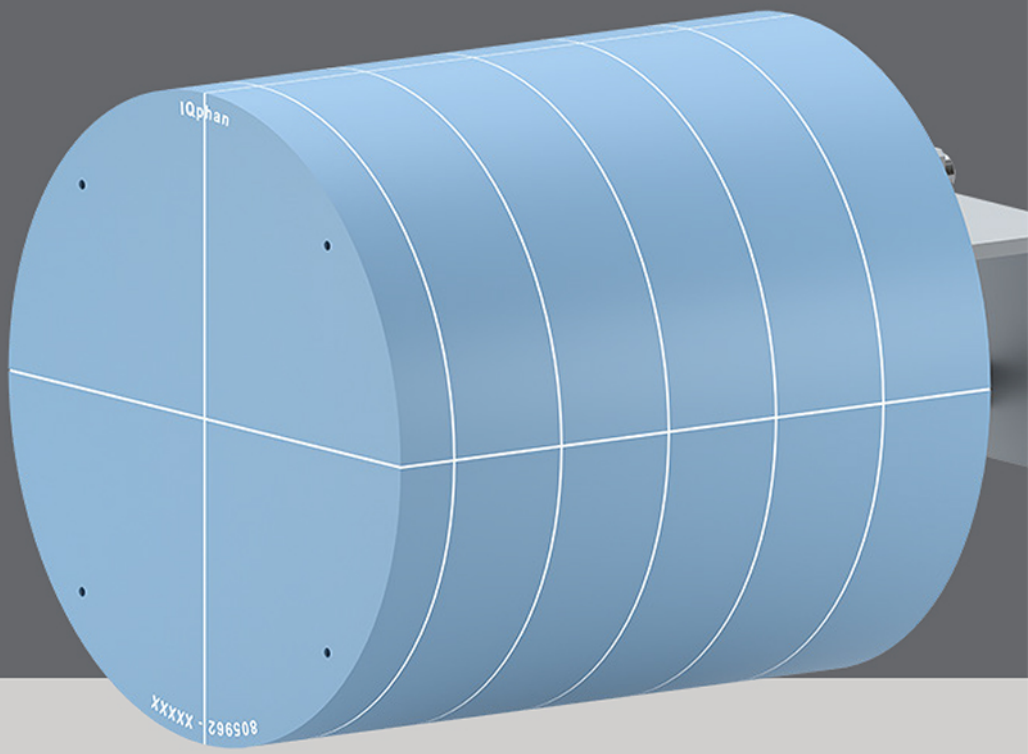
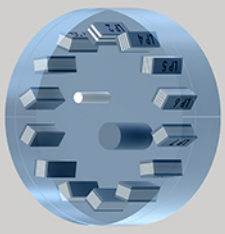
Introducing IQphan™

See IQphan at ECR:
Mirion Medical
booth # 102
Expo X1

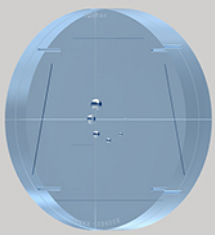
Comprehensive Image Quality Phantom

With IQphan, a single phantom addresses QA across the range of specifications of different CT scanners, enabling you to gain more QA information.



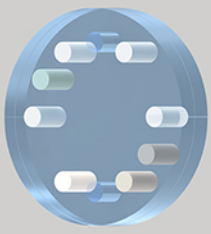


High-Contrast Resolution Module

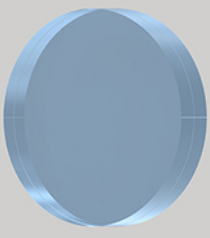


Slice Thickness & Geometric Evaluation Module

Low-Contrast Detectability Module



HU Module



Uniformity Module

sunnuclear.com

SUN NUCLEAR
A MIRION MEDICAL COMPANY

Received: 8 November 2022 | Revised: 9 January 2023 | Accepted: 31 January 2023

DOI: 10.1002/mp.16294

RESEARCH ARTICLE

MEDICAL PHYSICS

Morphological prognosis prediction of choroid neovascularization from longitudinal SD-OCT images

Jiayan Shen[1] | **Zhongyue Chen**[1] | **Yuanyuan Peng**[1] | **Siqi Zhang**[2] | **Chenan Xu**[1] | **Weifang Zhu**[1] | **Haiyun Liu**[2] | **Xinjian Chen**[1,3]

[1]MIPAV Lab, School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu Province, China

[2]Key Laboratory of Ocular Fundus Diseases, Shanghai General Hospital, Shanghai, China

[3]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

Correspondence
Xinjian Chen, MIPAV Lab, School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu Province, 215006, China.
Email: xjchen@suda.edu.cn

Haiyun Liu, Key Laboratory of Ocular Fundus Diseases, Shanghai General Hospital, Shanghai 200080, China.
Email: drliuhaiyun@163.com

Funding information
National Key R&D Program of China, Grant/Award Numbers: 2016YFC0904800, 2018YFA0701700, 2019YFC0840607; National Nature Science Foundation of China, Grant/Award Numbers: U20A20170, 62271337; Natural Science Research of Jiangsu Higher Education Institutions of China, Grant/Award Number: 21KJB510021; Science and Technology Commission of Shanghai Municipality, Grant/Award Number: 201409006700

Abstract

Background: Choroid neovascularization (CNV) has no obvious symptoms in the early stage, but with its gradual expansion, leakage, rupture, and bleeding, it can cause vision loss and central scotoma. In some severe cases, it will lead to permanent visual impairment.

Purpose: Accurate prediction of disease progression can greatly help ophthalmologists to formulate appropriate treatment plans and prevent further deterioration of the disease. Therefore, we aim to predict the growth trend of CNV to help the attending physician judge the effectiveness of treatment.

Methods: In this paper, we develop a CNN-based method for CNV growth prediction. To achieve this, we first design a registration network to rigidly register the spectral domain optical coherence tomography (SD-OCT) B-scans of each subject at different time points to eliminate retinal displacements of longitudinal data. Then, considering the correlation of longitudinal data, we propose a co-segmentation network with a correlation attention guidance (CAG) module to cooperatively segment CNV lesions of a group of follow-up images and use them as input for growth prediction. Finally, based on the above registration and segmentation networks, an encoder-recurrent-decoder framework is developed for CNV growth prediction, in which an attention-based gated recurrent unit (AGRU) is embedded as the recurrent neural network to recurrently learn robust representations.

Results: The registration network rigidly registers the follow-up images of patients to the reference images with a root mean square error (RMSE) of 6.754 pixels. And compared with other state-of-the-art segmentation methods, the proposed segmentation network achieves high performance with the Dice similarity coefficients (Dsc) of 85.27%. Based on the above experiments, the proposed growth prediction network can play a role in predicting the future CNV morphology, and the predicted CNV has a Dsc of 83.69% with the ground truth, which is significantly consistent with the actual follow-up visit.

Conclusion: The proposed registration and segmentation networks provide the possibility for growth prediction. In addition, accurately predicting the growth of CNV enables us to know the efficacy of the drug against individuals in advance, creating opportunities for formulating appropriate treatment plans.

KEYWORDS
choroid neovascularization, growth prediction, optical coherence tomography, registration, segmentation

1 | INTRODUCTION

Choroid neovascularization (CNV) is caused by the proliferation of blood vessels in the choroid, which mostly occurs in the macula. With the expansion or leakage of CNV, retinal pigment epithelium (RPE) layer detachment, and vitreous or retinal hemorrhage will be caused, resulting in serious visual loss.[1] Many fundus diseases can lead to the formation of CNV, of which age-related macular degeneration (AMD) and pathological myopia (PM) are the two main causes.[2,3]

At present, intravitreal injection of anti-vascular endothelial growth factor (VEGF) drugs has become an important way to treat CNV.[4] These drugs can inhibit the growth of CNV by binding and blocking VEGF receptors, reducing vascular permeability and the fluid area below the RPE layer. There are two schemes for the injection of anti-VEGF drugs. One is Pro Re Nata (PRN) scheme, that is, one injection per month. After three consecutive injections, whether to continue the injection is determined according to the patient's condition. The second is the T&E (Treat and Extend) scheme, that is, one injection per month. After three consecutive injections, the injection interval will be gradually extended according to the state of CNV. The datasets used in this paper adopt the first injection scheme. Theoretically, this treatment method is feasible, but in fact, its curative effect varies from person to person. The high-frequency injection of anti-VEGF drugs and the high cost bring a huge burden to patients. Moreover, intraocular injection is invasive and may cause some serious side effects, such as intraocular inflammation.[5]

Functional imaging plays an important role in today's diagnosis and treatment, promoting the rapid development of ophthalmic imaging technology. Among them, fundus fluorescein angiography (FFA), indocyanine green angiography (ICGA), optical coherence tomography (OCT), and optical coherence tomography angiography (OCTA) can be used to diagnose CNV.[6] The application of various auxiliary examinations not only improves the diagnostic accuracy of CNV but also provides a basis for the follow-up treatment plan. OCT is divided into time domain OCT and frequency domain OCT, where frequency domain OCT is divided into spectral domain optical coherence tomography (SD-OCT) and swept source optical coherence tomography (SS-OCT). SD-OCT has a fast imaging speed, providing three-dimensional reconstruction images of the retina with micron resolution, and revealing detailed information about retinal pathological changes. In current clinical practice, SD-OCT equipment is more commonly used to evaluate the retinal morphological changes for follow-up.[7] Figure 1 shows the OCT image slice taken by Heidelberg Spectralis OCT. The macular region of CNV patients is usually accompanied by the generation of subretinal fluid (SRF) and intraretinal fluid (IRF), as shown in Figure 1b.

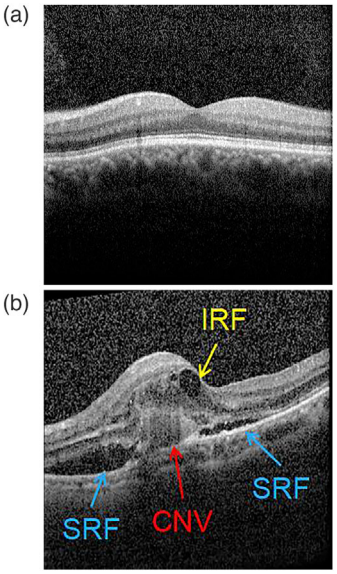


FIGURE 1 Examples of optical coherence tomography (OCT) image slice. (a) Macular region image of normal human eyes. (b) Macular region image of choroid neovascularization (CNV) patients caused by age-related macular degeneration (AMD). IRF, intraretinal fluid; SRF, subretinal fluid.

In recent years, many studies have been carried out on CNV due to its great damage to vision. Specifically, there have been some studies on retinal OCT registration.[8–10] However, these methods are unsuitable for data with serious pathologies, such as CNV. For example, Pan et al.[11] proposed a feature-based 3D registration method that can be used for both normal retinal OCT data and serious pathological OCT data. After the segmentation of seven retinal surfaces, they designed an intensity-based region feature, surface-based structure feature, and vessel-like feature for the registration. This method achieves high registration accuracy and can be extended to other pathological cases. Nevertheless, deformable registration will deform the lesion area, making it more similar to the reference image, which is contrary to our goal of growth prediction.

Most of the work focused on CNV lesion region segmentation[12–14] and retinal layers segmentation[15,16] in the past. Accurately segmenting the lesion area of CNV can quantify the area, volume, width, height, optical density value, and other properties of CNV. Using these quantitative characteristics, ophthalmologists can accurately evaluate the state and treatment effect of CNV, study the pathological characteristics of CNV, and design a more effective treatment plan.

Furthermore, automatic prediction for CNV cases has been applied in clinical trials. For example, Banerjee et al.[17] proposed a mixed sequence prediction model. Combined with the longitudinal OCT imaging radiology characteristics, demographics, and visual factors, the model can predict the probability of exudation in the eyes of patients in the short-term (within 3 months)

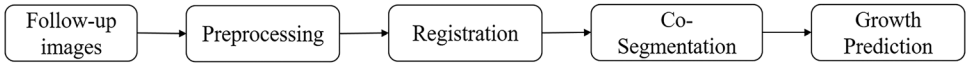


FIGURE 2 Overall flow chart of the proposed scheme.

and long-term (within 21 months) in the future. Zuo Chang et al.[18] combined the reaction-diffusion model with the hyperelastic biomechanical model to predict the growth morphology and volume changes of CNV after anti-VEGF treatment. However, their model only focuses on the independent changes of individual lesion areas, and cannot learn the CNV growth pattern in the whole study population. In addition, the optimization algorithm is very time-consuming, and the accuracy of prediction needs to be further verified on a large dataset. Recently, Zhang et al.[19] proposed a lesion attention map guidance network based on three-dimensional SD-OCT images, where they aimed to automatically predict the CNV volume of the next follow-up after treatment. Different types of CNVs caused by AMD and PM, and OCT images taken by different devices (Zeiss and RTVue) are used to test the performance of the network. The mean absolute error (MAE) between the predicted volume and the actual volume is 0.067 mm^3, 0.112 mm^3, and 0.070 mm^3 respectively, which has reliable prediction performance. It is worth noting that using a classification network to generate CNV attention maps requires high image quality and obvious CNV characteristics. However, the data taken by most hospitals do not meet these two conditions. They usually have a lot of noise, and the CNVs have different shapes or blurred boundaries.

Based on the above research results, we proposed the whole process of the CNV growth prediction method. The major contributions of this paper can be summarized as follows:

1. A rigid registration network based on key points is proposed to register the follow-up images of patients to the reference images, which can eliminate the offset of the longitudinal data in a competitive time.
2. The follow-up images of each patient at four-time points are segmented collaboratively by our proposed co-segmentation model. In the co-segmentation model, we also designed the CAG module. It is used to enhance common abstract features and suppress other irrelevant semantic information so that the module can guide the segmentation of the next time point image.
3. We use the reference image, the examination OCT images after the first injection, and after the second injection to predict the CNV morphology of the third injection. In the growth prediction model, we develop the AGRU module to recurrently learn robust representations from the spatial-temporal sequence. Results show that the predicted CNV is significantly consistent with the actual follow-up visit.

2 | METHODS

2.1 | Overview

To predict the effect of three injections of anti-VEGF drugs, we have proposed a complete pipeline, the process is shown in Figure 2. For each patient's follow-up images, the following four steps are taken to achieve this goal: preprocessing, longitudinal data registration, longitudinal data co-segmentation, and CNV growth prediction. A detailed description and functional interpretation of the proposed method are provided below.

2.2 | Preprocessing

Firstly, we conduct anisotropic filtering on these OCT images due to their large speckle noise. Secondly, the boundary of the retinal layers, namely the inner limiting membrane (ILM) layer and the RPE layer, are segmented to extract the volume of interest (VOI). However, some lesion areas are too large and their shadows cover the choroid, making it difficult to segment the RPE layer automatically. Therefore, the position of the RPE layer is manually adjusted with the help of OCT-Explorer,[20] a software specially used for retinal layer segmentation. The results are shown in Figure 3.

2.3 | Registration

If there is an offset or rotation between two OCT volume scans, the frame at the corresponding positions of the two OCT volumes will not correspond to the same position on the retina, which severely affects the final growth prediction accuracy. Given this situation, to guarantee the CNV growth prediction accuracy, it is necessary to solve the offset of the longitudinal data and keep the CNV area at the same position. Therefore, we need to register the follow-up OCT images to the first-time point. Considering that non-rigid registration may affect the shape and size of CNV, rigid registration is adopted in this paper. In addition, classical registration methods based on intensity similarity are sensitive to speckle noise which is commonly found in OCT images.[11] Hence, a registration net-

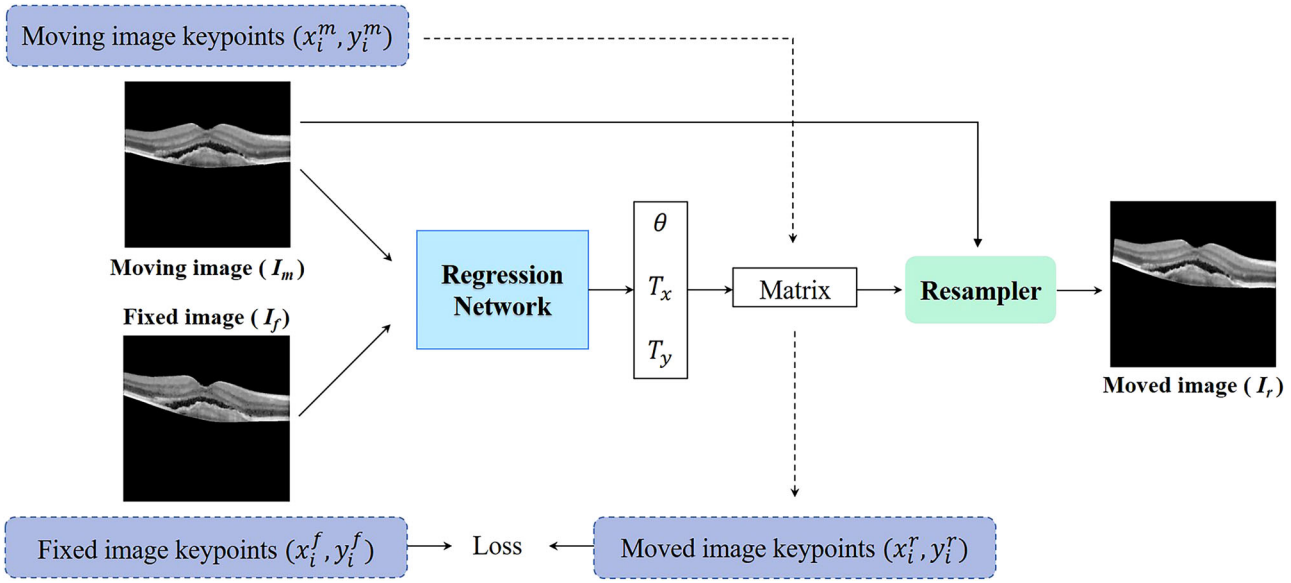


FIGURE 3 Registration structure diagram. The dotted line indicates that it is only required during the training process.

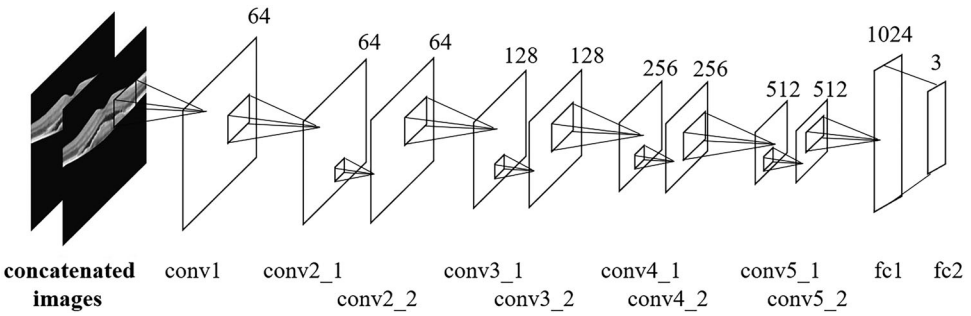


FIGURE 4 Regression network architecture.

work based on key points is proposed, as shown in Figure 3.

During training, the fixed image (I_f), the moving image (I_m), and their corresponding key points are input into a regression network. Usually, points with obvious features are selected as key points, but these key points cannot be accurately matched with pathological changes in follow-up images. Therefore, we selected 4 points on the RPE layer as the key points, and the values of abscissa are 102, 204, 306, and 408 respectively. In addition, the ordinate values can be calculated from the preprocessed images. In fact, two key points can determine the parameters of rigid registration. Here, to align a group of images to be registered on the RPE layer, two additional points on the RPE layer are taken for constraint, so as to avoid the situation that two groups of key points are aligned but the RPE layer is not aligned. The regression network uses ResNet-18[21] as its feature extractor, as shown in Figure 4.

After that, three transformation parameters are obtained: the rotation angle θ, the horizontal translation T_x and the vertical translation T_y. These three parameters constitute the rigid transformation matrix T_r:

$$T_r = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & x_0 - x_0\cos(\theta) + y_0\sin(\theta) + T_x \\ \sin(\theta) & \cos(\theta) & y_0 - x_0\sin(\theta) - y_0\cos(\theta) + T_y \end{bmatrix} \quad (1)$$

where (x_0, y_0) is the coordinate of the center point of the moving image. Transform the key points $(x_i^m,\ y_i^m)$ of the moving image according to the following formula to get the transformed key points $(x_i^r,\ y_i^r)$:

$$\begin{pmatrix} x_i^r \\ y_i^r \end{pmatrix} = T_r^{-1} \begin{pmatrix} x_i^m \\ y_i^m \\ 1 \end{pmatrix}, i = 1,\ 2,\ 3,\ 4 \quad (2)$$

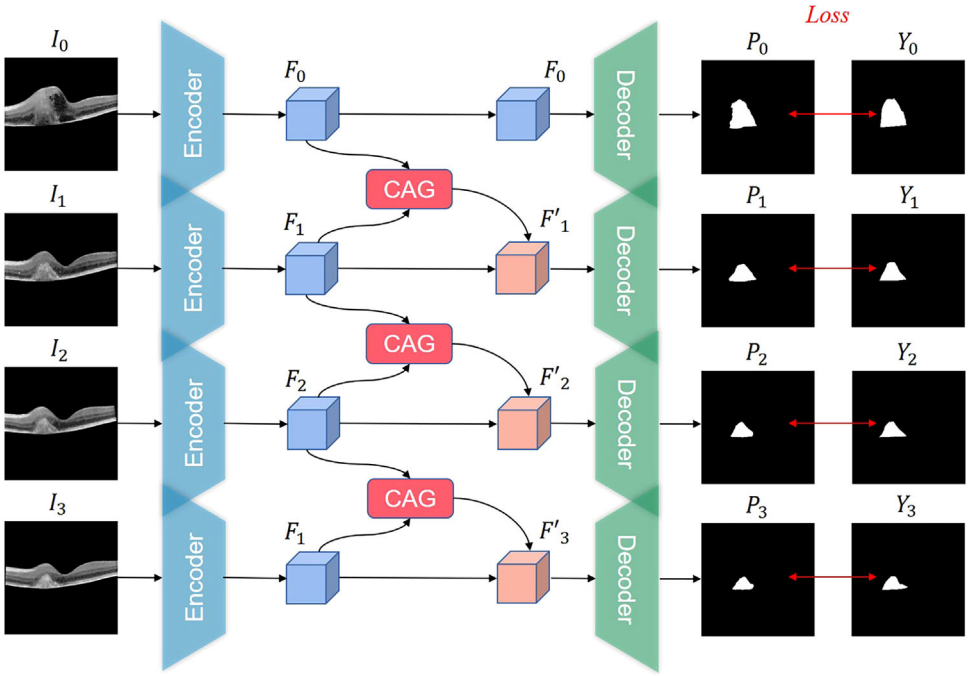


FIGURE 5 An overview of the proposed network for choroid neovascularization (CNV) co-segmentation. It contains three parts: Siamese encoder, correlation attention guidance (CAG) module, and Siamese decoder.

The loss function is defined as the root mean square error (RMSE) between $(x_i^r,\ y_i^r)$ and $(x_i^f,\ y_i^f)$:

$$Loss = \sqrt{\frac{\sum_{i=1}^{4}\left((x_i^f - x_i^r)^2 + (y_i^f - y_i^r)^2\right)}{4}} \tag{3}$$

Finally, the registered image is obtained by resampling the moving image. There is no need for key points during the test phase, only the fixed image and the moving image need to be input, and the transformation parameters can be automatically regressed to generate the registered image. Then, the registered CNV mask is acquired by implementing the same transformation on the moving image and moving mask, which prepares the input for the next step of co-segmentation.

2.4 | Co-segmentation

To study the growth process of CNV, we need to segment the registered images and extract the interested CNV lesion regions, which can be used as input for growth prediction. CNV segmentation is a very important step, because the segmentation results directly affect the accuracy of growth prediction results. Here, we designed a co-segmentation network to achieve this, as shown in Figure 5. Our model is mainly composed of three parts: Siamese encoder, correlation attention guidance (CAG) module, and Siamese decoder, where Siamese encoder and Siamese decoder are weight sharing. The former is used to extract spatial features from the input image, while the latter is adopted to generate the segmentation map from the features. In addition, the CAG module is designed to obtain the co-attention of two input images and apply it to the second feature map, as shown in Figure 6.

Firstly, we use I_0 to represent the reference image of each patient, and the three follow-up images after registration are represented by I_1, I_2, and I_3, respectively. These four images are integrated into a group as the input of the network. In the proposed method, ResNet-34[22] is adopted as the backbone of the feature encoder, which retains four feature-extracting blocks without the average pooling layer and the fully connected layer. The feature maps generated by the encoder are denoted as F_0, F_1, F_2, and F_3 and, the adjacent features among them are then input into the CAG module respectively, which is illustrated in Figure 6.

Supposing that the input feature map is $F_A \in \mathbb{R}^{C,H,W}$ and $F_B \in \mathbb{R}^{C,H,W}$, where C, H, W represent the channel numbers, height, and width of the input feature. As can be seen from Figure 6, the proposed CAG module mainly consists of six steps:

1. To reduce the dimension of weights and speed up the computations, the input frames are passed through a dimension reduction layer, which includes 1×1

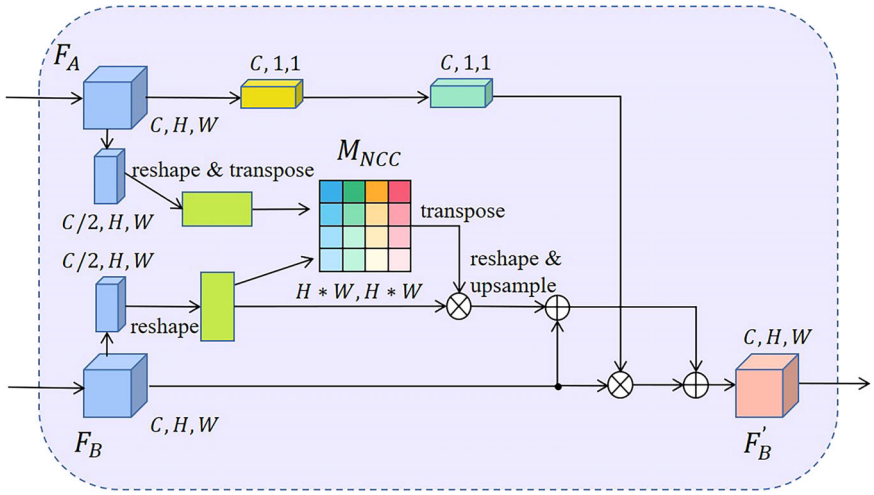


FIGURE 6 The illustration of the correlation attention guidance (CAG) module. F_A and F_B represent the pair of the input feature map. M_{NCC} is a comparison matrix, which is obtained by calculating the normalized cross correlation (NCC) of each pixel on the pair of feature maps. By performing the above operations, we can get the attended feature map F'_B.

convolution, batch normalization, and ReLU. In this way, we can get the feature maps of dimension C/2 × H × W as the output.

2. We reshape and transpose F_A to $F_{AR} \in \mathbb{R}^{H*W,C/2}$, and reshape F_B to $F_{BR} \in \mathbb{R}^{C/2,H*W}$.
3. Calculating the comparison matrix $M_{NCC} \in \mathbb{R}^{H*W,H*W}$ between F_{AR} and F_{BR} to compare the local descriptor at every spatial location (i,j). Here, the comparison is carried out using normalized cross correlation (NCC) since it is more robust than a simple correlation.[23] And then apply a softmax layer, as follows:

$$M_{NCC} = softmax\left(\frac{f_{AR}(i,j) - \overline{f_{AR}}}{\sqrt{\frac{1}{C/2}\sum_{j=0}^{C/2}\left(f_{AR}(i,j) - \overline{f_{AR}}\right)^2}} \times \frac{f_{BR}(j,i) - \overline{f_{BR}}}{\sqrt{\frac{1}{HW}\sum_{i=0}^{HW}\left(f_{BR}(j,i) - \overline{f_{BR}}\right)^2}}\right) \quad (4)$$

where $f_{AR}(i,j)$ represents the pixel value at (i,j) in feature map F_{AR}, $\overline{f_{AR}}$ represents the average pixel value of feature map F_{AR}, $f_{BR}(j,i)$ represents the pixel value at (j,i) in feature map F_{BR}, and $\overline{f_{BR}}$ represents the average pixel value of feature map F_{BR}.

4. Then, a matrix multiplication between F_{BR} and the transpose of M_{NCC} is performed to reshape and upsample the result to $\mathbb{R}^{C,H,W}$. We multiply the result by a scale parameter α and perform an element-wise sum operation with the feature map F_B to obtain the spatial attention output $E \in \mathbb{R}^{C,H,W}$, as follows:

$$E = \alpha \sum(F_{BR}M_{NCC}) + F_B \quad (5)$$

5. In the channel attention step, we apply Global Average Pooling (GAP) on the feature map F_A, and the result feature map is passed through a Multi-Layer Perceptron (MLP) followed by a sigmoid activation function to get the channel weight. By performing channel-wise multiplication with the channel weight and the feature map F_B, we can get the channel attention output $T \in \mathbb{R}^{C,H,W}$:

$$T = F_B * \sigma(MLP(GAP(F_A)) \quad (6)$$

It is worth noting that if F_A does not have high activation in channel i, F_B will not be activated in channel i since this channel activation is initially small.

6. Finally, an element-wise summation operation is done between E and T to generate the attended feature map $F'_B \in \mathbb{R}^{C,H,W}$:

$$F'_B = E + T \quad (7)$$

It can be seen from the above operations that the proposed CAG module is a channel spatial attention architecture. Although GAP can extract global information from the feature map, spatial information will be lost due to its pooling operation. To this end, we calculate the correlation of the feature maps as the spatial attention to improve the segmentation performance. Since the follow-up images to be segmented are registered, their appearance and position are related to a certain extent. Therefore, the CAG module can learn the synergetic relationships between image pairs to suppress noise information, like the irrelevant background and non-common objects.

Processed by the CAG module, we can obtain the attended feature map F'_1, F'_2, F'_3, and the origin feature map F_0, and input them to the decoder. The decoder aims to recover the high-level semantic features from

F_0, F'_1, F'_2, and F'_3, so we choose to use deconvolution as the upsampling method, which mainly consists of a 1×1 convolution, a 3×3 deconvolution and a 1×1 convolution consecutively. In addition, the skip connection is added to the network to integrate some detailed information from the encoder to the decoder. Finally, the decoder output 4 prediction masks: P_0, P_1, P_2, and P_3, with the same size as the input.

A joint loss L_{single} is adopted to measure each prediction mask P_i and its corresponding ground truth Y_i, which consists of Dice loss L_{Dice} and binary cross-entropy loss L_{BCE}:

$$L_{single} = L_{Dice} + L_{BCE} \tag{8}$$

where,

$$L_{Dice} = 1 - \frac{2|P * Y|}{|P| + |Y|} \tag{9}$$

$$L_{BCE} = -\sum_{h,w} (1 - Y) \log (1 - P) + Y log (P) \tag{10}$$

where P and Y are the segmentation results and their corresponding ground truth, h and w are the coordinates of the pixel in P and Y. Obviously, L_{Dice} and L_{BCE} are able to optimize the model at image level and pixel level, respectively. The total loss function combining the four single loss functions is as follows:

$$L_{total} = \frac{1}{4} \sum_{i=1}^{4} L_{single} (P_i, Y_i) \tag{11}$$

2.5 | Growth prediction

Based on the above registration and segmentation experiments, the growth prediction experiments are carried out in this section. Studies have shown that recurrent neural networks like long short-term memory (LSTM)[24] and Gated Recurrent Unit (GRU)[25] can deal with time series tasks. Compared with LSTM, the Convolutional LSTM (ConvLSTM)[26] is designed for spatio-temporal sequence prediction, which can perform convolution in multi-dimensional data to capture spatial features. The Convolutional GRU (ConvGRU)[27] performs similar operations on GRU. However, it has only two gates, which makes the computation less and the performance no worse than that of ConvLSTM. Its working mechanism is as follows:

$$R_t = \sigma (W_{xr} * X_t + W_{hr} * H_{t-1}) \tag{12}$$

$$Z_t = \sigma (W_{xz} * X_t + W_{hz} * H_{t-1}) \tag{13}$$

$$\tilde{H}_t = tanh (W_{xh} * X_t + R_t \circ (W_{hh} * H_{t-1})) \tag{14}$$

$$H_t = (1 - Z_t) \circ H_{t-1} + Z_t \circ \tilde{H}_t \tag{15}$$

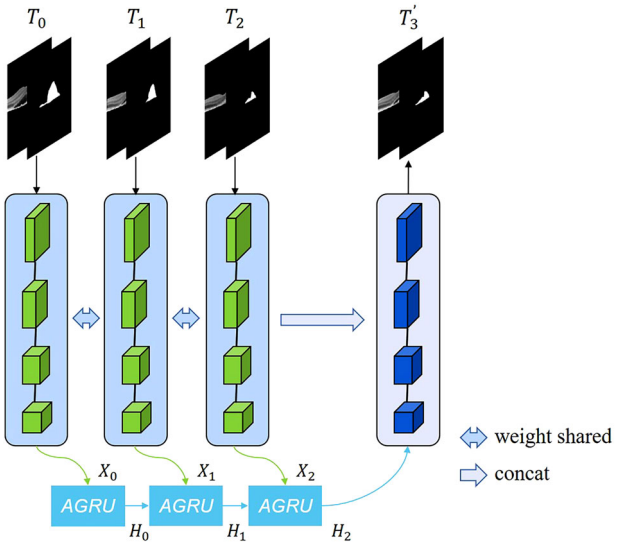


FIGURE 7 Overview of the proposed growth prediction model, where "AGRU" represent the attention-based gated recurrent unit (AGRU) module.

where X_t, R_t, Z_t, $\tilde{H}_t$, and H_t are the input tensor, reset gate, update gate, new information, and memory state, respectively. W_{**} represent the weight and the bias terms are omitted for notational simplicity. * is the convolution operation and $\circ$ is the Hadamard product. σ and *tanh* denote the sigmoid function and *tanh* function, respectively. Once a new input arrives, the reset gate will control whether to clear the previous state and the update gate will control how much the new information will be written to the state.

Inspired by this, we propose an encoder-recurrent-decoder framework to predict the future shape of CNV, which is shown in Figure 7. Firstly, we concatenate the first three time points OCT images and their segmentation results respectively denoted as T_0, T_1, and T_2. Then, we use UNet[28] as the backbone of the encoder. Specifically, on the encoder side, 4 convolution layers all have a filter size of 3×3, the stride of 1, followed by rectified linear unit (ReLU) activation function and zero padding of 1. Max-pooling has been performed over 2×2 spatial windows with stride 2. Furthermore, the obtained high dimension feature maps X_0, X_1, and X_2 are input to the attention-based gated recurrent unit (AGRU) module. This module is improved on the basis of ConvGRU, which can jointly consider the temporal and spatial characteristics of spatio-temporal sequence images while reducing the amount of computation. In particular, we use the attention mechanism to improve the update gate, so that the network can focus on the spatial-channel-wise variation of the current input. The working mechanism of the AGRU module is as follows:

$$R_n = G_{reset} (X_n, H_{n-1}) = \sigma (W_{xr} * X_n + W_{hr} * H_{n-1}) \tag{16}$$

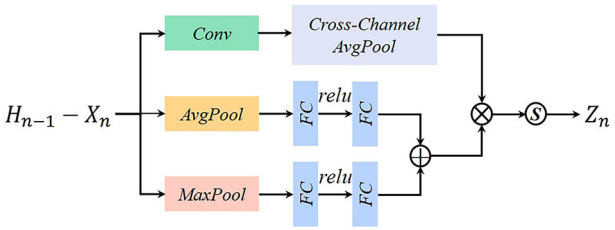


FIGURE 8 The architecture of the proposed update gate.

$$Z_n = G_{update}\ (H_{n-1} - X_n) = \sigma\left(F_{SA}\left(H_{n-1} - X_n\right) \circ F_{CA}\left(H_{n-1} - X_n\right)\right) \quad (17)$$

$$\tilde{H}_n = tanh\left(W_{xh} * X_n + R_n \circ (W_{hh} * H_{n-1})\right) \quad (18)$$

$$H_n = (1 - Z_n) \circ H_{n-1} + Z_n \circ \tilde{H}_n \quad (19)$$

where X_n, R_n, Z_n, $\tilde{H}_n$, and H_n are the input tensor, reset gate, update gate, new information, and memory state, respectively. Since the morphology of CNV is changing, we want to fully predict the areas that remain unchanged and the areas that will change with the attention mechanism to determine what previous information should be retained in H_{n-1} and what new information should be updated from $\tilde{H}_n$. So, we use $H_{n-1} - X_n$ as the input of the update gate. The architecture of the proposed update gate is illustrated in Figure 8.

As can be seen that for spatial attention, we perform 1×1 convolution and cross-channel average pooling to get the overall response in each spatial position. This operation is defined as F_{SA} in Equation (17). For channel attention, we introduce the global average pooling and global max pooling for the input, respectively. Then two fully connected layers are applied to get the channel attention maps. The summation of two-channel attention maps is defined as F_{CA} in Equation (17). The overall attention map is computed by element-wise multiplication of the spatial attention map and channel attention map. After a sigmoid operation, we can get the output of the update gate.

On the decoder side of Figure 7, there are four deconvolution layers and finally, the prediction results are achieved through the 1×1 convolution. In addition, considering that the CNV regions to be predicted are most relevant to the T_2 image, we perform the concatenation operation for the feature maps generated by T_2 in each stage and the feature maps restored by the decoder to make up for the loss of fine information caused by down-sampling.

3 | EXPERIMENTS AND RESULTS

3.1 | Dataset

In this study, there are 63 CNV subjects, and each subject has OCT follow-up images collected by Shanghai General Hospital at 4 time points (time intervals are within 30 ± 7 days [mean $\pm$ SD]). All the SD-OCT images are obtained by the Heidelberg scanner with a size of $512 \times 496 \times 19$ voxels ($12.38\ \mu m \times 3.87\ \mu m \times 261.9\ \mu m$), covering a volume of 6 mm $\times$ 2 mm $\times$ 5 mm. The third dimension is 19, which means each 3D OCT has 19 slices. Therefore, we have 4788 slices of 63 patients, and the following experiments are based on these two-dimensional images. The annotations of the CNV boundaries were all manually segmented by professional physicians. The injection of Anti-VEGF and the changes in CNV are shown in Figure 9.

3.2 | Experimental setup

3.2.1 | Parameter setting

The aforementioned models were implemented with Pytorch. All experiments were conducted on a computer with Intel Core i7-9700 CPU, GPU NVIDIA GeForce

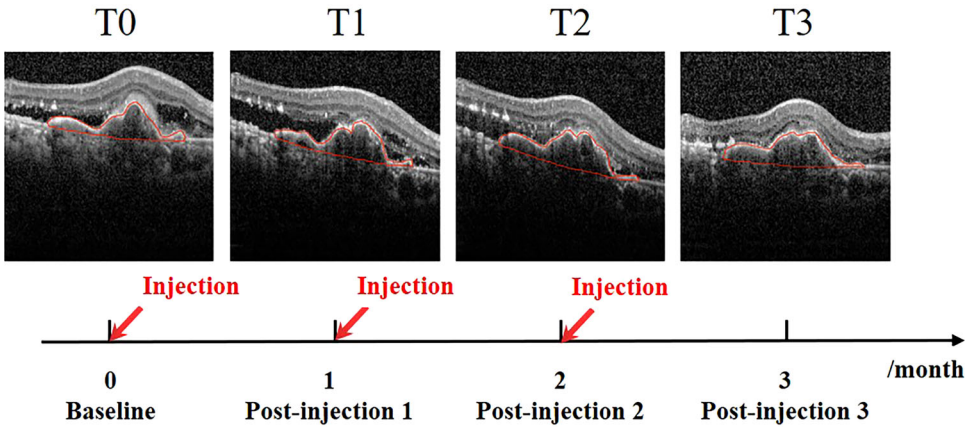


FIGURE 9 Optical coherence tomography (OCT) B-scans of the same patient at different time points. The red line is the choroid neovascularization (CNV) boundary.

RTX 2060 SUPER, and 16 GB RAM. For the registration model, 43/11/9 patients were respectively used for training/validation/testing. Adam optimizer was adopted to optimize the loss function, setting the learning rate to 0.0001, batch size to 16, and epoch to 400. For the segmentation model, a 4-fold cross-validation strategy was adopted. The first three folds include the data of 16 patients separately, and the last fold includes data of 15 patients. We trained the model for 200 iterations per-fold with a batch size of 2. For the growth prediction model, we used the same cross-validation method, but the batch size and epoch were set to 4 and 50, respectively. To prevent over-fitting and improve the robust ability of the model, we performed the online data augmentation strategy, including horizontal flipping, shifting, and rotations of −20 degrees to 20 degrees. For a fair comparison, we used the same training strategy in all experiments.

3.2.2 | Evaluation metrics

In addition to the RMSE mentioned above, we also use normalized mutual information (NMI) to measure the effect of registration, as shown below:

$$NMI\ (I_f, I_r) = \frac{H(I_f) + H(I_r)}{H(I_f, I_r)} \tag{20}$$

where I_f is the reference image, I_r is the registered image. $H(I_f)$ and $H(I_r)$ represent the information entropy of I_f and I_r, respectively. $H(I_f, I_r)$ is the joint entropy of I_f and I_r.

To comprehensively and fairly evaluate the segmentation performance of different methods, five evaluation indicators are used, including Dice similarity coefficient (Dsc), Jaccard index (Jac), Sensitivity (Sen), number of parameters, and test time, among which Dsc and Jac are the most commonly used metrics in validating the performance of segmentation algorithms.[29–31] Their definitions are as follows:

$$Dsc = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{21}$$

$$Jac = \frac{TP}{TP + TN + FN} \tag{22}$$

$$Sen = \frac{TP}{TP + FN} \tag{23}$$

For growth prediction, we also apply three additional indicators, including Precision (Pre), Specificity (Spec), and Accuracy (Acc).

$$Pre = \frac{TP}{TP + FP} \tag{24}$$

$$Spec = \frac{TN}{TN + FP} \tag{25}$$

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{26}$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

3.3 | Results

3.3.1 | Registration results

To prove the effectiveness of the proposed registration method, we compare our proposed method with other excellent registration algorithms, including a traditional image registration algorithm SimpleElastix[32] and a rigid registration network based on deep learning proposed by Sloan et al.,[33] as shown in Table 1. SimpleElastix is originally used for intensity-based registration of medical images. Here we perform a rigid registration experiment with default parameter settings, including mutual information measure, adaptive stochastic gradient descent (ASGD) optimization, and linear interpolator. Sloan et al. use real transformation parameters as supervisory information to train the full convolution regression network, of which the loss function used is the mean square error between the real transformation parameters and the predictive transformation parameters. Since the dataset used in this paper does not contain the real transformation parameters, we use two of the four key points mentioned above to calculate the rotation angle, horizontal and vertical displacement as the ground truth for experiments. In addition, compared with the unregistered images, the performance of the above methods is evaluated by RMSE, NMI, and test time. It can be observed from Table 1 that compared with the traditional SimpleElastix, our registration method has advantages in three indicators, especially in reducing the test time greatly. It is worth noting that Sloan et al.'s method that directly predicts transformation parameters has a better performance than the method based on intensity similarity. However, due to the different units of transformation parameters, simple objective functions

TABLE 1 Comparison of registration results.

METHODS	RMSE	NMI	Test time (s)
Without registration	43.182 ± 28.114	0.161 ± 0.094	–
SimpleElastix[32]	12.026 ± 22.757	0.632 ± 0.134	14.705
Sloan[33]	9.294 ± 10.242	0.650 ± 0.038	**0.079**
Proposed	**6.754 ± 5.292**	**0.666 ± 0.043**	0.083

Abbreviations: NMI, normalized mutual information; RMSE, root mean square error.
Values in bold indicate the best performance.

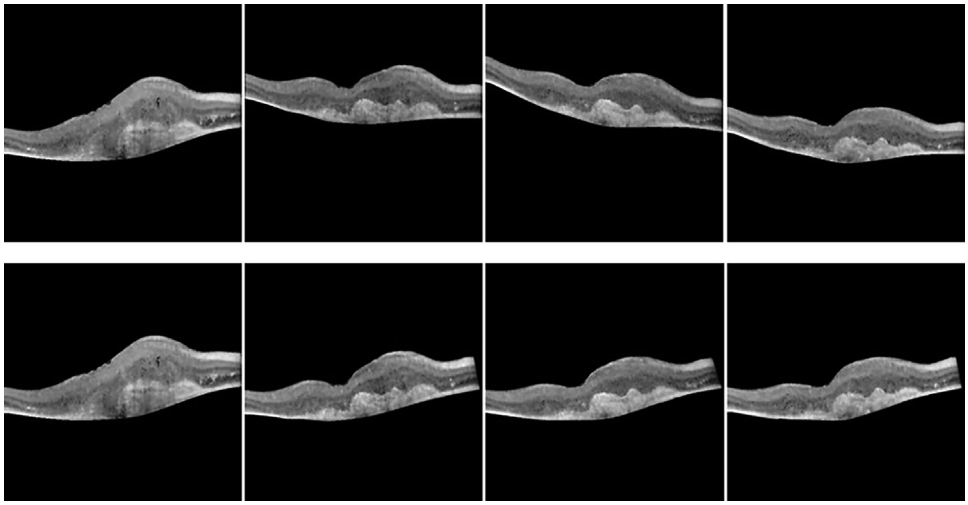


FIGURE 10 Comparison before and after registration. The first row is B-scans at four time points that are not registered, and the second row is B-scans at four time points after registration, in which the first column is used as the reference image.

cannot fully learn such differences, which makes the network difficult to converge.

Figure 10 shows the comparative B-scans of a randomly selected patient before and after registration. The B-scans in the first row are unregistered, from which can be seen that there is a large displacement deviation in different follow-up visits. The B-scans in the second row are registered, reducing the offset of the CNV area to a certain extent which is caused by different shooting positions or the patient's eyeball rotation. Based on the above analysis, this registration method can avoid the change of CNV morphology caused by non-rigid registration. This is conducive to observing the change of lesion area in the same position, and further reduces the obstacle for subsequent CNV growth prediction.

3.3.2 | Segmentation results

In this section, we compare the proposed co-segmentation network with single input segmentation methods and one common co-segmentation model. The U-shape based networks such as U-Net,[28] CE-Net,[34] CPFNet,[29] and Swin-Unet[35] are trained as the single input segmentation methods. Among them, UNet is the most classic network in the field of image segmentation. CE-Net and CPFNet proposed later are both improved on the basis of UNet and designed specifically for medical image segmentation. Swin-Unet is a U-shaped architecture based on the pure transformer, which has achieved good performance in many segmentation tasks. The transformer block can learn global and long-term semantic feature interaction well, thus avoiding the inherent limitations of the convolution operation. Table 2 gives the quantitative experimental results. Firstly, compared with single input segmentation methods, the proposed method has achieved better segmentation performance and the Dsc, Jac and Sen are improved by 0.55%, 0.61%, and 1.01% respectively, and the number of parameters is less. In addition, we also compared the proposed method with a common co-segmentation method, which is a semantic aware attention based co-segmentation model with few computing resources proposed by Chen et al.[36] As can be seen from Table 2, compared with the method of Chen et al., the proposed method gets an overall improvement in terms of all three indicators with fewer parameters. The proposed network without the CAG module is used as the baseline method. Compared with the baseline method, the performance of our method has been greatly improved and achieves 85.27% for Dsc, 80.62% for Jac, and 90.10% for Sen. In addition, it can be seen from Table 2 that the complexity and test time of our model are comparable or less than those of other methods. Here, the test time of co-segmentation refers to the total time of four images simultaneously segmented. The above experiments are completed on the basis of registration. Then, we input the unregistered follow-up images into the proposed co-segmentation network, and the segmentation performance is significantly decreased, with Dsc 83.69%, Jac 78.78%, and Sen 88.15%. The experimental results show the proposed CAG module has poor performance on unregistered images, and even plays a negative role. This may be because the position correlation of CNV in the unregistered follow-up images is not high, making the network incorrectly guide the segmentation of the next follow-up image. However, the location correlation of the registered images is improved, so the proposed CAG module can capture the location correlation and semantic correlation of a pair of images at the same time, which is conducive to guiding the segmentation of the next follow-up image.

TABLE 2 The Results of comparable experiments and ablation studies on choroid neovascularization (CNV) segmentation.

Methods	Networks	Dsc (%)	Jac (%)	Sen (%)	Parameters (M)	Time (ms)
Single input segmentation	Swin-Unet[35]	78.80 ± 0.78	73.60 ± 0.61	83.00 ± 2.33	17.9	16.4
	U-Net[28]	82.33 ± 2.12	77.34 ± 2.57	85.55 ± 2.52	31.0	**4.0**
	CE-Net[34]	84.48 ± 1.47	79.91 ± 1.44	89.59 ± 1.36	29.0	9.1
	CPFNet[29]	84.72 ± 1.14	80.01 ± 1.06	89.09 ± 1.45	43.3	12.6
Co-segmentation	Chen et al.[36]	83.55 ± 1.16	78.38 ± 1.44	88.81 ± 1.46	32.8	31.0
	Baseline	84.31 ± 0.87	79.75 ± 1.02	89.50 ± 1.98	**21.7**	29.3
	Proposed	**85.27 ± 0.91**	**80.62 ± 1.01**	**90.10 ± 1.54**	26.1	32.7

Values in bold indicate the best performance.

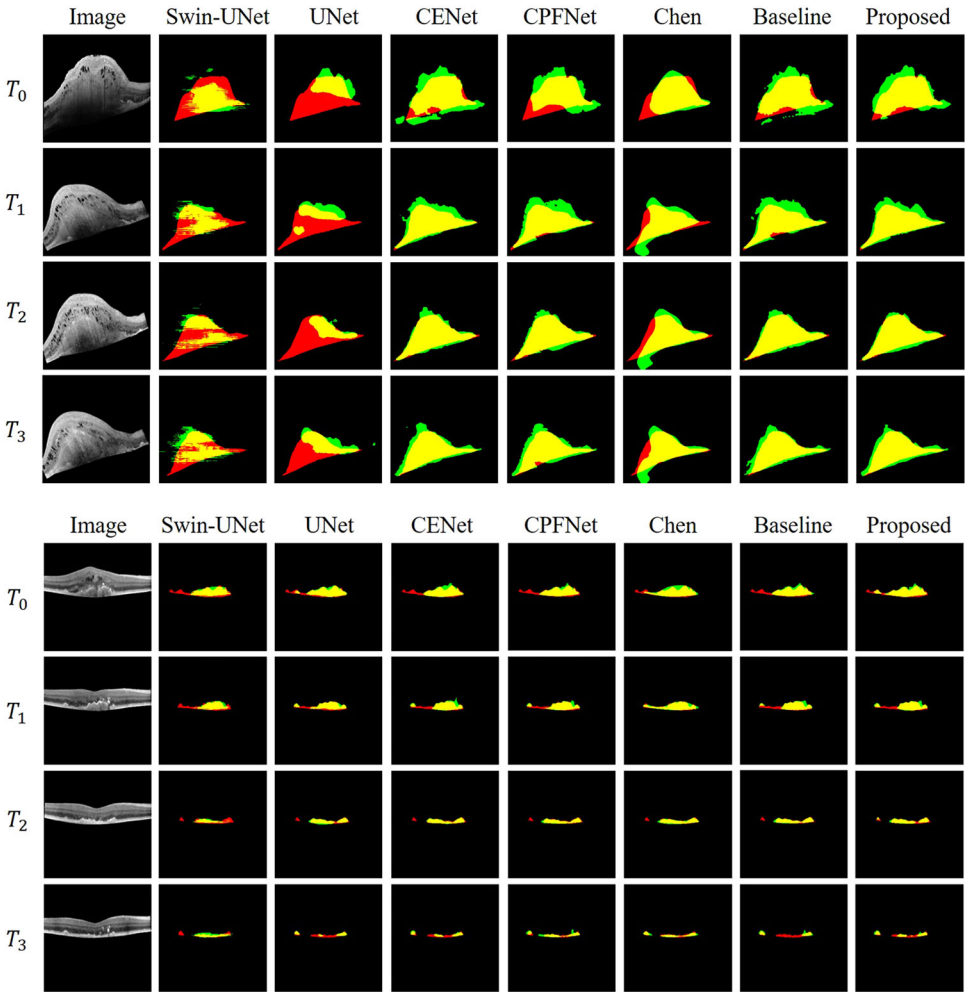


FIGURE 11 Two cases of choroid neovascularization (CNV) co-segmentation. From left to right: input image, segmentation result of Swin-Unet, UNet, CENet, CPFNet, Chen, baseline, and the proposed method. Yellow represents the correctly segmented region, while red and green are the results of false negative segmentation and false positive segmentation, respectively.

To qualitatively analyze the effectiveness of the segmentation method proposed in this paper, two cases of segmentation results of several methods are shown in Figure 11, where yellow represents the correctly segmented region while red and green are the results of false negative segmentation and false positive segmentation, respectively. Overall, Swin-Unet performs worst, probably because the transformer block lacks the ability to extract complex pathological features. In Figure 11a, the CNV has a large area and its boundary is not obvious, which poses a great obstacle to the segmentation task. Among these results, U-Net has a serious mis-segmentation problem, while the performance of CE-Net and CPFNet is better than U-Net, which may

TABLE 3 Experimental results of growth prediction with segmentation results as input.

Methods	Dsc (%)	Jac (%)	Pre (%)	Spe (%)	Acc (%)
TAPred[37]	79.69 ± 1.56	74.75 ± 1.85	86.06 ± 3.11	99.68 ± 0.12	99.46 ± 0.15
Baseline	82.19 ± 2.66	77.22 ± 2.97	83.49 ± 4.33	99.63 ± 0.16	99.51 ± 0.20
Baseline+CLSTM	83.17 ± 2.51	78.24 ± 2.69	84.14 ± 4.54	99.65 ± 0.19	99.51 ± 0.23
Baseline+CGRU	83.38 ± 2.51	78.75 ± 2.76	**87.10 ± 5.14**	99.73 ± 0.16	**99.57 ± 0.18**
Baseline+AGRU (proposed)	**83.69 ± 2.50**	**78.85 ± 2.55**	86.91 ± 4.75	**99.73 ± 0.15**	99.54 ± 0.21

benefit from the fused context information. However, there are still some false negative segmentation and false positive segmentation. The results of Chen et al. indicate that only channel-wise attention may underestimate the correlation between the two images when there are large changes in them, reducing the channel weights and causing some mis-segmentation problems. It is worth noting that with the help of the CAG mechanism, our method can guide the extraction of features at the next time point according to the existing features, thus reducing part of the wrong segmentation. Moreover, the proposed method also works well in the case of small CNV regions, which can be seen in Figure 11b.

3.3.3 | Growth prediction results

To achieve the growth prediction of CNV, the segmentation results obtained by our co-segmentation network are input into the growth prediction network. In this part, we conduct a series of comparative experiments and ablation experiments to verify the effectiveness of the proposed method. The quantitative results of different methods for the growth prediction of CNV are shown in Table 3. TAPred was proposed by Zhang et al.[37] for automated prediction of geographic atrophy (GA) lesion growth. Here we adapt their model to fit the CNV growth prediction task. In addition, we set the Baseline method by using concatenate operations instead of recurrent neural networks in this study. Baseline+CLSTM, Baseline+CGRU, and Baseline+AGRU represent ConvLSTM, ConvGRU and AGRU used as recurrent neural networks, respectively. As can be seen from Table 3, with the high accuracy of the above registration and segmentation network, the growth prediction experiments have basically achieved a considerable effect, especially our method has made an overall improvement in terms of most metrics. Firstly, compared with TAPred method, the proposed method improves the Dsc, Jac, Pre, Spre and Acc by 4.00%, 4.10%, 0.85%, 0.05%, and 0.08% respectively, and the main indicators Dsc and Jac reach 83.69% and 78.85%, which indicates that our prediction results are closer to the actual follow-up. In addition, compared with the Baseline, the existence of recurrent neural networks is crucial for the growth prediction task which can capture the dynamic changes of the input. And among these recurrent neural networks, our AGRU network has better performance due to the integration of spatial attention and channel attention of longitudinal data. Similarly, when the unregistered images are inputted into the growth prediction network, the resulting Dsc, Jac, Pre, Spe and Acc are 76.43%, 71.25%, 82.28%, 99.67%, and 99.38%. This proves the effectiveness of the previous registration work, which can enable the prediction network to better capture the dynamic changes of pixels at the same location.

Figure 12 shows two examples of CNV growth prediction results. The first row is the patient's follow-up OCT B-scans and the second row is the input segmentation results, followed by the prediction results. It can be observed from Figure 12a and b that compared with TAPred method, our model is able to represent subtle reduction and strong growth in size well. The predicted results are closely related to the segmentation results, and the results of our proposed method are largely consistent with the ground truth, which further shows the effectiveness of the proposed method.

4 | CONCLUSION AND DISCUSSION

In this paper, we propose a CNV growth prediction method based on deep learning, including registration, co-segmentation, and growth prediction. Firstly, the registration model is trained by matching key points to register the follow-up image to the reference image. Then, the co-segmentation network embedded with the CAG module collaboratively segments the longitudinal SD-OCT images. Finally, the growth prediction model integrating CNN and AGRU modules is designed, which can extract the features of each OCT B-Scan and learn the spatio-temporal information of the longitudinal data, respectively. To the best of our knowledge, it is the first time to develop a CNN-based method for CNV growth prediction. The experimental results show that our prediction results of CNV are largely consistent with the ground truth and may become a potential technology to provide useful prognostic information for clinicians and patients. In addition, we calculate the efficiency of the three networks as a whole. The input is four 3D OCT follow-up images of a patient, that is, each 3D OCT has 19 slices. After the preprocessing of these

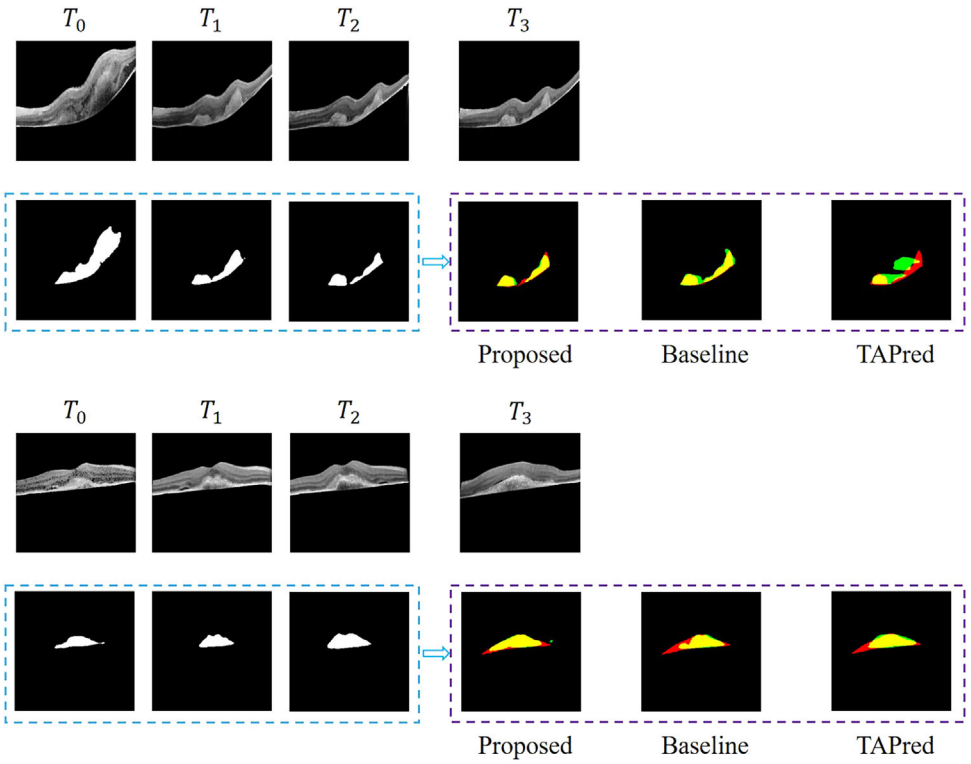


FIGURE 12 Two examples of choroid neovascularization (CNV) growth prediction results. The first row is the patient's follow-up optical coherence tomography (OCT) B-scans. The blue box in the second row is the input CNV segmentation results, followed by the prediction results of different methods in the purple box. In the prediction results, yellow represents the correctly predicted region, while red and green are the results of false negative prediction and false positive prediction, respectively.

four images, the time required from registration to segmentation to obtaining growth prediction results is about 18.7 s. Note that this time is based on OCT volume, and also includes data reading and storage. The proposed method is highly efficient and can meet clinical needs.

Although the CNV morphology of T_2 and T_3 is similar, their Dsc and Jac are 80.44% and 75.51% respectively after registration, so it is valuable to carry out the growth prediction task. However, there is a question about whether a single follow-up visit can provide enough prior information to predict subsequent follow-up visits. To answer this question, we do not use T_0, T_1, and T_2 as the input of growth prediction, but only use T_2 to predict T_3, and the resulting Dsc is 82.86% and Jac is 77.86%. It can be seen that the recursive network combines the growth trend of CNV at the previous two time points to provide a reliable reference for the subsequent prediction. Besides, when the manually labeled masks are used as the input of the growth prediction model instead of the segmentation results, the prediction results are more similar to the ground truth, with the Dsc of 92.32% and the Jac of 88.08%, which shows that if the segmentation performance can be improved, our growth prediction effect will be better and closer to reality. According to the predicted CNV morphology, combined with some clinical indicators, the attending physician can judge the efficacy of the drug and decide whether to continue the injection next month or change the treatment plan.

In the future, we will explore strategies to improve the accuracy of segmentation and expand our prediction model to 3D. Furthermore, ophthalmologists are more concerned about quantitative indicators of patient prognosis, such as best corrected visual acuity (BCVA) and central retina thickness (CRT). Therefore, if these values can be predicted directly, it will be of great help to ophthalmologists and patients, so we leave this for our future work.

ACKNOWLEDGMENTS

This study was supported in part by the National Key R&D Program of China (2016YFC0904800, 2018YFA0701700, 2019YFC0840607), in part by the National Nature Science Foundation of China (U20A20170, 62271337), in part by Natural Science Research of Jiangsu Higher Education Institutions of China (21KJB510021), and in part by Science and Technology Commission of Shanghai Municipality (201409006700).

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The dataset underlying the results presented in this paper is not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. Ohno-Matsui K, Lai Y, Lai C, Cheung C. Updates of pathologic myopia. *Prog Retin Eye Res*. 2016;52(1):156-187.
2. Klein R, Klein B, Jensen S, Meuer S. The five-year incidence and progression of age-related maculopathy: the beaver dam eye study. *Ophthalmology*. 1997;104(1):7-21.
3. Voykov B, Ziemssen F. Myopic CNV. *Klinische Monatsblatter fur Augenheilkunde*. 2011;228(9):762-770.
4. Wang E, Chen Y. Intravitreal anti-vascular endothelial growth factor for choroidal neovascularization secondary to pathologic myopia: systematic review and meta-analysis. *Retina (Philadelphia, Pa)*. 2013;33(7):1375-1392.
5. Kubicka-Trząska A, Wilańska J, Romanowska-Dixon B, Sanak M. Circulating antiretinal antibodies predict the outcome of anti-VEGF therapy in patients with exudative age-related macular degeneration. *Acta Ophthalmol (Copenh)*. 2012;90(1):e21-e24.
6. Mathew R, Pefkianaki M, Kopsachilis N, Brar M, Richardson M, Sivaprasad S. Correlation of fundus fluorescein angiography and spectral-domain optical coherence tomography in identification of membrane subtypes in neovascular age-related macular degeneration. *Ophthalmologica*. 2014;231(3):153-159.
7. Michalewski J, Nawrocki J, Trebinska M, Michalewska Z. Spectral-domain optical coherence tomography features preceding new-onset neovascular membrane formation. *Can J Ophthalmol*. 2014;49(4):339-344.
8. Niemeijer M, Garvin M, Lee K, Ginneken B, Abràmoff M, Sonka M. Registration of 3D spectral OCT volumes using 3D SIFT feature point matching. In Proc. SPIE 7259, Medical Imaging 2009: Image Processing.
9. Xu J. Alignment of 3-D optical coherence tomography scans to correct eye movement using a particle filtering. *IEEE Trans Med Imaging*. 2012;31(7):1337-1345.
10. Ricco S, Chen M, Ishikaw H, Wollstein G, Achuman J. Correcting motion artifacts in retinal spectral domain optical coherence tomography via image registration. International Conference on Medical image computing and computer-assisted intervention. Springer; 2009:100-107.
11. Pan L, Shi F, Xiang D, et al. OCTRexpert: a feature-based 3D registration method for retinal OCT images. *IEEE Trans Image Process*. 2020;29(1):3885-3897.
12. Xi X, Meng X, Yang L, et al. Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior. *Multimedia Syst*. 2019;25(1):95-102.
13. Fahmy S, Abdelmoula M, Mahfouz E, Shah M. Segmentation of choroidal neovascularization lesions in fluorescein angiograms using parametric modeling of the intensity variation. Proceedings of the IEEE International Symposium on Biomedical Imaging: from Nano to Macro. 2011:665-668.
14. Xue J, Yan S, Wang Y. Unsupervised segmentation of choroidal neovascularization for optical coherence tomography angiography by grid tissue-like membrane systems. *IEEE Access*. 2019;7(1):143058-143066.
15. Xiang D, Tian H, Yang X, et al. Automatic segmentation of retinal layer in OCT images with choroidal neovascularization. *IEEE Trans Image Process*. 2018;27(12):5880-5891.
16. Yang X, Chen X, Xiang D. Attention-guided channel to pixel convolution network for retinal layer segmentation with choroidal neovascularization. Proc. SPIE 11313, Medical Imaging 2020: Image Processing.
17. Banerjee I, Sisternes D, Hallak A, et al. Prediction of age-related macular degeneration disease using a sequential deep learning approach on longitudinal SD-OCT imaging biomarkers. *Sci Rep*. 2020;10(1):15434.
18. Zuo C, Shi F, Zhu W, Chen H, Chen X. 3D choroid neovascularization growth prediction with combined hyperelastic biomechanical model and reaction-diffusion model. *Fetal Infant Ophthalmic Med Image Anal*. 2017;10554(1):142-149.
19. Zhang Y, Ma X, Li M, Ji Z, Yuan S, Chen Q. LamNet: a lesion attention maps-guided network for the prediction of choroidal neovascularization volume in SD-OCT images. *IEEE J Biomed Health Inf*. 2022;26(4):1660-1671.
20. Kafieh R, Rabbani H, Abramoff M, Sonka M. Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. *Med Image Anal*. 2013;17(8):907-928.
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
22. Apostolopoulos S, Zanet S, Ciller C, Wolf S, Aznitman R, Pathological OCT retinal layer segmentation using branch residual u-shape networks. International Conference on Medical image computing and computer-assisted intervention. Springer; 2017:294-301.
23. Subramaniam A, Chatterjee M, Mittal A. Deep neural networks with inexact matching for person re-identification. Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016:2675-2683.
24. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780.
25. Cho K, Merrienboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014:1724-1734.
26. Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Proceedings of the 29th International Conference on Neural Information Processing Systems. 2015:802-810.
27. Shi X, Gao Z, Lausen L, Wang H, Yeung D. Deep learning for precipitation nowcasting: a benchmark and a new model. Proceedings of the 31th International Conference on Neural Information Processing Systems. 2017:5617-5627.
28. Ronneberger O, Fischer P, Brox T, U-Net: convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer; 2015:234-241.
29. Feng S, Zhao H, Shi F, et al. CPFNet: context pyramid fusion network for medical image segmentation. *IEEE Trans Med Imaging*. 2020;39(10):3008-3018.
30. Crum W, Camara O, Hill D. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging*. 2006;25(1):1451-1461.
31. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:6230-6239.
32. Marstal K, Berendsen F, Staring M, Klein S. SimpleElastix: a user-friendly, multi-lingual library for medical image registration. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016:574-582.

33. Sloan JM, Goatman KA, Siebert JP. Learning rigid image registration. Proceedings of the the 11th International Joint Conference on Biomedical Engineering Systems and Technologies; 2018:89-99.
34. Gu Z, Cheng J, Fu H, et al. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imaging*. 2019;38(10):2281-2292.
35. Cao H, Wang Y, Chen J, et al. Swin-Unet: unet-like pure transformer for medical image segmentation. European Conference on Computer Vision; 2022:205-218.
36. Chen H, Huang Y, Nakayama H. Semantic aware attention based deep object co-segmentation. Asian Conference on Computer; 2018:435-450.
37. Zhang Y, Zhang X, Ji Z, et al. An integrated time adaptive geographic atrophy prediction model for SD-OCT images. *Med Image Anal*. 2021;68(1):101893.

How to cite this article: Shen J, Chen Z, Peng Y, et al. Morphological prognosis prediction of choroid neovascularization from longitudinal SD-OCT images. *Med Phys*. 2023;1-15. https://doi.org/10.1002/mp.16294