Received: 13 March 2025 | Revised: 2 July 2025 | Accepted: 25 July 2025 | Published online: 13 August 2025

RESEARCH ARTICLE

Medinformatics
2025, Vol. 00(00) 1–12
DOI: 10.47852/bonviewMEDIN52025662

BON VIEW PUBLISHING

Deep Learning for Joint Scleral and Scleral Vessel Segmentation: A Comparative Study

Yongbin Qi[1], Baochen Zhen[1], Jiaming Wang[1], Shilin Zhao[1], Yansuo Yu[1,*] and Qiang Liu[1]

[1]*Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, China*

Abstract: Scleral segmentation and scleral vessel segmentation are increasingly recognized as critical components in medical image analysis, with broad applications in ocular disease diagnosis and biometric identification. In particular, scleral vessel segmentation contributes significantly to the early detection of conditions such as diabetic retinopathy and glaucoma. However, the intricate structure of scleral vessels and the scarcity of high-quality annotated datasets continue to present major challenges. To address these issues, an ensemble deep learning-based approach is proposed, integrating three segmentation models—UNet, NestNet, and DeepLabv3_Plus—to evaluate and quantitatively analyze the unified task of scleral segmentation and scleral vessel segmentation. The input ocular images undergo preprocessing steps including denoising, contrast-limited adaptive histogram equalization, and cropping. Scleral regions are extracted to enhance the explicit representation of vascular structures. Experiments are conducted on the publicly available SBVPI dataset. The DeepLabv3_Plus model achieves the highest performance in scleral segmentation, with an accuracy of 0.9656, sensitivity of 0.9694, specificity of 0.9421, and a Dice coefficient of 0.9723. For scleral vessel segmentation, the same model achieves an accuracy of 0.9285 and a specificity of 0.9536. These results demonstrate the effectiveness of the proposed ensemble framework and highlight its potential for advancing scleral vessel segmentation research. Future work will focus on further model optimization and the exploration of clinical and real-world deployment scenarios.

Keywords: scleral vessel segmentation, deep neural network, UNet, NestNet, Deeplabv3_Plus

1. Introduction

Scleral segmentation [1] and scleral vessel segmentation [2] are important tasks in the field of medical image analysis. They play an important role in the diagnosis of eye diseases and in biometrics. Diabetic retinopathy (DR) is a common complication of diabetes and is one of the leading causes of blindness among the elderly worldwide [3]. The risk of blindness can be effectively reduced through early detection and active treatment. According to a report by the World Health Organization [4], early detection and treatment of eye diseases are crucial strategies for reducing the global burden, saving healthcare costs, and improving quality of life. Glaucoma, another major cause of blindness, is characterized by damage to the optic nerve, and changes in the scleral vessels are closely related to the health of the optic nerve [5]. By analyzing the morphology of scleral vessels, doctors can identify early signs of DR and glaucoma. These changes may include dilated, tortuous, or neovascularized blood vessels. Morphological characteristics of scleral vessels, such as vessel length, width, curvature, branching pattern, and angle, can be used to diagnose DR and glaucoma. In addition, biometrics is a technology that uses an individual's physiological or behavioral characteristics to identify their identity [6]. The scleral vessel pattern is highly individual, with each person having a unique and relatively stable scleral vessel pattern that does not change significantly over time, making it suitable as a basis for long-term biometric recognition.

High-precision results for scleral segmentation and scleral vessel segmentation are important for the diagnosis of DR and glaucoma, as well as for biometrics. For sclera and scleral vessel segmentation, an image of the eye needs to be obtained using professional ophthalmic imaging equipment. These images are usually high resolution and clearly show the structures of the eye, including the sclera and vascular tissue structure. The acquired image needs to be preprocessed to improve image quality. Preprocessing steps may include denoising [7], contrast enhancement [8], standardization [9], contrast-limited adaptive histogram equalization (CLAHE) [10], etc., to facilitate subsequent segmentation processing. In the absence of automated tools, researchers need to manually label scleral and vascular areas, which is time-consuming and prone to errors. It not only requires researchers to have professional knowledge and experience to accurately identify and label vascular structures but also faces large-scale segmentation tasks. The efficiency and accuracy of manual labeling are difficult to meet actual needs. This not only greatly limits the timeliness of disease diagnosis but also hinders progress in subsequent research and applications.

With the continuous development of deep learning technology, especially in the field of image segmentation, deep neural networks have already demonstrated excellent performance. The development

*Corresponding author: Yansuo Yu, Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, China. Email: yuyansuo@bipt.edu.cn

of this technology has undoubtedly brought great potential and hope to the field of scleral vessel segmentation. In the field of image segmentation and other vascular segmentation, various models have proven their efficiency and accuracy. The successful application of these models provides new ideas and methods for scleral vessel segmentation.

We tried to apply the model that performed well in retinal segmentation to scleral vessel segmentation. In order to obtain more accurate scleral vessel segmentation, the eye image first needs to be accurately segmented in the scleral region. By extracting the sclera part of the eye image, we can focus the scleral vessel segmentation on the sclera area, thereby effectively reducing the interference of non-sclera areas on the segmentation results. This step is crucial for improving the accuracy of vascular segmentation. In order to more clearly show the vascular structure, we enhance the blood vessel area without sacrificing image quality and with as little loss of information as possible. This process aims to enhance the contrast between blood vessels and the surrounding background so that blood vessel structures can be more effectively identified and analyzed. A cropping strategy [11] was used to improve the ability to recognize the characteristics of small blood vessels and to alleviate the extreme imbalance between vascular and non-vascular areas. In terms of task nature, scleral segmentation can be classified as a region segmentation problem [12], while scleral vessel segmentation belongs to the more complex tree segmentation problem.

Although significant progress has been made in retinal vessel segmentation, systematic and quantitative research on scleral vessel segmentation remains limited. The accurate and efficient segmentation of scleral vessels continues to pose substantial challenges due to their distinct anatomical characteristics compared to retinal vessels, including pronounced curvature and complex overlap between superficial and deep vascular structures.

The goal of this study is to use three excellent models, UNet [13], NestNet [14], and Deephlabv3_Plus [15], to evaluate and quantitatively analyze the unified task of scleral segmentation and scleral vessel segmentation, with the aim of promoting research progress in this field. In this way, we aim to fill gaps in existing research and provide a more precise analytical tool for scleral vessel segmentation. Meanwhile, on the public dataset SBVPI [16], we will use evaluation metrics [17] such as accuracy, Dice, Intersection over Union (IoU), loss, sensitivity, specificity, and area under the curve (AUC) to comprehensively evaluate the performance and effectiveness of these three models in the scleral segmentation and scleral vessel segmentation tasks. The paper is organized as follows: The following section will provide an overview of relevant research progress in the fields of scleral segmentation and vessel segmentation. Section 3 will elaborate on the overall architecture of the task design, provide an in-depth interpretation of the preprocessing techniques used, and systematically introduce the framework structure of the usage model. Section 4 will present the experimental results of the methods used and analyze them in detail. Section 5 expands on the relevant issues and discusses them in depth. Finally, Section 6 summarizes the text and draws conclusions.

2. Related Works

The traditional scleral segmentation and scleral vessel segmentation rely on manual segmentation performed by experts with extensive knowledge. In recent years, with the significant advancements of deep learning in the medical field, automation in both scleral segmentation and scleral vessel segmentation has made remarkable progress.

2.1. Scleral segmentation

In the process of scleral vessel segmentation, scleral segmentation is a crucial step, as it involves the accurate separation of the scleral region from eye images. Traditional scleral segmentation methods include threshold-based approaches [18], edge detection methods [19], and region growing techniques [20]. Lucio et al. [21] applied Fully Convolutional Networks (FCNs) to address the scleral segmentation problem. FCNs are end-to-end deep learning models capable of learning directly from image pixels to segmentation masks without the need for manual feature extraction, allowing the model to adapt to complex image variations. Although some success has been achieved, there are still notable limitations in the accuracy of the results. To further improve segmentation accuracy, Lucio et al. [21] introduced Generative Adversarial Networks (GANs). GANs consist of a generator and a discriminator; the generator produces samples close to real data, while the discriminator distinguishes between real and generated data. GANs can be used to generate higher-quality training data, enhancing the model's ability to generalize to scleral images under varying conditions. However, GANs are prone to instability during training, and the adversarial process between the generator and discriminator may lead to mode collapse. Additionally, inaccuracies are observed in the segmentation of certain local regions in the results. Mistry and Ramakrishnan [22] explored methods for the automatic detection of ocular cancer in healthcare through machine learning and image analysis, which also involves scleral segmentation and analysis. However, machine learning methods typically handle only shallow features of images, struggling to capture deeper semantic information. Additionally, when processing complex image features and tasks, they often exhibit lower accuracy, flexibility, and robustness. Naqv and Loh [23] propose a model named Sclera-net, which effectively leverages the advantages of residual networks to learn complex features through the network and process image data from various sensors. Sclera-net is highly adaptable and can be tailored to diverse image data sources. However, its performance may degrade under varying lighting conditions or in the presence of image noise.

2.2. Vessel segmentation

In the field of vessel segmentation, although there is limited research directly applied to scleral vessel segmentation, the successful application of deep learning in retinal and other vascular segmentation fields suggests significant potential for its use in scleral vessel segmentation. Prior to the use of deep neural networks, researchers typically relied on traditional image processing techniques and machine learning methods for retinal vessel segmentation [24]. Staal et al. [25] proposed an automated method for segmenting blood vessels in the retina by generating feature vectors for each pixel, utilizing the scale and orientation-selective properties of Gabor filters. The extracted features were then classified using a Gaussian mixture model and a support vector machine classifier to separate vessels from non-vessels. In their study on detecting DR in retinal images, Al-Rawi et al. [26] employed a matched filter to extract vessel features, which were further analyzed to segment vessels in retinal images. You et al. [27] introduced a novel retinal vessel segmentation method that combines radial projection with semi-supervised learning,

improving segmentation accuracy, particularly in detecting fine capillaries.

Traditional machine learning methods for vessel segmentation still exhibit notable limitations. First, these approaches heavily rely on manually designed image features and are incapable of automatically capturing deep semantic information. As a result, their performance is restricted when dealing with complex vascular structures, such as fine branches, curved morphologies, or overlapping regions. Second, these methods are highly sensitive to variations in image quality and demonstrate poor robustness under conditions such as uneven illumination, noise interference, or image blurring. Furthermore, they generally lack strong generalization capabilities, making them difficult to adapt to diverse data sources or clinical environments.

With the emergence of deep neural networks, these models have demonstrated substantial potential in retinal vessel segmentation. Long et al. [28] proposed an FCN for semantic segmentation, defining a skip architecture that combines semantic information from deep, coarse layers with appearance information from shallow, fine layers to produce accurate and detailed segmentation results. Fine-tuned for segmentation tasks, FCN has shown promising potential for retinal vessel segmentation. In another study, Ronneberger et al. [29] introduced an innovative convolutional neural network architecture, UNet, which features a symmetric structure with an encoder to capture contextual information and a symmetric decoder for precise localization, employing skip connections to reduce information loss during upsampling. By combining high-resolution and low-resolution features, UNet has excelled in various biomedical segmentation applications and has become an essential tool in the biomedical segmentation field. Wang et al. [30] enhanced the UNet architecture by incorporating an additional encoder and channel attention mechanisms in the skip connections for retinal vessel segmentation. This dual-encoder and multi-scale feature fusion approach significantly improved the accuracy of retinal vessel segmentation. A model named “Claw U-Net,” developed by Chang Yao et al. [31], is proposed for scleral vessel segmentation through the integration of UNet and deep feature cascade technology. The capture capability of key image features is enhanced through deep feature cascades, significantly improving the detail handling of vessel segmentation and reducing information loss. However, the approach increases the computational complexity and the number of parameters, thereby elevating the demand for computational resources.

Although deep learning methods demonstrate outstanding performance in retinal vessel segmentation tasks, their application to scleral vessel segmentation remains limited, and the performance of related models requires further validation. Furthermore, systematic research and validation are urgently needed for the evaluation and quantitative analysis of integrated scleral segmentation and scleral vessel segmentation tasks.

3. Proposed Methodology

Scleral vessel segmentation plays a crucial role in machine vision and medical image analysis. Its primary goal is to accurately identify and extract vascular structures from eye images, which is vital for applications such as medical diagnosis and biometric identification. Although various models have been developed for retinal vessel segmentation, each with its unique strengths and limitations, scleral vessel segmentation remains a relatively underexplored area. Therefore, this study aims to explore the potential application of three advanced deep learning models—UNet, NestNet, and DeepLabv3_Plus—for the task of scleral vessel segmentation. In this section, the overall architecture of the task design is discussed in depth, the preprocessing techniques employed are thoroughly explained, and the framework structures of the models used are systematically introduced. To evaluate the performance of these models in scleral segmentation and scleral vessel segmentation tasks, a series of quantitative metrics is used to comprehensively assess and quantify the models’ performance.

3.1. Architecture overview

High-precision results in scleral vessel segmentation are of critical importance for medical diagnosis and biometric identification. To achieve this, precise scleral segmentation is essential to eliminate interference from non-scleral regions around the eye. This study employs three models—UNet, NestNet, and DeepLabv3_Plus—to conduct a preliminary quantitative analysis of scleral segmentation and scleral vessel segmentation, aiming to advance research in this field. As shown in Figure 1, the publicly available SBVPI (ScleraSeg) scleral segmentation dataset [15] is

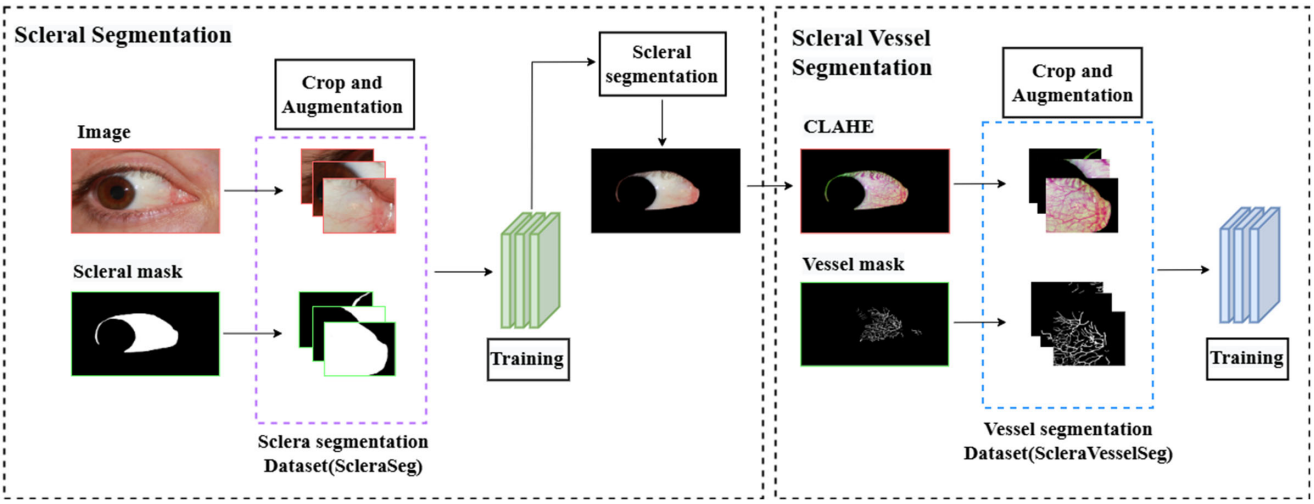

Figure 1. Overview of the joint sclera and scleral vessel segmentation

first preprocessed with cropping and image augmentation. Then, the three models are trained to meet the scleral segmentation task requirements. Next, the SBVPI (ScleraVesselSeg) scleral vessel segmentation dataset [15] undergoes cropping, CLAHE, and image enhancement, followed by training on the three models to achieve the scleral vessel segmentation task objectives. The input images are first cropped to meet the model's input requirements. Afterward, scleral segmentation is performed on the cropped images, and the results are stitched together to obtain the complete scleral segmentation output. The scleral segmentation result is then processed alongside the original image, retaining only the scleral region to focus the view on the scleral vessels. CLAHE enhancement is applied to the processed scleral region to improve contrast and detail. The image is then cropped again to meet the input requirements for the model, and scleral vessel segmentation is performed on the cropped image. Finally, the segmentation results are stitched together to obtain the scleral vessel segmentation output for the original input image. Through this series of operations, the models not only perform scleral segmentation but also achieve high-precision scleral vessel segmentation, meeting the core requirements of scleral vessel segmentation tasks. This provides a more accurate and reliable image analysis tool for medical diagnosis and biometric identification, with the potential for broad application in related fields.

3.2. Preprocessing

3.2.1. Contrast-limited adaptive histogram equalization

To improve the contrast between vessels and background regions, reduce noise, and make small and weakened vessel structures more prominent—while preserving image quality and minimizing information loss—we apply an image enhancement technique known as CLAHE.

CLAHE is an image enhancement method [32]. The fundamental principle involves dividing the image into several small blocks (local regions) and performing histogram equalization on each block individually to enhance local contrast. A threshold (clip limit) is set for each block's histogram to prevent excessive amplification of noise resulting from over-enhanced local contrast. Finally, interpolation is applied to smooth the boundaries between adjacent blocks, producing an enhanced image with improved overall contrast and controlled noise.

It is important to note that the CLAHE technique is not directly applied to the original image but is instead applied only to the green channel of the image. Compared to applying CLAHE to the entire image, focusing on enhancing the green channel more effectively improves the contrast of the vessels, highlighting the vascular details without significantly affecting the structure of the other color channels [33].

3.2.2. Cropping strategy

The images processed through scleral segmentation and CLAHE enhancement are still subject to certain limitations when used for scleral vessel segmentation model training. For instance, it is challenging to capture the local features of small vessels, and there is a significant imbalance between vascular and non-vascular regions. To address these issues and improve model performance while reducing the model's burden of learning from non-scleral vessel regions, a cropping strategy is applied to the dataset images. The images are cropped into smaller, fixed-size patches, which are then used to train the model.

Cropping the image into smaller patches offers several benefits for our model training, particularly in computer vision tasks [34]. For example: (1) patch cropping enables the model to focus on smaller regions, allowing it to learn the local features of fine vessels; (2) for datasets with limited scleral vessel segmentation data, the cropped patches increase dataset diversity, reduce overfitting, and improve the model's generalization ability; (3) it effectively alleviates the extreme imbalance between vascular and non-vascular regions in the image, helping to train a more balanced model. The model makes local decisions based on these cropped patches, and the results are later stitched together to form the global output.

To minimize interference from irrelevant regions, a cropping strategy is primarily focused on vessel areas. Specifically, random cropping from the central region of the image is performed to highlight scleral vessel features, and additional small patches are randomly cropped to increase sample diversity. Patches with a low proportion of vessel pixels are removed through threshold filtering, thereby improving both the training effectiveness and generalization capability of the model.

3.2.3. Augmentation

In image processing, data augmentation is a widely used technique designed to enhance a model's generalization ability and robustness by generating more diverse training samples [35]. In this study, several image enhancement techniques are employed to improve the model's performance in real-world applications.

During the image preprocessing stage, horizontal flipping is first applied. This technique, a simple and effective data augmentation method, generates a new image symmetric to the original one along the horizontal direction by mirroring it across the vertical center axis. Next, elastic deformation is applied. Elastic deformation is an image enhancement technique that simulates the surface deformation of objects in the real world [36]. Finally, a local distortion operation is introduced through grid distortion. Grid distortion modifies the shape of local regions in the image by shifting the grid points.

In summary, by incorporating horizontal flipping, elastic deformation, and local distortion through grid distortion, a large number of diverse training samples with real-world characteristics are generated. This approach significantly enhances the model's generalization ability, allowing it to perform tasks more accurately and stably when confronted with various complex scenarios and changes in real-world applications, thereby offering more reliable support for research and applications in related fields.

3.3. Model and architecture

3.3.1. UNet

The UNet model is an encoder-decoder architecture commonly used for image segmentation tasks. The encoder extracts features from the input image, while the decoder reconstructs these features through upsampling, producing an output image with the same dimensions as the input. Each decoder block utilizes skip connections to concatenate high-resolution feature maps from the encoder with low-resolution feature maps from the decoder, allowing for improved restoration of image details.

The model primarily consists of two components: the Encode_block and Decoder_block. The Conv_block involved is uniformly defined as follows: a convolution operation is performed using a 3×3 convolutional kernel, with padding set to "same" to ensure that the output size matches the input size. A batch normalization layer is then applied to accelerate training and improve stability, followed by the ReLU activation function to introduce

non-linearity. This process is repeated twice, forming a standard convolution block with two convolutional layers. The Encoder_block extracts features through the Conv_block and reduces the spatial dimensions of the feature map using MaxPool2D with 2 × 2 pooling. The Decoder_block first performs upsampling using Conv2DTranspose, which doubles the spatial dimensions of the feature map. This operation is the reverse of pooling and restores the image resolution based on the learned weights. The upsampled feature map is then concatenated with the skip connection feature map from the encoder using the Concatenate layer, allowing the decoder to utilize high-resolution information from the encoder to recover image details. The concatenated features are then processed further through the Conv_block to extract additional features.

The model's image processing flow is described as follows. First, four encoder blocks, each employing 64, 128, 256, and 512 convolution filters, are used to extract features, while a max pooling layer is applied to reduce the spatial dimensions of the feature maps. A bottleneck with 1024 convolution filters is then employed to capture the deepest level of features (i.e., high-level image information). Next, four decoder blocks progressively restore the spatial resolution of the feature maps, and skip connections are used to integrate detailed information from the corresponding encoder layers. In each decoder block, a transposed convolution is performed for upsampling, and spatial details are recovered by concatenating the skip-connected feature maps and applying additional convolution operations. Finally, a 1 × 1 convolution reduces the number of channels to one, generating a single-channel output. A Sigmoid activation function is then applied to produce a probability value indicating whether each pixel belongs to a vessel.

The UNet model is a classic convolutional neural network architecture specifically designed for image segmentation tasks. Its distinctive encoder-decoder structure enables the extraction of rich multi-scale features while efficiently recovering spatial details. By introducing skip connections between corresponding layers of the encoder and decoder, UNet effectively integrates high-level semantic information with fine-grained spatial details, which significantly improves segmentation accuracy—particularly for small or structurally complex targets. Due to its strong contextual capture ability and precise localization performance, UNet is widely employed as a highly effective solution for medical image segmentation, remote sensing, and related fields.

3.3.2. NestNet

Both the NestNet model and the aforementioned UNet model are deep learning models employed for image segmentation tasks. Many similarities in structure and design philosophy are observed between them; however, NestNet is regarded as an extension and improvement of UNet. A U-shaped architecture is adopted by NestNet, preserving the encoder–decoder framework. In the decoding stage, additional skip connections and upsampling strategies are introduced to enhance feature utilization and improve model performance. A detailed description of NestNet is provided below.

A Conv_batchnorm_relu_block is first constructed, which defines a convolution block that employs a 3 × 3 kernel with "same" padding to maintain the spatial dimensions of the input. Subsequently, a batch normalization layer is applied to accelerate training and enhance stability, followed by a ReLU activation function to introduce non-linearity. In the encoder block, features are extracted using the Conv_batchnorm_relu_block and are downsampled via a 2 × 2 pooling operation using AvgPool2D, thereby reducing the spatial dimensions of the feature maps. In the decoder block, upsampling is performed using Conv2DTranspose, which enlarges the spatial dimensions of the feature maps; this operation restores the image resolution through learned weights, counteracting the effect of pooling. Next, the upsampled feature maps are concatenated with skip connection feature maps from the encoder via the Concatenate operation. Unlike the UNet model, NestNet incorporates not only the encoder feature maps at each decoding stage but also the feature maps from the preceding decoder layer. This approach enables the early decoder layers to acquire features from multiple locations, thereby facilitating the capture of multi-scale information.

The model processes image information as follows: First, five encoder blocks are employed to extract features using 32, 64, 128, 256, and 512 convolutional kernels, respectively, while the spatial dimensions of the feature maps are reduced by an average pooling layer. Multiple convolution and pooling layers are constructed to extract multi-level features. At each encoder, the extracted features are progressively upsampled using transposed convolution, and the upsampled feature maps are concatenated with the feature maps from the preceding level to obtain richer information. Finally, a 1 × 1 convolution is applied to generate the final output feature map, and a Sigmoid activation function is employed to produce a probability value that indicates the likelihood of a pixel belonging to a vessel.

NestNet is characterized by the introduction of dense skip connections, which effectively facilitate the comprehensive fusion of features across different depths and scales. By integrating shallow fine-grained features with deep high-level semantic features, NestNet substantially enhances the ability to capture object boundaries and details, thereby improving accuracy in segmentation tasks involving structurally complex or variably sized targets. Additionally, the dense multi-level skip connections strengthen the internal information flow within the network, alleviating the gradient vanishing problem commonly observed in traditional deep networks. As a result, the model is better able to learn and represent complex structural features.

3.3.3. Deeplabv3_Plus

The architecture of the Deeplabv3_Plus model differs from the UNet and NestNet models in that it is not strictly U-shaped. However, it incorporates design elements similar to the U-shape, particularly by combining low-level and high-level features to improve boundary details in segmentation results. Dilated convolutions [37] and Atrous Spatial Pyramid Pooling (ASPP) [38] are employed to capture multi-scale contextual information, resulting in superior performance in complex scenes and multi-scale challenges. A detailed description of Deeplabv3_Plus follows.

The ASPP module is first constructed by the model, in which global information is extracted by pooling layers and mapped via 1 × 1 convolutions. Convolutions with various dilation rates are employed to process different feature regions to capture multi-scale information. All processed feature maps are concatenated through a Concatenate operation and then integrated using a 1 × 1 convolution. ResNet50 [39] is utilized as the backbone feature extraction network, without pre-trained weights and excluding the top fully connected layers. Output from the conv4_block6_out layer is extracted as high-level features and passed to the ASPP module for processing and upsampling. The output from the conv2_block2_out layer is processed as low-level features through a 1 × 1 convolution, batch normalization, and ReLU activation. High-level and low-level features are concatenated to fuse semantic information across different layers. In the Decoder_block, features undergo two 3 × 3 convolutions with ReLU activations, followed by four-fold upsampling to restore the feature map to match the input's spatial dimensions.

The model processes image information as follows: First, high-level features are extracted from the conv4_block6_out layer of ResNet50 and passed to the ASPP module to capture information from different receptive fields, including global context. Next, low-level features from the conv2_block2_out layer are processed with a 1 × 1 convolution, batch normalization, and ReLU activation. The features are then concatenated to integrate semantic information. The Decoder_block restores the feature map to match the input's spatial dimensions. Finally, a 1 × 1 convolution with Sigmoid activation outputs a probability value indicating the likelihood that a pixel belongs to a scleral vessel.

Deeplabv3_Plus is characterized by the innovative integration of the ASPP module, which significantly enhances the network's ability to extract multi-scale contextual information. The ASPP module effectively fuses features from different receptive fields by applying parallel atrous convolutions with varying dilation rates, enabling the model to simultaneously capture both global structures and local details of targets within an image. This design not only markedly improves the recognition and segmentation of objects with diverse sizes and shapes but also expands the receptive field of the feature maps, thereby further increasing the accuracy of segmentation boundaries. The network architecture of Deeplabv3_Plus is illustrated in Figure 2.

4. Experiments

In the experimental section of this paper, a detailed analysis is conducted on the performance of the UNet, NestNet, and Deeplabv3_Plus models in the tasks of scleral segmentation and scleral vessel segmentation, in order to evaluate their application potential and limitations.

4.1. Experimental setup

4.1.1. Dataset

The publicly available SBVPI (scleral vessel, periocular, and sclera) medical imaging dataset is used in this study. It consists of 1,858 high-resolution RGB eye images from 55 subjects. The dataset takes into account variations in gaze direction and viewpoint, as even small changes can result in significant differences in the appearance of scleral vessels. To maintain consistency in image size across the entire dataset, all images are resized to 3000 × 1700 pixels, corresponding to the average size of the region of interest extracted.

4.1.2. Implementation details

We first perform precise matching of the image data for scleral segmentation and scleral vessel segmentation. The dataset for scleral segmentation research includes 1,830 eye images along with their corresponding scleral region masks, while the dataset for scleral vessel segmentation research contains 127 eye images and their respective scleral vessel region masks. Both datasets are strictly divided into training, validation, and test sets in a ratio of 80%, 10%, and 10%, respectively.

In processing the scleral vessel segmentation dataset, scleral segmentation is first performed, followed by the application of CLAHE to enhance the contrast of the vascular structures in the images. Subsequently, a refined cropping strategy is applied to both datasets, and the image size is uniformly adjusted to 256 × 256 pixels. Additionally, basic data augmentation techniques, including horizontal rotation, elastic transformation, and grid distortion, are applied to both datasets.

The proposed model is implemented in TensorFlow and initialized according to the selected architecture (UNet, NestNet, or DeepLabv3_Plus). Throughout the experiment, all input images and corresponding masks were resized to 256 × 256 pixels using bilinear interpolation to ensure consistency and facilitate batch processing. The Adam optimizer was employed to optimize the model, with a batch size of 4 and an initial learning rate of 0.0001. The learning rate followed a step decay schedule, where it was multiplied by 0.5 every 10 epochs. The total number of training epochs was set to 100. The loss function used was Dice loss. Early stopping was applied with a patience of 100 epochs to

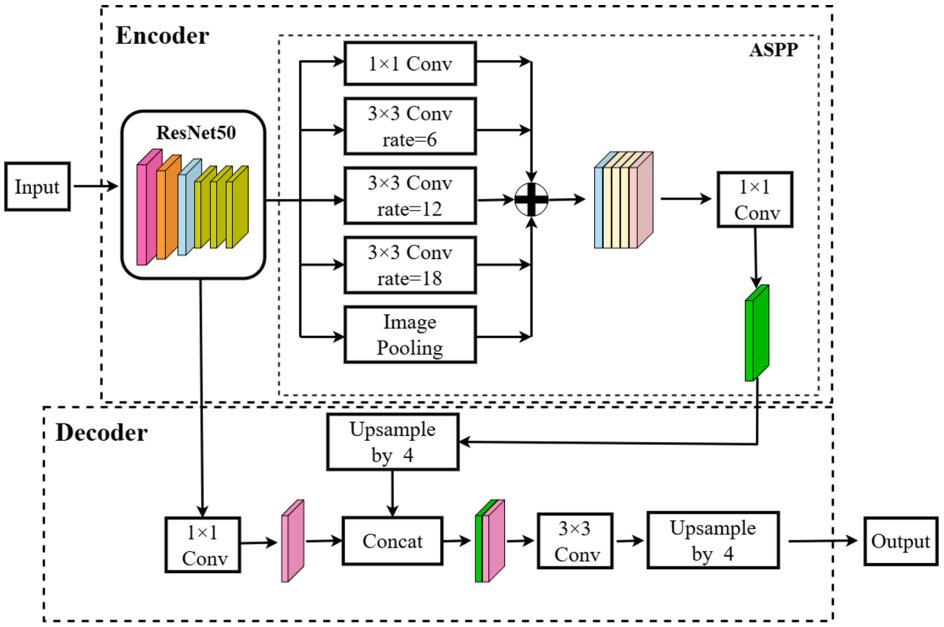

Figure 2. The architecture of Deeplabv3_Plus. In the encoder, ResNet50 is employed as the backbone network to extract high-level features, while the ASPP module captures multi-scale contextual information from various receptive fields. In the decoder, the extracted features are fused, and the feature maps are subsequently restored to the original spatial dimensions of the input images

prevent overfitting. All experiments were conducted using a single NVIDIA RTX 3080 GPU.

4.1.3. Performance metrics

In this study, a series of key evaluation metrics is used to comprehensively assess the model's performance. First, accuracy is employed to measure the overall classification correctness, while the area under the receiver operating characteristic curve (AUC-ROC) provides a quantitative assessment of the model's ability to distinguish between positive and negative samples. Additionally, sensitivity and specificity are introduced to analyze the model's performance across different categories. Sensitivity measures the model's ability to correctly identify positive samples, while specificity evaluates the model's ability to identify negative samples. To further analyze the model's segmentation performance, two key metrics—Dice coefficient (Dice_coef) and IoU—are incorporated. The IoU metric focuses on measuring the overlap between predicted results and ground truth labels, while the Dice coefficient is used to assess the similarity between two sample sets. Given the relatively small target areas in the scleral vessel segmentation task, the Dice coefficient, due to its high sensitivity to small targets, more effectively evaluates the segmentation performance of vessels. For the scleral segmentation task, due to the importance of shape matching, we prefer to use the IoU metric to assess the model's performance. By combining multiple metrics, we can comprehensively evaluate the model's performance across different segmentation tasks.

4.2. Performance comparison

Evaluating the model's performance in scleral segmentation and scleral vessel segmentation is crucial for improving disease diagnosis techniques and biometric methods. Our evaluation process first focuses on the model's quantitative performance in these two tasks and then further explores the visual representation of the model's results for scleral segmentation and scleral vessel segmentation. Through this comprehensive evaluation approach, the potential and effectiveness of the model in handling both scleral segmentation and scleral vessel segmentation tasks can be fully assessed.

The performance metrics of UNet, NestNet, and Deeplabv3_Plus for scleral segmentation and scleral vessel segmentation tasks are provided below. It is important to note that the dataset used for the scleral segmentation task is ScleraSeg, while the dataset for the scleral vessel segmentation task is ScleraVesselSeg.

4.2.1. Scleral segmentation performance

As shown in Table 1, the accuracies of the three models are 0.9653, 0.9614, and 0.9656, respectively. However, model performance should not be evaluated solely based on accuracy; other metrics such as sensitivity, specificity, Dice coefficient, and IoU are also essential. In particular, sensitivity and the Dice coefficient provide a more precise assessment of segmentation performance. These comprehensive metrics collectively reflect the model's ability to distinguish between scleral and non-scleral regions, as well as its performance in shape matching.

To provide a more intuitive and comprehensive analysis of model performance, the visual results of the three models in the scleral segmentation task, along with their ROC curves, are presented in Figures 3 and 4. These images and curves offer an insightful basis for further evaluating the models' performance.

In Figure 3, (A) shows the overall image used as input for scleral segmentation, (B) shows the corresponding scleral segmentation mask for the overall image, (C) presents the scleral segmentation result from the UNet model applied to the input image, (D) displays the scleral segmentation result from the NestNet model, and (E) illustrates the scleral segmentation result from the Deeplabv3_Plus model.

A detailed comparison between the scleral segmentation results and the mask images of the input images reveals the following observations: when examining the scleral segmentation of the overall images, all three models accurately identify most of the scleral region boundaries. Notably, Deeplabv3_Plus demonstrates exceptional performance in the eye-corner region, showing not only high accuracy but also smooth edges, which highlights its advantage in handling fine details.

Figure 4 presents the ROC curves and AUC for UNet, NestNet, and Deeplabv3_Plus in scleral segmentation; the AUC values are 0.98, 0.97, and 0.97, respectively.

4.2.2. Scleral vessel segmentation performance

In Table 2, the three models achieved accuracy values of 0.9328, 0.9267, and 0.9285, respectively. Compared to the scleral segmentation task, the performance of the three models in sensitivity and specificity decreased—likely due to the higher complexity of the scleral vessel segmentation task, in which the vascular structures are fine and difficult to identify. In addition, Dice coefficient and sensitivity are critical metrics for evaluating model performance in the high-sensitivity detection of small targets. The three models achieve Dice coefficients of 0.4821, 0.5062, and 0.4586 and sensitivities of 0.4821, 0.5062, and 0.4586, respectively. These results indicate that NestNet slightly outperforms the other two models on both metrics, demonstrating superior sensitivity and segmentation accuracy in scleral vessel segmentation tasks. Therefore, NestNet is more suitable for precise segmentation applications involving small targets.

To provide a more intuitive and comprehensive analysis of model performance, the visual results of three models in the scleral vessel segmentation task, along with their ROC curves, are presented in Figures 5 and 6. These images and curves offer a direct basis for further analysis of model performance.

In Figure 5, (A) shows the input image and its zoomed-in view, (B) displays the input image after scleral segmentation and the subsequent image enhanced using CLAHE, along with its zoomed-in view; (C) presents the ground truth image for scleral vessel segmentation of the input image and its zoomed-in view; (D) illustrates the result of the UNet model for scleral vessel segmentation on the input image and its zoomed-in view; (E) shows the result of the NestNet model

Table 1. Direct performance comparison on scleral segmentation tasks

Model	Accuracy	Sensitivity	Specificity	Dice	IoU
UNet	0.9653	0.9692	0.9412	0.9706	0.9546
NestNet	0.9614	0.9661	0.9347	0.9671	0.9511
Deeplabv3_Plus	0.9656	0.9694	0.9421	0.9723	0.9484

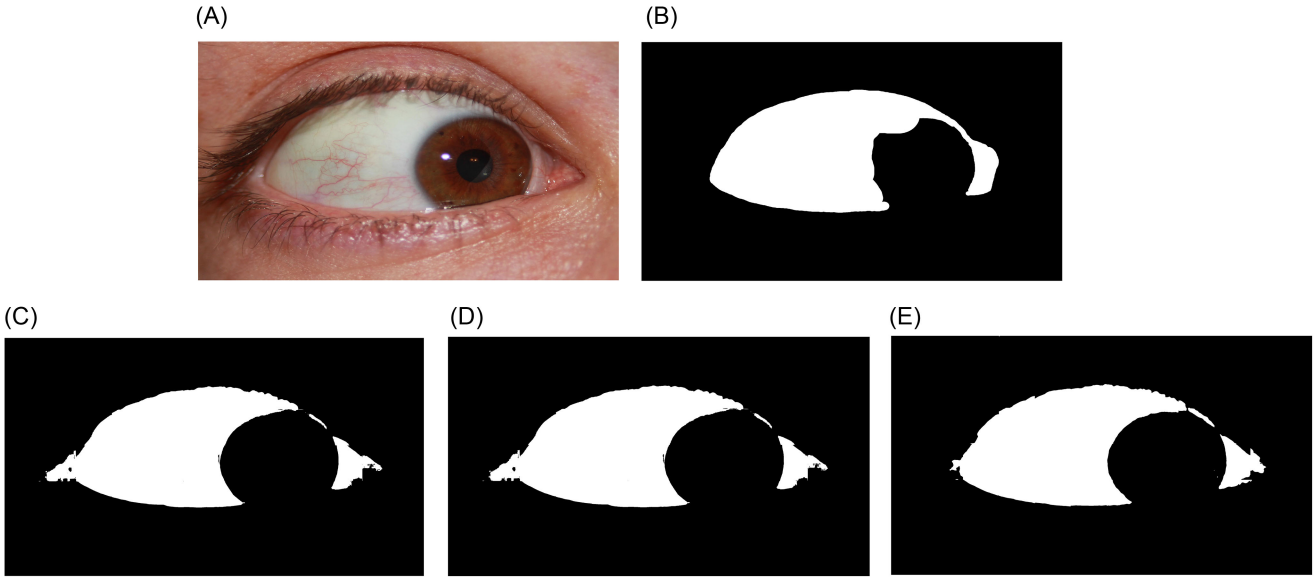

Figure 3. Visualization results for the scleral segmentation task. Panel (A) displays an ocular image, while panel (B) shows its corresponding ground truth. Panels (C), (D), and (E) depict the scleral segmentation outcomes obtained using the UNet, NestNet, and Deeplabv3_Plus models, respectively

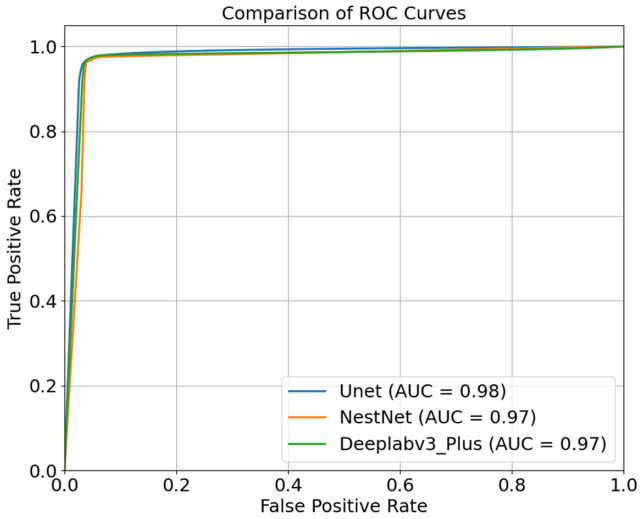

Figure 4. Receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) values for the scleral segmentation task

for scleral vessel segmentation on the input image and its zoomed-in view; and (F) provides the result of the Deeplabv3_Plus model for scleral vessel segmentation on the input image and its zoomed-in view.

Figure 6 presents the ROC curves and AUC values for scleral vessel segmentation using the UNet, NestNet, and Deeplabv3_Plus models; the corresponding AUC values are 0.85, 0.84, and 0.77, respectively.

4.2.3. Ablation experiments and statistical analysis

To evaluate the impact of CLAHE preprocessing and cropping strategies on segmentation outcomes and to eliminate the possibility of incidental experimental results, relevant ablation experiments and statistical analyses are conducted for the scleral vessel segmentation task. Given that the primary focus of the proposed integrated sclera and scleral vessel segmentation method lies in scleral vessel segmentation, this section specifically analyzes the effectiveness and role of CLAHE and cropping strategies in scleral vessel segmentation.

The experimental design is as follows: This section aims to evaluate the impact of CLAHE image enhancement and cropping strategies on scleral vessel segmentation performance. The experiment is divided into two groups, ensuring that both groups use the exact same configuration parameters. In the NoCLAHECrop group, CLAHE image enhancement and cropping strategies are not applied to the data, while in the CLAHECrop group, these strategies are implemented. By comparing the results from both groups, the actual effectiveness of CLAHE and cropping strategies in improving scleral vessel segmentation performance can be determined.

When evaluating the image segmentation performance of the model, sensitivity and Dice are considered two crucial metrics. Therefore, special attention is paid to the variations in these parameters. As shown in Table 3, the segmentation performance of the three models is significantly improved when CLAHE image enhancement and cropping strategies are applied, compared to when these strategies are not used. This indicates that CLAHE image enhancement and cropping strategies effectively enhance the model's performance in scleral vessel segmentation tasks.

The experimental design for statistical analysis is structured as follows: Each of the three models is subjected to five independent

Table 2. Direct performance comparison on scleral vessel segmentation tasks

Model	Accuracy	Sensitivity	Specificity	Dice	IoU
UNet	0.9328	0.5473	0.9420	0.4821	0.3371
NestNet	0.9267	0.6371	0.9343	0.5062	0.3397
Deeplabv3_Plus	0.9285	0.5097	0.9536	0.4586	0.2943

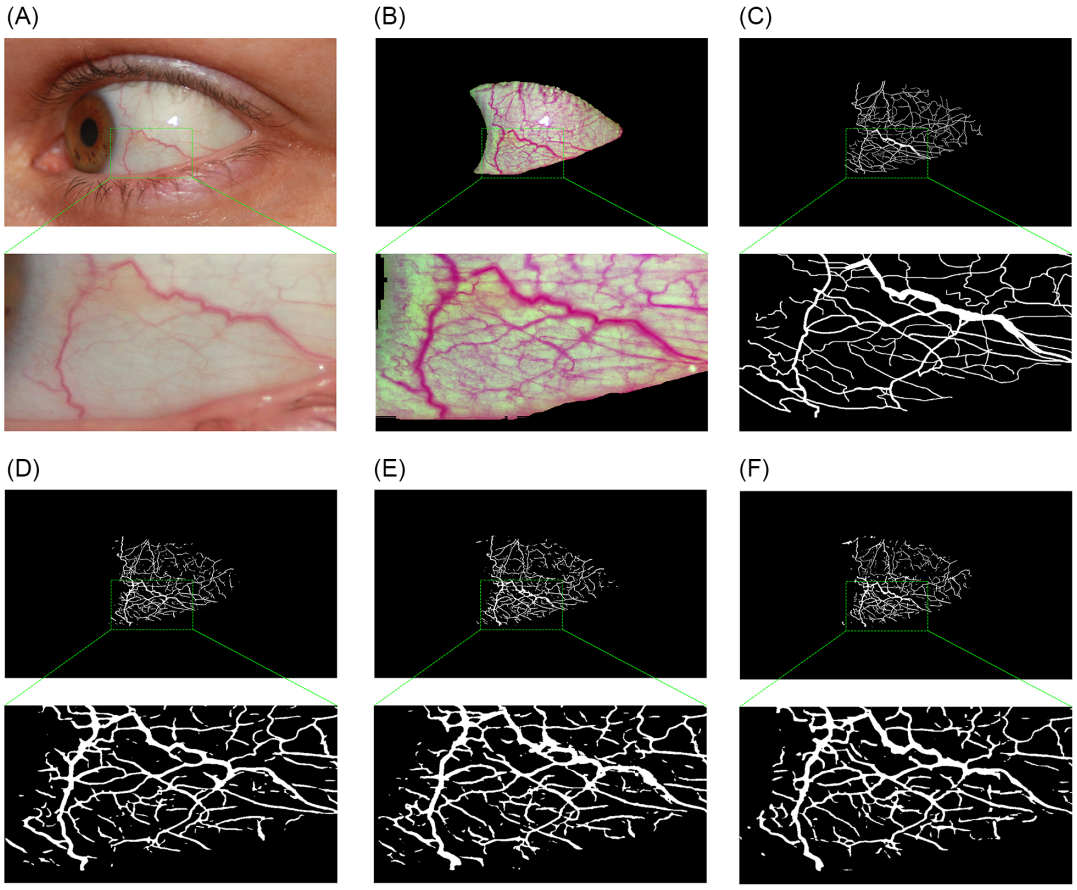

Figure 5. Visualization results for the scleral vessel segmentation task. Image (A) shows the ocular image, while image (B) depicts the ocular image following scleral segmentation and CLAHE processing. Image (C) represents the corresponding ground truth. Images (D), (E), and (F) display the scleral vessel segmentation results produced by the UNet, NestNet, and Deeplabv3_Plus models, respectively

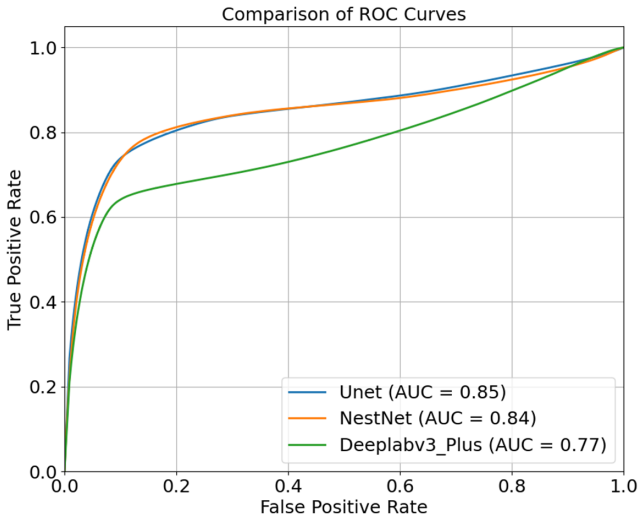

Figure 6. Receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values for the scleral vessel segmentation task

repeat experiments. In these experiments, all settings are kept consistent, except for the variation in random seed, encompassing dataset loading, weight parameters, and test set selection. Upon completing the five independent repeat experiments, a statistical analysis of the models' performance on the test set is conducted, and confidence intervals are calculated to ensure the results' reliability and accuracy. The resulting boxplots are illustrated in Figure 7.

Table 3. Comparison of CLAHE and clipping strategy ablation

Model	Strategy	Sensitivity	Dice
UNet	NoCLAHECrop	0.4732	0.2403
	CLAHECrop	0.5473	0.4821
NestNet	NoCLAHECrop	0.4070	0.3321
	CLAHECrop	0.6371	0.5062
Deeplabv3_Plus	NoCLAHECrop	0.2449	0.2161
	CLAHECrop	0.5097	0.4586

5. Discussion

In tasks involving scleral segmentation and scleral vessel segmentation, Deeplabv3_Plus and NestNet models demonstrate superior performance. In contrast, the performance of the Sclera-net model declines in the presence of image noise, while the Claw U-Net model is limited by increased computational complexity and a higher number of parameters. Moreover, these models still lack precise evaluation and systematic analysis for integrated tasks of scleral segmentation and scleral vessel segmentation. This study evaluates and conducts a quantitative analysis of the combined tasks of scleral segmentation and scleral vessel segmentation, confirming

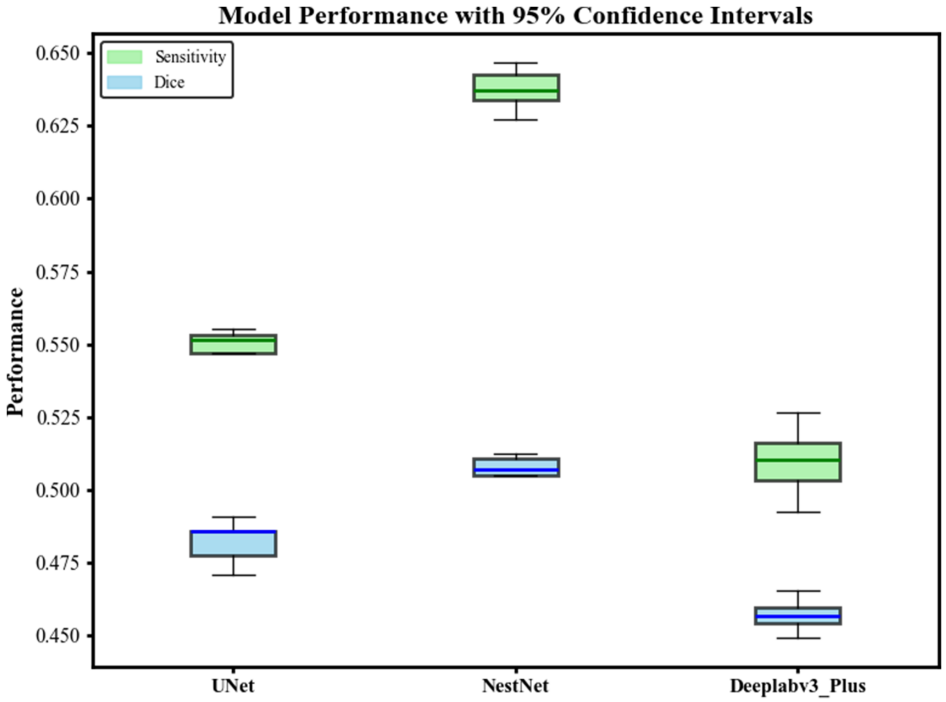

Figure 7. Analysis of the confidence intervals of model performance. The confidence interval analysis of model performance is carried out to compare the three models in terms of their experimental performance. This involves analyzing the 95% confidence intervals for sensitivity and Dice metrics

their feasibility and demonstrating significant potential, certain limitations persist. First, high-quality scleral vessel segmentation datasets are relatively scarce, restricting both the scale of datasets used in experiments and the evaluation of model generalization. Although the SBVPI dataset offers high-resolution images, its size remains limited; thus, optimizing performance on small-scale datasets during model training remains an important research direction. Second, the observation of scleral vessels is challenging—particularly for deep vessels, which are finer and less visible compared with superficial vessels—complicating their identification during segmentation. Additionally, the tortuous nature of scleral vessels, along with their varying lengths and diameters, further complicates segmentation. Given the small proportion of vessels relative to the background, even minor deviations in predictions can lead to suboptimal performance metrics. Future research may explore the integration of advanced attention mechanisms or segmentation models to enhance the segmentation of fine and deep vessels and improve vessel smoothness, thereby increasing their practical utility.

6. Conclusion

This study investigates the application of several models, which have demonstrated superior performance in retinal vessel segmentation, to the integrated task of sclera and scleral vessel segmentation. Through evaluation and quantitative analysis, these models are shown to exhibit significant potential for this task. Furthermore, applying CLAHE to the green channel of images for enhancement, along with cropping strategies, significantly improves segmentation performance. Experiments conducted on the publicly available SBVPI dataset indicate that this approach holds considerable promise for sclera and scleral vessel segmentation tasks, providing a solid foundation for medical diagnosis and biometric fields. This method is expected to achieve higher precision in scleral vessel segmentation results, further advancing the exploration and application of deep learning in this domain and promoting the practical implementation of related technologies.

Funding Support

This work was supported by the fund of Beijing Municipal Education Commission (No. 22019821001) and the National College Students Innovation and Entrepreneurship Training Program (No. 2025J00105).

Ethical Statement

This study does not require any kind of ethical approval as it completely relies on the use of computational techniques and models. This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the first author, Yongbin Qi, at peninsulazz@163.com.

Author Contribution Statement

Yongbin Qi: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Baochen Zhen:** Investigation, Writing – review & editing. **Jiaming Wang:** Investigation. **Shilin Zhao:** Investigation. **Yansuo Yu:** Conceptualization, Methodology, Formal analysis,

Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Qiang Liu:** Resources, Supervision, Funding acquisition.

References

[1] Vitek, M., Das, A., Lucio, D. R., Zanlorensi, L. A., Menotti, D., Nourmohammadi Khiarak, J., . . . , & Štruc, V. (2022). Exploring bias in sclera segmentation models: A group evaluation approach. *IEEE Transactions on Information Forensics and Security*, *18*, 190–205. https://doi.org/10.1109/TIFS.2022.3216468

[2] Das, S., de Ghosh, I., & Chattopadhyay, A. (2021). An efficient deep sclera recognition framework with novel sclera segmentation, vessel extraction and gaze detection. *Signal Processing: Image Communication*, *97*, 116349. https://doi.org/10.1016/j.image.2021.116349

[3] Kropp, M., Golubnitschaja, O., Mazurakova, A., Koklesova, L., Sargheini, N., Vo, T. T. K. S., . . . , & Thumann, G. (2023). Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—Risks and mitigation. *Epma Journal*, *14*(1), 21–42. https://doi.org/10.1007/s13167-023-00314-8

[4] World Health Organization. (2019). *World report on vision*. https://www.who.int/docs/default-source/documents/publications/world-vision-report-accessible.pdf

[5] Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T., & Pfeiffer, N. (2020). The diagnosis and treatment of glaucoma. *Deutsches Ärzteblatt International*, *117*(13), 225–234. https://doi.org/10.3238/arztebl.2020.0225

[6] Abdulrahman, S. A., & Alhayani, B. (2023). A comprehensive survey on the biometric systems based on physiological and behavioural characteristics. *Materials Today: Proceedings*, *80*, 2642–2646. https://doi.org/10.1016/j.matpr.2021.07.005

[7] Goyal, B., Dogra, A., Agrawal, S., Sohi, B. S., & Sharma, A. (2020). Image denoising review: From classical to state-of-the-art approaches. *Information Fusion*, *55*, 220–244. https://doi.org/10.1016/j.inffus.2019.09.003

[8] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, *65*(5), 545–563. https://doi.org/10.1111/1754-9485.13261

[9] Lee, S. B., Hong, Y., Cho, Y. J., Jeong, D., Lee, J., Yoon, S. H., . . . , & Cheon, J. E. (2023). Deep learning-based computed tomography image standardization to improve generalizability of deep learning-based hepatic segmentation. *Korean Journal of Radiology*, *24*(4), 294–304. https://doi.org/10.3348/kjr.2022.0588

[10] Zhao, X., Tao, S., Ku, R., & Wei, Z. (2024). Application and analysis of medical image processing based on improved CLAHE. In *International Conference on Artificial Intelligence in China*, 583–591. https://doi.org/10.1007/978-981-99-7545-7_59

[11] Magboo, M. S. A., & Coronel, A. D. (2024). Effects of cropping vs resizing on the performance of brain tumor segmentation models. In *2024 International Conference on Computer, Information and Telecommunication Systems*, 1–8. https://doi.org/10.1109/CITS61189.2024.10608031

[12] Xu, D., Dong, W., & Zhou, H. (2022). Sclera recognition based on efficient sclera segmentation and significant vessel matching. *The Computer Journal*, *65*(2), 371–381. https://doi.org/10.1093/comjnl/bxaa051

[13] Han, J., Wang, Y., & Gong, H. (2022). Fundus retinal vessels image segmentation method based on improved U-Net. *IRBM*, *43*(6), 628–639. https://doi.org/10.1016/j.irbm.2022.03.001

[14] Chen, Y., Ma, B., & Xia, Y. (2020). α-UNet++: A data-driven neural network architecture for medical image segmentation. In *MICCAI Workshop on Domain Adaptation and Representation Transfer*, 3–12. https://doi.org/10.1007/978-3-030-60548-3_1

[15] Tang, M. C. S., Teoh, S. S., & Ibrahim, H. (2022). Retinal vessel segmentation from fundus images using DeepLabv3+. In *2022 IEEE 18th International Colloquium on Signal Processing & Applications*, 377–381. https://doi.org/10.1109/CSPA55076.2022.9781891

[16] Das, S., de Ghosh, I., & Chattopadhyay, A. (2022). Sclera biometrics in restricted and unrestricted environment with cross dataset evaluation. *Displays*, *74*, 102257. https://doi.org/10.1016/j.displa.2022.102257

[17] Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1), 210. https://doi.org/10.1186/s13104-022-06096-y

[18] Kako, N. A., & Abdulazeez, A. M. (2022). Peripapillary atrophy segmentation and classification methodologies for glaucoma image detection: A review. *Current Medical Imaging*, *18*(11), e080322201870. https://doi.org/10.2174/1573405618666220308112732

[19] Adeniyi, T. T., Omolewa, O. T., & Adeniyi, J. K. (2024). Sclera boundary localization using circular hough transform and a modified run-data based algorithm. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *22*(3), 720–731. http://doi.org/10.12928/telkomnika.v22i3.24588

[20] Cheng, Z., & Wang, J. (2020). Improved region growing method for image segmentation of three-phase materials. *Powder Technology*, *368*, 80–89. https://doi.org/10.1016/j.powtec.2020.04.032

[21] Lucio, D. R., Laroca, R., Severo, E., Britto, A. S., & Menotti, D. (2018). Fully convolutional networks and generative adversarial networks applied to sclera segmentation. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems*, 1–7. https://doi.org/10.1109/BTAS.2018.8698597

[22] Mistry, J., & Ramakrishnan, R. (2023). The automated eye cancer detection through machine learning and image analysis in healthcare. *Journal of Xidian University*, *17*(8), 763–772.

[23] Naqvi, R. A., & Loh, W. K. (2019). Sclera-Net: Accurate sclera segmentation in various sensor images based on residual encoder and decoder network. *IEEE Access*, *7*, 98208–98227. https://doi.org/10.1109/ACCESS.2019.2930593

[24] Khandouzi, A., Ariafar, A., Mashayekhpour, Z., Pazira, M., & Baleghi, Y. (2022). Retinal vessel segmentation, a review of classic and deep methods. *Annals of Biomedical Engineering*, *50*(10), 1292–1314. https://doi.org/10.1007/s10439-022-03058-0

[25] Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, *23*(4), 501–509. https://doi.org/10.1109/TMI.2004.825627

[26] Al-Rawi, M., Qutaishat, M., & Arrar, M. (2007). An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology and Medicine*, *37*(2), 262–267. https://doi.org/10.1016/j.compbiomed.2006.03.003

[27] You, X., Peng, Q., Yuan, Y., Cheung, Y. M., & Lei, J. (2011). Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognition*, *44*(10–11), 2314–2324. https://doi.org/10.1016/j.patcog.2011.01.007

[28] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

[29] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention: 18th International Conference*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[30] Wang, B., Qiu, S., & He, H. (2019). Dual encoding U-Net for retinal vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention: 22nd International Conference*, 84–92. https://doi.org/10.1007/978-3-030-32239-7_10

[31] Yao, C., Tang, J., Hu, M., Wu, Y., Guo, W., Li, Q., & Zhang, X. P. (2021). Claw U-Net: A unet variant network with deep feature concatenation for scleral blood vessel segmentation. In *Artificial Intelligence: First CAAI International Conference*, 67–78. https://doi.org/10.1007/978-3-030-93049-3_6

[32] Rao, B. S. (2020). Dynamic histogram equalization for contrast enhancement for digital images. *Applied Soft Computing*, *89*, 106114. https://doi.org/10.1016/j.asoc.2020.106114

[33] Abdulsahib, A. A., Mahmoud, M. A., Mohammed, M. A., Rasheed, H. H., Mostafa, S. A., & Maashi, M. S. (2021). Comprehensive review of retinal blood vessel segmentation and classification techniques: Intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *10*, 20. https://doi.org/10.1007/s13721-021-00294-7

[34] Takahashi, R., Matsubara, T., & Uehara, K. (2018). RICAP: Random image cropping and patching data augmentation for deep CNNs. In *Proceedings of The 10th Asian Conference on Machine Learning, 95*, 786–798.

[35] Khosla, C., & Saini, B. S. (2020). Enhancing performance of deep learning models with different data augmentation techniques: A survey. In *2020 International Conference on Intelligent Engineering and Management*, 79–85. https://doi.org/10.1109/ICIEM48762.2020.9160048

[36] Argelaguet, F., Gómez Jáuregui, D. A., Marchal, M., & Lécuyer, A. (2013). Elastic images: Perceiving local elasticity of images through a novel pseudo-haptic deformation effect. *ACM Transactions on Applied Perception*, *10*(3), 17. https://doi.org/10.1145/2501599

[37] Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv Preprint:1511.07122*. https://doi.org/10.48550/arXiv.1511.07122

[38] Qiu, X. (2022). U-Net-ASPP: U-Net based on atrous spatial pyramid pooling model for medical image segmentation in COVID-19. *Journal of Applied Science and Engineering*, *25*(6), 1167–1176. https://doi.org/10.6180/jase.202212_25(6).0012

[39] Elharrouss, O., Akbari, Y., Almadeed, N., & Al-Maadeed, S. (2024). Backbones-review: Feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision. *Computer Science Review*, *53*, 100645. https://doi.org/10.1016/j.cosrev.2024.100645

How to Cite: Qi, Y., Zhen, B., Wang, J., Zhao, S., Yu, Y., & Liu, Q. (2025). Deep Learning for Joint Scleral and Scleral Vessel Segmentation: A Comparative Study. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN52025662