ 0.505

The bold values indicate the optimal results.

that iris motion features are important to predict the binary classification (angle status) result. The testing accuracy of CNN-LSTM shows the best performance at the 50th layer with the increase in depth. The performance of both 3D CNN and CNN-LSTM on the data splits of 5 % and 10 % is much better than those of 20 % . Therefore, in the follow-up experiments, we conduct training on the two dataset splits and take the average testing values as final results.

3.4. Ablation Study. To evaluate the effectiveness of four modules in our framework, including alignment, ACA extraction, image difference, 3D CNN, and CNN-LSTM, we provide an ablation study. Based on the baseline experiments, we employ 3D ResNet18 and ResNet50-LSTM as baselines in the following experiments, and the results are reported in Table 3.

The scleral spur localization is very important for the classification, Thus, in the article, we adopt UNet++ to get accurate SS localization. The model is trained based on the public AGE dataset [6], which is similar to our dataset. For very few video frames that cannot locate SS, we get it from the SS position of the frame preceding the current frame of the aligned video.

(i) For the volume spatial information, video alignment and ACA region extraction improve the classification results of 3D CNN to a certain extent compared with the baselines. When the two preprocesses are combined, all the evaluation metrics increase. It is noted that the results combined with the multiviews are better than those from only one general view.

(ii) For temporal information, it illustrates the importance of global change in the iris regions for improving classification performance. For CNN-LSTM, although its testing performance is not promoted much after extracting the iris motion information (image difference), it significantly improves after the video is aligned. When image difference is combined with video alignment, the evaluation metrics further increase, which indicates the negative effect of video jitter on the extraction of iris dynamic features. The temporal information is helpful for the classification.

(iii) For volume spatial and temporal information, the alignment, ACA extraction, and image difference improve the results, as shown in Table 3. The results in the last line achieve optimal performance by integrating the multiview spatial, temporal, and preprocessing, which is our proposed framework, MT-net.

3.5. Performance on the Two Private AS-OCT Video Datasets. To prove the superiority of classification based on the AS-OCT videos, we compare our framework with the present algorithm based on single AS-OCT images. We select frames from the beginning and end of our videos taken under a dark environment, which is the same as the datasets of most of the present classification algorithms [4, 11, 12]. For the Casia dataset, the selected images are combined into a training set with a total of 2160 AS-OCT images (1230 angle-closure and 930 normal images) and a testing set with 520 AS-OCT images (250 angle-closure and 270 normal images) with the same distribution as the video dataset. For the Zeiss dataset, the extracted image dataset contains a training set with 3380 AS-OCT images (1360 angle-closure and 2020 normal images) and a testing set with 500 AS-OCT images (200 angle-closure and 300 normal images) with the same distribution as the video dataset.

We use 2D ResNet50, which has the best performance in the baseline experiments, as the comparison algorithm based on the AS-OCT image datasets. The ACA extraction is also combined with 2D ResNet50, and the results are shown in Table 4. To ensure the fairness of comparison, for AS-OCT image datasets, we get final classification results based on each video in the test stage; that is, if the number of correctly classified images accounts for more than 50 % of the total frames of the video, we will give the correct judgment.

As shown in Table 4, for the two datasets, the ACA extraction is helpful for the ACA status classification for all two datasets. But our proposed MT-net based on AS-OCT videos gives the best evaluation metrics. For the Casia dataset, the classification accuracy for our MT-net is 0.866 with a sensitivity of 0.857 and a specificity of 0.875, which achieves superior performance compared with the results of the algorithms based on AS-OCT images with an obvious gap. For the Zeiss dataset, our method based on AS-OCT videos also gets better performance against those based on AS-OCT images with a classification accuracy of 0.833, a sensitivity of 0.860 and a specificity of 0.800. Although the values of sensitivity and specificity are not the highest in Table 4 for the Zeiss dataset, we achieve the highest Kappa

Table 3: Classification performance of the angle-closure glaucoma by different module combinations on private Casia video dataset.

AL[1]	ACA	Diff[2]	C3D[3]	ConvL[4]	B-Acc	Pre	Recall	F1 score	Sen	Spe	Kappa
			✓		0.692	0.704	0.697	0.695	0.718	0.673	0.370
✓			✓		0.712	0.735	0.703	0.701	0.848	0.576	0.493
	✓		✓		0.719	0.720	0.720	0.720	0.706	0.733	0.518
✓	✓		✓		0.755	0.756	0.753	0.754	0.777	0.732	0.587
✓	✓		✓[5]		0.763	0.767	0.767	0.766	0.714	0.813	0.529
				✓	0.781	0.823	0.777	0.789	0.821	0.625	0.545
✓				✓	0.813	0.819	0.816	0.817	0.750	0.860	0.629
		✓		✓	0.607	0.615	0.600	0.596	0.714	0.500	0.210
✓		✓		✓	0.830	0.834	0.833	0.833	0.786	**0.875**	0.664
			✓	✓	0.777	0.804	0.767	0.763	0.728	0.625	0.542
✓	✓	✓	✓	✓	0.820	0.838	0.817	0.814	**0.857**	0.780	0.636
✓	✓	✓	✓[5]	✓	**0.866**	**0.867**	**0.867**	**0.867**	**0.857**	**0.875**	**0.732**

[1]AL: Alignment; [2]Diff: Difference; [3]C3D: 3D CNN; [4]ConvL: CNN-LSTM; [5]3D CNN (multiview). The bold values indicate the optimal results.

Table 4: Comparison of the classification performance on private two AS-OCT video datasets and image datasets.

Casia	B-Acc	Pre	Recall	F1 Score	Sen	Spe	Kappa
ResNet50 (images)	0.767	0.768	0.767	0.766	0.800	0.733	0.533
ACA + ResNet50 (images)	0.774	0.830	0.759	0.748	0.810	0.547	0.530
Our MT-net	**0.866**	**0.867**	**0.867**	**0.867**	**0.857**	**0.875**	**0.732**
Zeiss	B-Acc	Pre	Recall	F1 score	Sen	Spe	Kappa
ResNet50 (images)	0.750	0.775	0.750	0.744	**0.900**	0.600	0.500
ACA + ResNet50 (images)	0.795	0.804	0.800	0.798	0.714	**0.875**	0.594
Our MT-net	**0.833**	**0.840**	**0.840**	**0.840**	0.860	0.800	**0.600**

The bold values indicate the optimal results.

value and F1 score, which are used to reflect the trade-offs between sensitivity and specificity.

4. Discussion

In this study, after extracting multiview spatial information and modelling motion, we develop the MT-net to learn to discriminate 3D spatial and temporal features from AS-OCT videos. Our proposed method is shown to be a promising technology for serving clinicians in faithfully identifying angle-closure in AS-OCT videos with a high classification accuracy. The proposed framework opens the door to further enhance the screening ability of angle-closure-related disease from a brand new perspective. More research is needed to explore the employment of deep learning algorithms deployed in diverse population settings, with the use of multiple devices and larger AS-OCT datasets.

The effectiveness of our proposed MT-net is proved in the above experimental parts. The AS-OCT videos can be a more comprehensive means for angle-closure diagnosis. But the study still has two limitations. One limitation of this study is that it assesses two specific Asian populations (Chinese and Singaporeans) due to the high prevalence of primary glaucoma in Asia, so the results may not be applicable to other ethnic groups. But this effect can be mitigated by increasing the diversity of ethnic data. Another potential limitation is that the AS-OCT videos are captured from Casia and Zeiss, the two famous manufacturers in the world. Because of the difference between the capturing machines, this may adversely affect the quality and performance when our network is applied to videos from other AS-OCT acquisition devices, which did not happen in our present two datasets. If more data can be acquired from other devices in the future, the performance of our model may become more stable and more powerful.

5. Conclusions

We first proposed to detect the ACA status based on light-changing AS-OCT videos in this article. A multiview volume and temporal difference framework (MT-net) is proposed to learn to discriminate spatial and temporal features on the ROIs of AS-OCT videos, which include ACA and iris dynamic changes in the dark-light-dark environment. The ablation experiments prove the effectiveness of our MT-net. The evaluation metrics based on videos are better than those based on 2D AS-OCT images, manifesting that the chamber angle status analysis in a changing environment could improve the ability of angle-closure related disease screening.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to restrictions in the ethical permit but are partly available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (8210072776), the Science and Technology Innovation Committee of Shenzhen City (JCYJ20200109140820699 and 20200925174052004), Guangdong Basic and Applied Basic Research Foundation (2021A1515012195), Guangdong Provincial Department of Education (2020ZDZX3043), and Guangdong Provincial Key Laboratory (2020B121201001). The authors thank the doctors from Xinhua Hospital for the data collection and analysis. The authors also thank the help of our imed Group for the support.

References

[1] X. Li, E. Chan, J. Liao, T. Wong, T. Aung, and C. -Y. Cheng, "Number of people with glaucoma in Asia in 2020 and 2040: a hierarchical bayesian meta-analysis," *Investigative Ophthalmology & Visual Science*, vol. 54, p. 2656, 2013.

[2] Y.-C. Tham, L. Xiang, T. Y. Wong, H. A. Quigley, T. Aung, and C. Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.

[3] D. H. W. Su, D. S. Friedman, J. L. S. See et al., "Degree of angle closure and extent of peripheral anterior synechiae: an anterior segment OCT study," *British Journal of Ophthalmology*, vol. 92, no. 1, pp. 103–107, 2008.

[4] P. J. Foster, F. T. S. Oen, D. Machin et al., "The prevalence of glaucoma in Chinese residents of SingaporeA cross-sectional population survey of the tanjong pagar district," *Archives of Ophthalmology*, vol. 118, no. 8, pp. 1105–1111, 2000.

[5] R. Sihota, D. Ghate, S. Mohan, V. Gupta, R. M. Pandey, and T. Dada, "Study of biometric parameters in family members of primary angle closure glaucoma patients," *Eye*, vol. 22, no. 4, pp. 521–527, 2008.

[6] H. Fu, F. Li, X. Sun et al., "Age challenge: angle closure glaucoma evaluation in anterior segment optical coherence tomography," *Medical Image Analysis*, vol. 66, Article ID 101798, 2020.

[7] M. E. Nongpiur, L. M. Sakata, D. S. Friedman et al., "Novel association of smaller anterior chamber width with angle closure in Singaporeans," *Ophthalmology*, vol. 117, no. 10, pp. 1967–1973, 2010.

[8] Y. Xu, L. Jiang, J. Cheng et al., "Automated anterior chamber angle localization and glaucoma type classification in OCT images," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7380–7383, Osaka, Japan, July 2013.

[9] Y. Xu, L. Jiang, W. K. W. Damon et al., "Similarity-weighted linear reconstruction of anterior chamber angles for glaucoma classification," in *Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 693–697, Prague, Czech Republic, April 2016.

[10] H. Fu, Y. Xu, S. Lin et al., "Segmentation and quantification for angle-closure glaucoma assessment in anterior segment OCT," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1930–1938, 2017.

[11] H. Fu, Y. Xu, S. Lin et al., "Angle-closure detection in anterior segment OCT based on multilevel deep network," *IEEE Transactions on Cybernetics*, vol. 2019, Article ID 2897162, 2019.

[12] H. Fu, M. Baskaran, Y. Xu et al., "A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images," *American Journal of Ophthalmology*, vol. 203, no. 37–45, 2019.

[13] H. A. Quigley, "Angle-closure glaucoma-simpler answers to complex mechanisms: LXVI Edward Jackson memorial lecture," *American Journal of Ophthalmology*, vol. 148, no. 5, pp. 657–669, 2009.

[14] H. A. Quigley, D. M. Silver, D. S. Friedman et al., "Iris cross-sectional area decreases with pupil dilation and its dynamic behavior is a risk factor in angle closure," *Journal of Glaucoma*, vol. 18, no. 3, pp. 173–179, 2009.

[15] A. Narayanaswamy, C. Zheng, S. A. Perera et al., "Variations in iris volume with physiologic mydriasis in subtypes of primary angle closure glaucoma," *Investigative Opthalmology & Visual Science*, vol. 54, no. 1, pp. 708–713, 2013.

[16] F. Aptel and P. Denis, "Optical coherence tomography quantitative analysis of iris volume changes after pharmacologic mydriasis," *Ophthalmology*, vol. 117, no. 1, pp. 3–10, 2010.

[17] H. A. Quigley, "The iris is a sponge: a cause of angle closure," *Ophthalmology*, vol. 117, no. 1, pp. 1-2, 2010.

[18] C. Zheng, C. Y. Cheung, A. Narayanaswamy et al., "Pupil dynamics in Chinese subjects with angle closure," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 250, no. 9, pp. 1353–1359, 2012.

[19] C. Zheng, C. Y. Cheung, T. Aung et al., "In vivo analysis of vectors involved in pupil constriction in Chinese subjects with angle closure," *Investigative Opthalmology & Visual Science*, vol. 53, no. 11, pp. 6756–6762, 2012.

[20] D. Williams, Y. Zheng, P. G. Davey et al., "Reconstruction of 3D surface maps from anterior segment optical coherence tomography images using graph theory and genetic algorithms," *Biomedical Signal Processing and Control*, vol. 25, pp. 91–98, 2016.

[21] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, "Registration of challenging image pairs: initialization, estimation, and decision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1973–1989, 2007.

[22] J. Hao, F. Li, H. Hao et al., "Hybrid variation-aware network for angle-closure assessment in As-Oct," *IEEE Transactions on Medical Imaging*, vol. 41, 2021.

[23] H. Hao, H. Fu, Y. Xu et al., "Open-appositional-synechial anterior chamber angle classification in as-oct sequences," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Manhattan, NY, USA, October 2020.

[24] H. Fu, Y. Xu, S. Lin et al., "Multi-context deep network for angle-closure glaucoma screening in anterior segment OCT," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Manhattan, NY, USA, 2018.

[25] H. Hao, Y. Zhao, Q. Yan et al., "Angle-closure assessment in anterior segment oct images via deep learning," *Medical Image Analysis*, vol. 69, Article ID 101956, 2021.

[26] Z. Zhou, Md M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

[27] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1933–1941, Las Vegas, NV, USA, September 2016.

[28] Y.-H. Joe, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: deep networks for video classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4694–4702, Boston, MA, USA, March 2015.

[29] L. Sun, K. Jia, D.-Y. Yeung, and B. E. Shi, "Human action recognition using factorized spatio-temporal convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4597–4605, Santiago, Chile, December 2015.

[30] L. Wang, Y. Xiong, Z. Wang et al., "Temporal segment networks: towards good practices for deep action recognition," in *Computer Vision - ECCV 2016,* Springer, Manhattan, NY, USA, 2016.

[31] J. Yue-Hei Ng and L. S. Davis, "Temporal difference networks for video action recognition," in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1587–1596, IEEE, Lake Tahoe, NV, USA, March 2018.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[33] K. Hara, H. Kataoka, and Y. Satoh, "Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and Imagenet?" in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6546–6555, Salt Lake City, UT, USA, June 2018.

[34] W. Kay, J. Carreira, K. Simonyan et al., "The kinetics human action video dataset," 2017, https://arxiv.org/abs/1705.06950.

[35] H. Kataoka, T. Wakamiya, K. Hara, and Y. Satoh, "Would mega-scale datasets further enhance spatiotemporal 3D CNNs?," 2020, https://arxiv.org/abs/2004.04968.

[36] J. Donahue, L. A. Hendricks, S. Guadarrama et al., "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2625–2634, Boston, MA, USA, June 2015.

[37] S. Karlos, G. Kostopoulos, and S. Kotsiantis, "A soft-voting ensemble based co-training scheme using static selection for binary classification problems," *Algorithms*, vol. 13, no. 1, 2020.

[38] S. Liu, M. Yu, C. Ye, D. S. C. Lam, and C. Leung, "Anterior chamber angle imaging with swept-source optical coherence tomography: an investigation on variability of angle measurement," *Investigative Opthalmology & Visual Science*, vol. 52, no. 12, pp. 8598–8603, 2011.

[39] Y. Zhang, S. Z. Li, L. Li, M. G. He, R. Thomas, and N. L. Wang, "Dynamic iris changes as a risk factor in primary angle closure disease," *Investigative Opthalmology & Visual Science*, vol. 57, no. 1, pp. 218–226, 2016.