Prediction of 3D Strain Field in Posterior Sclera Based on Attention 3D U-Net

Xubo Zhang[1,*], Yusong Wang[1,*], Siyuan Shao[1], Chongbao Zhou[1], Mingming Gong[2]
[1]School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, 450001, China
[2]iFLYTEK Co., Ltd., Hefei, Anhui, 230088, China
**Corresponding Author.*

Abstract: To analyze the biomechanical properties of the posterior sclera, we denoised OCT images and segmented them using the nnU-Net model. After obtaining the segmented tissues, we performed registration and calculated the deformation of the posterior eye tissues using mathematical methods. Finally, we combined the baseline images and the force-loaded images and trained an Attention 3D U-Net model to predict the deformation of the posterior sclera, i.e., to predict its 3D strain field. The model performance was evaluated using the mean absolute error (MAE), which was below 0.2.

Keywords: Posterior Sclera; 3D Strain Field; Attention Mechanism; 3D U-Net; Deep Learning

1. Introduction

Myopia, as the most common refractive error, has become a significant global public health issue. Studies have shown that the number of myopic individuals was approximately 1.406 billion in 2000 and is projected to reach 4.758 billion by 2050 [1].

In 2010, the leading causes of blindness were cataracts, uncorrected refractive errors, and age-related macular degeneration. Although the number of people blinded by cataracts and trachoma has decreased, the number of people with uncorrected refractive errors has increased [2]. Genetic factors are often associated with earlier onset of myopia. Studies have shown that the prevalence of myopia is highest when both parents are myopic and lowest when neither parent is myopic.

Currently, there is a lack of direct measurement methods for the biomechanical properties of the posterior sclera in clinical settings. With the development of deep learning, deep learning models can automatically learn complex features from data, offering an advantage over traditional formula-based methods. Therefore, this study aims to address the challenges in AI-based myopia risk prediction using multimodal imaging by investigating the deformation of the posterior sclera, i.e., the 3D strain field, and proposing an Attention 3D U-Net model.

2. Biomechanical Properties of the Sclera

Axial elongation, a hallmark of refractive errors, is closely related to the biomechanical properties of the posterior sclera. Therefore, understanding the changes in the biomechanical properties of the posterior sclera is crucial for early myopia prediction. The sclera is a complex connective tissue with elasticity and provides a sturdy and stable base for the retina, protecting the delicate internal structures of the eye. The sclera can resist deformation caused by changes in the retina or lens-iris diaphragm due to its regionally specialized connective tissues that endow it with unique biomechanical properties [3].

As a biological soft tissue, the biomechanical properties of the sclera primarily include general mechanical properties such as stiffness and strength, as well as soft tissue-specific properties like elasticity, viscoelasticity, and anisotropy [4].

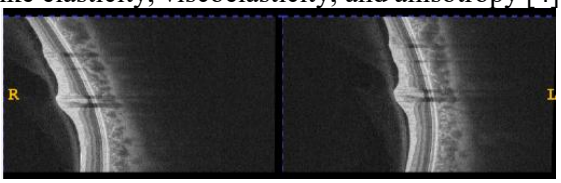

Figure 1. Comparison between Baseline and Force-loaded OCT Images of the Right Eye

Figure 1 presents a comparison between the baseline and force-loaded OCT images of the posterior sclera. By processing and analyzing these two images, the deformation of the eyeball can be obtained.

3. Calculation of the 3D Strain Field

In the data processing stage, the OCT images of the patient's fundus were denoised, and then the deformed tissue was extracted through manual segmentation. A large amount of manually segmented data was organized and trained using the nnU-Net model to achieve automatic segmentation. For the segmented baseline and force-loaded images, mathematical formulas were used to calculate the tissue deformation and obtain the 3D strain field.

To obtain the deformation between the baseline and force-loaded images, registration was performed between the two images, and the displacement field was obtained from the registration process.

The displacement field is typically represented as a vector that indicates the direction and distance of movement of each point in an object at a given time. By calculating the displacement gradient, as shown in Equation (1), the deformation of the posterior eye tissue can be reflected. The strain tensor can be calculated using Equations (2) and (3) based on the displacement gradient. The 3D strain field, including effective strain, first principal strain, and third principal strain, can be further calculated using Equation (5). The specific formulas are as follows:

$$\nabla u = \frac{\partial u_i}{\partial x_j} \tag{1}$$

$$\varepsilon_{ii} = \frac{\partial u_i}{\partial i} \tag{2}$$

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i}\right) \tag{3}$$

$$\varepsilon_{eff} = \sqrt{\frac{2}{3}\left(\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{zz}^2 + 2\left(\varepsilon_{xy}^2 + \varepsilon_{xz}^2 + \varepsilon_{yz}^2\right)\right)} \tag{4}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{xx} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{bmatrix} \tag{5}$$

In the formula: ∂u_i - displacement field; ∇u - displacement gradient; ε_{ii}- principal strain in the i direction; ε_{ij} - principal strain in the i and j directions; ε_{eff}- effective variable field; ε is the strain tensor matrix.

The registration process is illustrated in Figure 2. It was implemented using relevant methods from SimpleITK, a Python package for processing common medical image formats. Due to the large size and significant blank areas in the images, cropping and padding were applied to the data before registration. Cropping involved obtaining the coordinates of the non-zero regions in both the baseline and force-loaded images. Then, slices of the non-zero regions were generated using the top-left and bottom-right coordinates, resulting in cropped images. Subsequently, the smaller images were padded to the same size as the larger image, using zero padding centered around the cropped image.

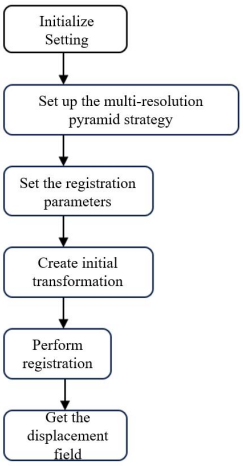

Figure 2. Registration Flow Chart

A multi-resolution pyramid strategy is an effective approach for optimizing image registration. By processing images at different levels, it can improve both the efficiency and accuracy of registration.

When defining the scaling factor of the pyramid, different scaling ratios can be set for different levels. This allows registration to start at a low resolution, accelerating the process and helping the algorithm quickly find a rough initial match. By defining different smoothing parameters for each pyramid level, more smoothing can be applied to low-resolution images, effectively reducing noise interference and enhancing the stability of the registration algorithm. The smoothing parameter is specified in physical units, allowing for more precise control of the smoothing level and ensuring consistency and comparability of the smoothing effect, especially when dealing with images with different spatial resolutions.

A conventional gradient descent method with a fixed step size is used for optimization. This ensures that the optimization algorithm has sufficient opportunities to find the optimal solution. The similarity metric for registration is the mean squared error (MSE), which evaluates the similarity between two images by calculating the sum of the squared differences between corresponding pixel values. Additionally, linear interpolation is used as the interpolation method for resampling pixel values during image transformation and registration, ensuring the smoothness and accuracy of the resampled

results.

To initialize the registration based on the geometric centers of the two images, an initial transformation is created using an affine transformation. This transformation includes translation, scaling, rotation, and shearing, and aims to provide a reasonable starting point for the registration algorithm, helping to improve convergence speed and accuracy.

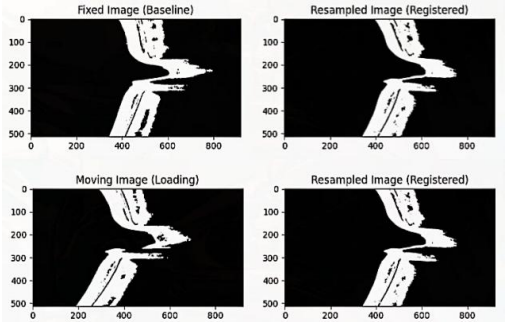

Figure 3. Before and after Registration

Figure 3 shows the comparison results of the baseline image, force-loaded image, and the registered image, indicating a good registration effect. Through the above registration process, the displacement field of the posterior eye tissue can be obtained, which allows for the calculation of the 3D strain field.

4. Attention3D U-Net Model

4.1 Introduction to Model

Since OCT images of the posterior fundus are 3D images, the U-Net model, commonly used for biomedical image segmentation, was adopted for processing. As a deep learning model, U-Net has a symmetrical encoder-decoder structure. The encoder is responsible for extracting image features, while the decoder gradually restores the spatial dimension through upsampling layers and convolutional layers. Skip connections are used between the encoder and decoder, directly connecting feature maps with the same resolution to help restore spatial information and enhance details [5]. The basic structure is shown in Figure 4.

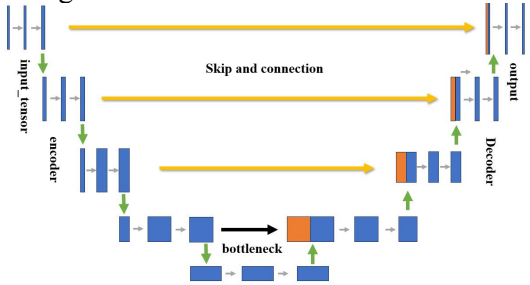

Figure 4. U-Net Structure Diagram

Traditional U-Net models perform poorly on complex OCT images, often resulting in edge losses or failing to correctly focus on tissue lesions [6]. To focus on more important features during processing, an attention mechanism was introduced during model training. By dynamically adjusting the weights of feature maps, the model can more effectively distinguish which features are more critical for a specific task [7], thus enhancing the model's robustness in handling complex medical images. For example, when encountering images with high background noise, the model can ignore unnecessary information, thereby improving segmentation accuracy.

The Attention3D U-Net model can better predict the 3D strain field of the original image, significantly reducing the time required to obtain the deformation of the eye tissue. By introducing the attention mechanism, the number of unnecessary parameters in the model can be reduced, which allows the model to maintain high performance while reducing computational resource requirements, facilitating deployment in resource-constrained environments [8].

4.2 Model Improvements

The attention mechanism is integrated into the 3D U-Net at the decoder part. During the upsampling process, the attention mechanism is applied to fuse features with the corresponding layers of the encoder. This design allows the decoder to not only rely on the feature maps provided by the encoder but also enhance important features through attention-weighted fusion, thereby better restoring image details [9].

The 3D U-Net part of the Attention3D U-Net model uses multiple convolutional blocks and upsampling blocks. Compared to the traditional U-Net, there are different numbers of convolutional layers in each encoder and decoder module. This flexibility allows the model to extract richer features at specific levels. It allows the number of convolutional layers in each convolutional block to be adjustable, enabling the network to learn at a deeper level during feature extraction, thus better capturing complex spatial features. The upsampling layer uses ConvTranspose3d, which not only improves spatial resolution but also introduces more non-linear transformations during the upsampling process, enhancing the model's expressive power.

In the Attention3D U-Net model, dynamic

padding is used to ensure the consistency of feature map sizes, improving the model's flexibility. To better control the spatial resolution of the output, interpolation is used, followed by output size adjustment to ensure that the model output meets specific requirements. In terms of model optimization, the ReLU function is used as the activation function with the parameter inplace set to True, thereby improving computational efficiency.

Figure 5 shows the architecture of the main part of the Attention3D U-Net model.

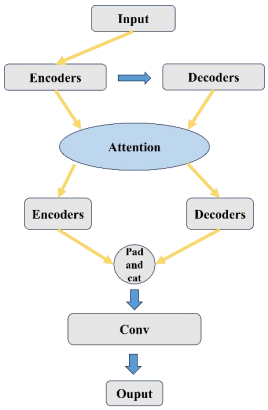

Figure 5. Partial Model Architecture Diagram of Attention 3D U-Net

5. Prediction Results and Evaluation

The training set of the model consists of 3D strain fields of the posterior sclera, including three types of strain. Therefore, the Attention3D U-Net model was trained separately for effective strain, first principal strain, and third principal strain, enabling the prediction of three different strain fields.

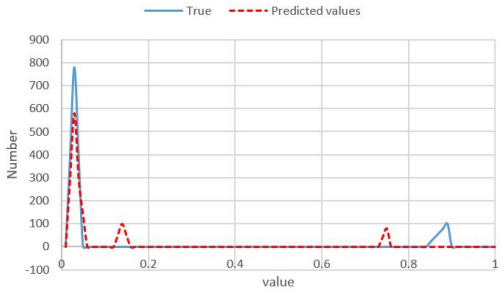

Figure 6. Histogram of the Numerical Distribution of the True and Predicted Values

Figure 6 shows the histograms of the numerical distribution of the true and predicted values. The left side shows the distribution of the true values, and the right side shows the distribution of the test values. It can be seen that most of the numerical distribution of the test values is correct, with a small amount of error.

To evaluate the calculation results and the efficiency of the model, we used Mean Absolute Error (MAE), which represents the average of the absolute differences between the predicted values and the test values. The smaller the value of MAE, the better the prediction performance of the model. The formula for calculating MAE is shown in Equation (6).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (6)$$

Where, N - the total number of data samples, y_i- the actual value of the i-th sample, $\hat{y}_i$ - the predicted value of the i-th sample.

Table 1. Loss Value and MAE of Model Training

Index	Effective strain	First principal strain	Third principal strain
LOSS	0.117	0.0046	0.0069
MAE	0.2058	0.1384	0.1653

As shown in Table 1, the model's prediction performance on the 3D strain field is presented. The first principal strain shows the best performance, while the effective strain shows relatively poorer performance. The analysis indicates that insufficient training data is the primary reason. If a larger dataset is used for model training, the prediction performance of the model can be effectively improved.

6. Conclusion

This paper innovatively applies the Attention3D U-Net model to predict the biomechanical properties of the posterior sclera of the eyeball, specifically focusing on the prediction of its 3D strain field. By employing traditional mathematical formulas to calculate the baseline image and the mechanically loaded image, effective strain, first principal strain, and third principal strain are obtained. Subsequently, the 3D strain field is used as training data for the model, combined with the baseline image and the mechanically loaded image. This enables the prediction of 3D strain fields through image uploading.

By combining traditional methods with deep learning, we can more effectively address current medical challenges. This will significantly contribute to the prediction of myopia risk, saving both time and costs. Moreover, it provides a new perspective and direction for future research on ocular tissues.

With the increase in data and the optimization of the model, it is believed that in the near future, the prediction of the biomechanical properties of

the posterior sclera will become more accurate.

References

[1] Holden B A, Fricke T R, Wilson D A, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology, 2016, 123(5): 1036-1042.

[2] Bourne R R A, Stevens G A, White R A, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. The lancet global health, 2013, 1(6): e339-e349.

[3] Boote C, Sigal I A, Grytz R, et al. Scleral structure and biomechanics. Progress in retinal and eye research, 2020, 74: 100773.

[4] Xing Fei, Ma Jianxiong, Ma Xinlong. Research Progress on the Biomechanical Properties of Soft Tissue. Chinese Journal of Orthopaedics, 2017, 37(22): 1432-1440.

[5] Huang Xiaoming, He Fuyun, Tang Xiaohu, et al. A Review of U-Net and Its Variants in Medical Image Segmentation. Chinese Journal of Biomedical Engineering, 2022, 41(05): 567-576.

[6] Yin Xiaohang, Wang Yongcai, Li Deying. A Review of Medical Image Segmentation Techniques Based on Improved U-Net Structure. Journal of Software, 2021, 32(02): 519-550.

[7] Zhou Tao, Dong Yali, Liu Shan, et al. Cross-Modal Multi-Encoded Hybrid Attention U-Net for Lung Tumor Image Segmentation. ACTA PHOTONICA SINICA, 2022, 51(04): 376-392.

[8] Zhang Y, Chen J, Ma X, et al. Interactive medical image annotation using improved Attention U-net with compound geodesic distance. Expert systems with applications, 2024, 237: 121282

[9] Yang Q, Chen J. An intelligent model for benign and malignant pulmonary nodule analysis using U-Net networks and multilevel attention mechanisms. Journal of Mechanics in Medicine and Biology, 2024: 2440032.

