Article

A Segmentation-Based Automated Corneal Ulcer Grading System for Ocular Staining Images Using Deep Learning and Hough Circle Transform

Dulyawat Manawongsakul [1] and Karn Patanukhom [2,*]

[1] Data Science Consortium, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand; dulyawat_mana@cmu.ac.th

[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

[*] Correspondence: karn.patanukhom@cmu.ac.th

Abstract: Corneal ulcer is a prevalent ocular condition that requires ophthalmologists to diagnose, assess, and monitor symptoms. During examination, ophthalmologists must identify the corneal ulcer area and evaluate its severity by manually comparing ocular staining images with severity indices. However, manual assessment is time-consuming and may provide inconsistent results. Variations can occur with repeated evaluations of the same images or with grading among different evaluators. To address this problem, we propose an automated corneal ulcer grading system for ocular staining images based on deep learning techniques and the Hough Circle Transform. The algorithm is structured into two components for cornea segmentation and corneal ulcer segmentation. Initially, we apply a deep learning method combined with the Hough Circle Transform to segment cornea areas. Subsequently, we develop the corneal ulcer segmentation model using deep learning methods. In this phase, the predicted cornea areas are utilized as masks for training the corneal ulcer segmentation models during the learning phase. Finally, this algorithm uses the results from these two components to determine two outputs: (1) the percentage of the ulcerated area on the cornea, and (2) the severity degree of the corneal ulcer based on the Type–Grade (TG) grading standard. These methodologies aim to enhance diagnostic efficiency across two key aspects: (1) ensuring consistency by delivering uniform and dependable results, and (2) enhancing robustness by effectively handling variations in eye size. In this research, our proposed method is evaluated using the SUSTech-SYSU public dataset, achieving an Intersection over Union of 89.23% for cornea segmentation and 82.94% for corneal ulcer segmentation, along with a Mean Absolute Error of 2.51% for determining the percentage of the ulcerated area on the cornea and an Accuracy of 86.15% for severity grading.

Keywords: corneal ulcer; convolutional neural network; deep learning; Hough transform; image processing; image segmentation; U-Net

1. Introduction

Corneal ulcer is an ocular condition associated with eye diseases that can result from various causes, including inflammation, infection, and injury [1]. Corneal ulcers can be categorized into three main types. The first type, point-like corneal ulcer, is characterized by small scar dots distributed on the corneal surface [2]. The severity of this condition correlates with the number of dots on the corneal area. One of the causes of this type is dry eye syndrome [3]. The second type, flaky corneal ulcer, consists of scar areas on the corneal surface. The severity is determined by the size and location of the corneal ulcer [2]. This type can be caused by many factors, such as bacterial infection or corneal surgery. The last type, mixed point–flaky corneal ulcer, has both small scar dots and scar patches on the cornea.

In diagnosing the severity of corneal ulcers, ophthalmologists commonly use fluorescein dye on the ocular surface. The corneal ulcer becomes stained green, while the non-ulcer area remains unstained. Subsequently, clinicians examine and evaluate the corneal ulcer severity by analyzing images captured through the slit lamp microscope by themselves [3–5]. However, manual grading is a time-consuming task and can yield inconsistent results depending on the evaluator [6]. Therefore, developing a reliable end-to-end algorithm for grading the severity of corneal ulcers can significantly improve treatment outcomes and save valuable time for ophthalmologists.

In this study, we propose an automated grading system for flaky corneal ulcers based on their size and location. The algorithm comprises three main components. First, we employ deep learning models to segment the cornea area from eye staining images and apply the Hough Circle Transform (HCT) to refine the result area into a circular shape. The corneal area identified in this process can be used in the corneal ulcer segmentation and severity grading process. Second, the corneal ulcer area is segmented using deep learning techniques. In this process, the novelty of our approach lies in using the predicted corneal area from HCT as additional training data during the learning phase to develop the corneal ulcer segmentation models. This method benefits our models by guiding the weight adjustments to focus on the corneal areas. Third, the severity of the corneal ulcer is graded based on two properties: the percentage of the ulcerated area on the cornea, and the ulcer's distribution across five corneal zones.

2. Related Work

Traditionally, the severity of corneal ulcers has been assessed by ophthalmologists through manual grading. For instance, the Oxford Grading System (OGS) is widely used as a standard for evaluating point-like corneal ulcers. Specialists need to compare the corneal ulcer area with the OGS index, which is divided into five grades, for manual assessment of symptoms [3,7]. This method may yield imprecise diagnostic results, as patients with varying symptoms could be categorized within the same severity grade. In addition, to assess the severity of flaky corneal ulcers, ophthalmologists use specialized software such as Image-J [5] to locate or count the pixels in the corneal ulcer region for diagnosis. These manual processes are notably time-consuming and may lead to inconsistencies in subsequent diagnostic results.

In response to these challenges, Ayşe Bağbaba et al. [3] proposed an automated grading system for Dry Eye Syndrome (DES) based on the OGS using fluorescein-stained images. Their method involved using the HCT for corneal area segmentation and utilizing the green channel to identify damaged areas caused by DES. They applied a connected component algorithm to count dots in the corneal area and determined a correlation between the number of dots and DES severity using linear regression. However, this technique is only suitable for point-like corneal ulcers, not flaky corneal ulcers.

There has been increasing interest in the development of automated segmentation algorithms for identifying flaky corneal ulcer areas. Zhenrong Liu et al. [8] utilized the Gaussian Mixture Model (GMM) and Otsu thresholding techniques to segment corneal ulcers. They used 150 fluorescein-stained corneal ulcer images from Zhongshan Ophthalmic Center, Sun Yat-sen University. Their algorithm first converted these images to grayscale and applied gamma correction, Otsu binarization, and morphological operations to obtain the segmentation result. Additionally, they transformed the same images into a Hue Saturation Value (HSV) color space and utilized the GMM to segment the area of interest. Finally, they combined the results from both methods.

In fluorescein-stained eye images, fluorescent staining outside the corneal area can lead to incorrect localization of the corneal ulcer area. To enhance the accuracy of identifying the corneal ulcer area, it is crucial to first establish a reliable delineation of the cornea area. In research on iris segmentation, Juš Lozej et al. [9] identified the iris area in eye images using the U-Net model, emphasizing performance enhancement through adjustments to depth and batch normalization layers. On the other hand, Anis Farihan Mat Raffei et al. [10]

implemented a corneal detection algorithm using edge detection and HCT to locate the cornea and pupil. Additionally, they employed the Linear Hough Transform to locate the eyelid boundary to specify the iris area.

Lijie Deng et al. [5] proposed a superpixel-based method for corneal ulcer segmentation using 150 fluorescein-stained images from Zhongshan Ophthalmic Center. Their methodology involved employing an iris locating algorithm from previous research and applying the coordinates to an ellipse equation. Corneal ulcer segmentation was achieved using a superpixel segmentation technique and a Support Vector Machine (SVM) classifier, followed by morphological operations to enhance system performance. Isam Abu Qasmieh et al. [11] applied image processing and eye-border recognition to locate the eye border in the image. They determined the iris region by fitting a circle into the eye border area. Finally, they located the corneal ulcer area by extracting the green channel from the image, which was masked by the iris region.

Recently, deep learning techniques have been employed in eye disease for classification tasks [12–14] and segmentation tasks [4,11,15,16]. Qichao Sun et al. [15] manually selected the cornea area and trained a Deep Neural Network (DNN) model to segment the corneal ulcer area, selecting the result from the intersection area. Tingting Wang et al. proposed two studies focused on end-to-end corneal ulcer segmentation, utilizing the SUSTech-SYSU dataset [2] for training and testing. The first work, CU-SegNet [16], utilized the Multiscale Adaptive-aware Deformation Module (MAD) and the Multiscale Global Pyramid Feature Aggregation Module (MGPA) to reduce the influence of large variations in the morphology and size of ulcers and guide the model towards multiscale deformation. In a subsequent study [4], they utilized the Multiscale Self-Transform Network (MsSTNet) for corneal ulcer segmentation. They enhanced performance by feeding the results into a Generative Adversarial Network (GAN) discriminator to predict whether images were real or fake during training. Isam Abu Qasmieh et al. [11] selected ResNet-18 to develop a corneal ulcer segmentation model. The model was trained and evaluated using 354 fluorescein-stained images from the SUSTech-SYSU dataset.

In the context of our study, we propose a novel end-to-end algorithm for identifying the severity degree of corneal ulcers in fluorescein-stained images. This algorithm leverages a combination of techniques:

1. Cornea Segmentation Module: This component is developed through the integration of deep learning techniques, image processing, and the HCT.
2. Corneal Ulcer Segmentation Module: This module is developed using deep learning methodologies, utilizing the predicted cornea mask during the model learning phase.

The utilization of these two components is fundamental to our approach to accurately determining two grading results:

1. The percentage of the ulcerated area on the cornea.
2. The severity degree based on the ulcer's distribution across five corneal zones.

3. Methods

This section explains the methods and techniques applied in this research to develop the proposed automated corneal ulcer grading system. The entire contents can be separated into three sections: (1) cornea segmentation, (2) corneal ulcer segmentation, and (3) severity grading. An overview of this research is illustrated in Figure 1.

3.1. Cornea Segmentation

3.1.1. Simple CNN Encoder–Decoder Model

To segment the cornea area in ocular-stained images, a Convolutional Neural Network (CNN) model is an effective choice [17,18]. The boundary of the cornea region is circular, which is a simple shape for image segmentation tasks. Therefore, we utilize the Simple CNN Encoder–Decoder (SCED) network to address this problem [19]. The SCED architecture comprises two crucial paths: (1) the encoder network and (2) the decoder network. There is

no skip-connection from the encoder network to the decoder network in this architecture. The encoder network consists of convolutional layers and pooling layers. Convolutional layers extract the feature maps from the input image, while pooling layers reduce the feature maps' spatial resolution. These processes are repeated according to the depth (*n*) of the designed architecture. In the decoder network, upsampling layers increase the spatial resolution of the feature maps, while convolutional layers refine the features and extract more details. Finally, the feature maps are passed to the output layer and the activation function in the output layer assigns labels to each pixel to generate the output image. The SCED model architecture for cornea segmentation is shown in Figure 2.
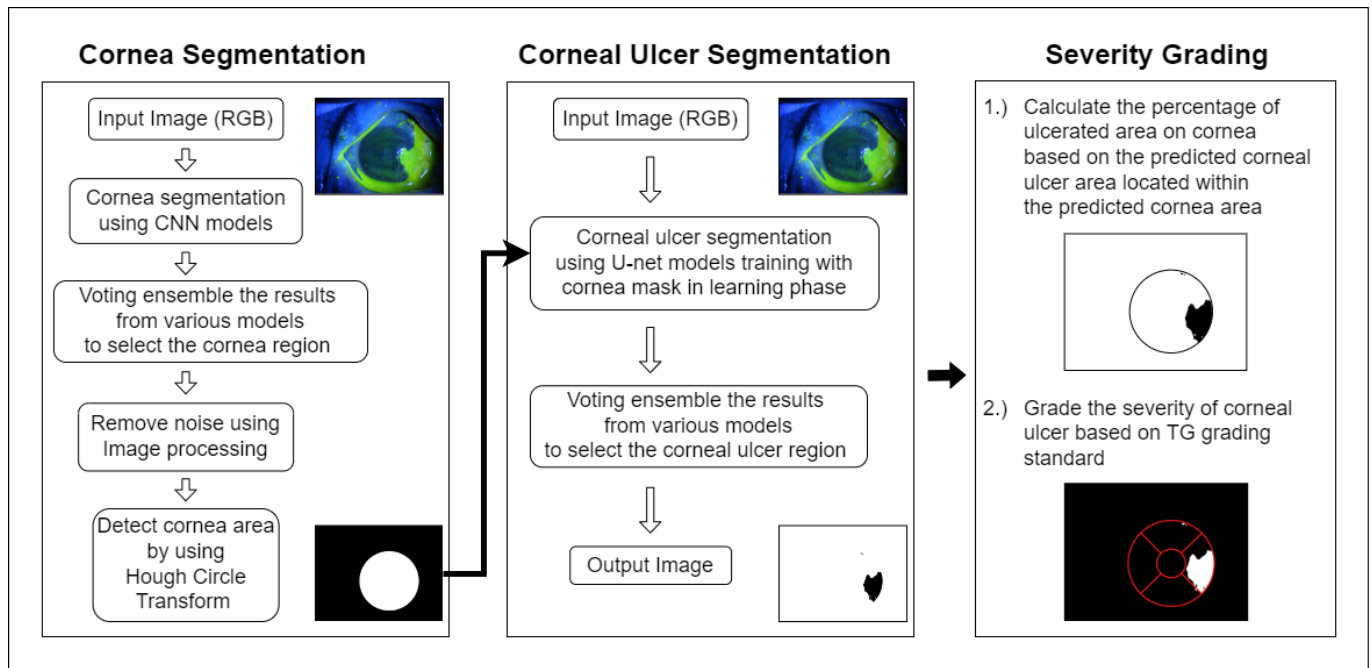

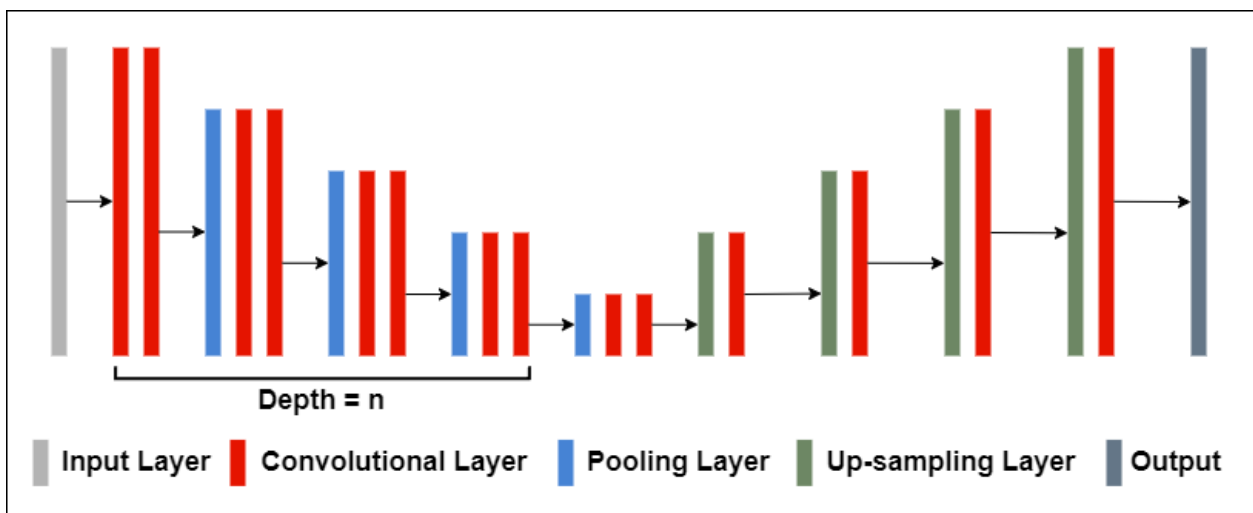**Figure 1.** Overview of the proposed method.

Figure 2. Simple CNN Encoder–Decoder network architecture.

3.1.2. Image Processing

The outputs of the SCED segmentation model are expressed in grayscale images, with values ranging from 0 to 1 for each pixel. To segment the cornea area, we apply image thresholding to transform grayscale images into binary images. The pixels segmented as 1 represent the cornea area, while the pixels segmented as 0 denote non-cornea areas.

However, there is an issue with noise pixels, which are predicted as 1 for the non-cornea area at the edges of the predicted cornea region. To address this concern, we utilize image processing techniques to minimize noise in the model outcomes. Morphological operations are employed on the predicted cornea area in this work. These operations aim to modify the structure of the input images by utilizing a kernel to interact with the objects within the image. The basic operations used in this context consist of erosion and dilation to respectively decrease and increase the target area. The variation from adjustments on the output region is based on the size and shape of the kernel defined in the operations [20,21].

Another approach to addressing noise problem involves utilizing the connected component labeling technique. This method aims to detect the group of connected pixels defined as the same label within a binary image. The algorithm scans the image pixel-by-pixel to determine the connectivity between each pixel and its neighboring pixels. Common connectivity types used to identify the connected components include 4-connectivity and 8-connectivity. A group of connected pixels is counted as a component [22,23]. In this experiment, we apply the connected component algorithm to detect all components from morphological operation outputs and remove artifact regions that are smaller than a specific size.

3.1.3. Hough Circle Transform

The HCT is an adaptation of the Hough Transform algorithm. This technique is used to detect circular objects in images [24,25]. The algorithm utilizes edge detection to identify potential edges within the image and evaluates all possible coordinates that are able to fit the center and radius of the circles. In general, the cornea and iris are different structural components; however, they share the same boundary in the front portion of the eye [26]. The characteristics of both components are almost circular in shape; therefore, the HCT is widely used for iris segmentation tasks and yields excellent results [3,10,27–29]. In this research, we employ the HCT with the output images from the image processing module to refine the segmented cornea area into a circular shape. The algorithm detects all circles in the input images and stores them in a circle list. Then, the circles with low confidence scores are eliminated by thresholding. The remaining circles are stored with their center coordinates and radius values as the cornea candidates. Finally, we select a circle C that maximizes our proposed objective function $F_\alpha(C)$ to represent the cornea area, as shown in Figure 3. The proposed objective function is defined as follows:

$$F_\alpha(C) = \frac{N_C/M}{(1 + (r_C/W))^\alpha} \tag{1}$$

where N_C denotes the cornea predicted area that intersects with the circle from HCT, M denotes the cornea predicted area, r_C denotes the radius of the circle from the HCT, W denotes the input image width, and α denotes the fine-tuning parameter.

3.1.4. Voting Ensemble

To solve a specific problem using machine learning, multiple individual models are usually developed in order to determine the most appropriate method for the task. Instead of relying on a prediction from one model, a voting ensemble can be used to combine the results of all models and make a final decision. The voting ensemble technique can be applied to improve performance, particularly in regression and classification tasks. There are various implementations of this method, such as hard voting, which make a final prediction based on the majority vote, and soft voting, which uses average of probabilities from all individual models [30,31]. In this research, we apply the concept of a voting ensemble to enhance segmentation performance. All corresponding predicted images of each method are combined through the hard voting ensemble approach. The resulting area from this technique is determined using the pixels voted by more than half of all corresponding models in each method. This methodology is then applied to both cornea segmentation and corneal ulcer segmentation.
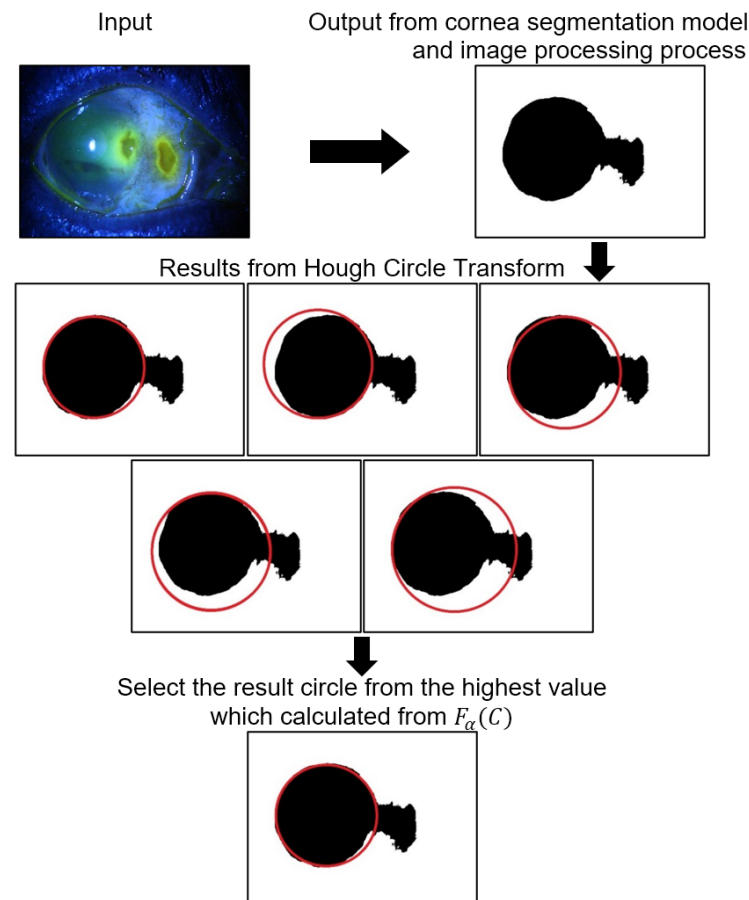
Figure 3. HCT and circle selection for cornea segmentation.

3.2. Corneal Ulcer Segmentation

Corneal ulcer segmentation demands a highly precise model due to the rich details on ocular staining images. To achieve accurate results, we adopt the U-Net model as the base model architecture for our proposed method. The U-Net model is a CNN encoder–decoder network specifically designed for image segmentation tasks, and is widely used in biomedical image analysis [9,32–36]. In terms of architecture, the U-Net model diverges from the SCED model primarily due to its implementation of skip connections. The skip connections in U-Net connect the convolutional layers in the encoder and decoder networks, operating on the corresponding feature maps. This integration aims to retain data resolution and improve segmentation performance [37]. During the experiment, the U-Net model is initially trained to identify the corneal ulcer regions. The model outputs are multiplied by the raw cornea segmentation result in the inference phase. This method serves as the baseline for our work.

For the proposed model, we modify U-Net by including the cornea segmentation result from the HCT as a second input. The final model layer is multiplied by the predicted cornea mask in the learning phase. This approach benefits our corneal ulcer segmentation model by directing the model's attention to learning features within specific regions and guiding weight adjustments to the cornea areas. The proposed model architecture for corneal ulcer segmentation is shown in Figure 4.
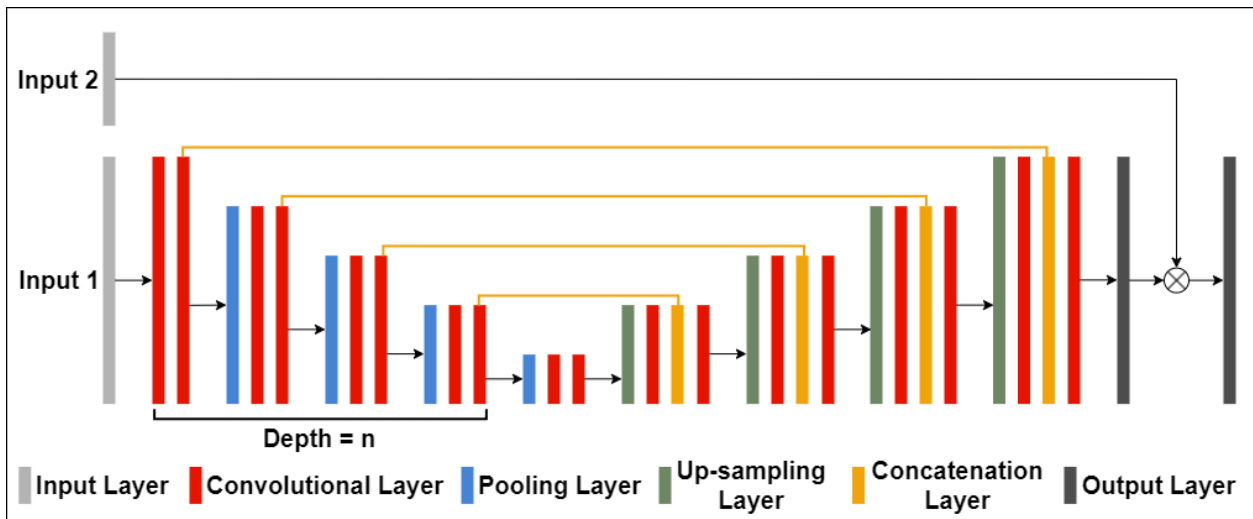
Figure 4. Model architecture of U-Net with HCT mask multiplication in the learning phase (UNET-H-L).

3.3. Severity Grading

There are several methods for identifying the severity of ocular staining images [38]. In this work, we utilize the same ulcer position-based severity grading as in the Type–Grade (TG) grading system, an accepted method for assessing the severity of corneal ulcers based on the stained area within the cornea region [2,39–41]. This method categorizes the severity of corneal ulcers into five grades (Grade0-4). The cornea area is divided into five zones, comprising four surrounding quadrants and one central area, as shown in Figure 5. The threshold used for grading in this experiment is one pixel. The criteria for each grade are as follows [2]:

- Grade4: Ulceration is found in the central area of the cornea.
- Grade3: Ulceration is found in three or all four quadrants surrounding the central area.
- Grade2: Ulceration is found in two quadrants surrounding the central area.
- Grade1: Ulceration is found in one quadrant surrounding the central area.
- Grade0: The cornea area is completely clear with no ulcerated area present.

According to the definition, the classification is based on the number of quadrants where ulcers are present and whether the ulcer appears in the pupil area or the central zone of the cornea, which is critical for vision. In such cases, the severity is graded as the highest degree (Grade4).
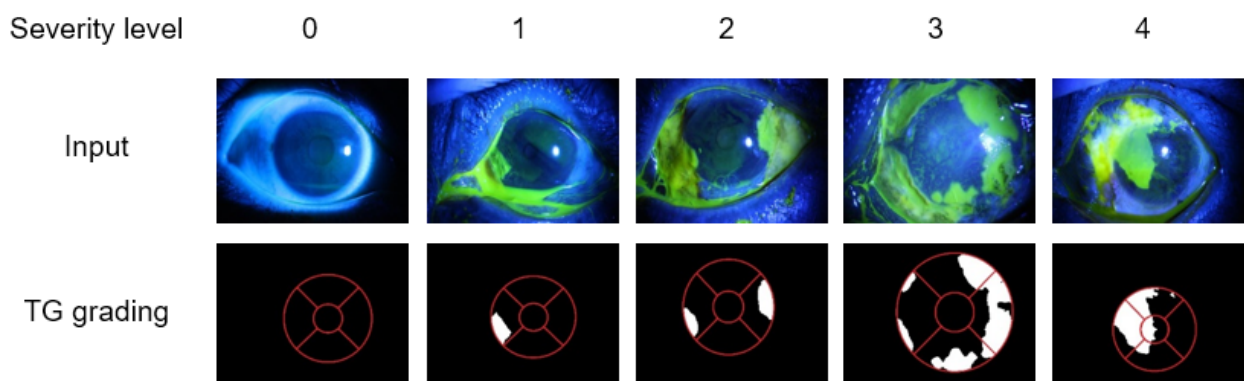

Figure 5. Samples of corneal ulcer severity grading based on the TG grading system.

However, this methodology categorizes symptoms into discrete degrees of severity, meaning that the results from this process may lead to inaccuracies in assessing the actual severity of the symptoms. Therefore, we include the percentage of the ulcerated area on the

cornea in order to support the severity grading of the patient's condition. The calculation can be performed as follows:

$$P = \frac{A_U}{A_C} \times 100 \tag{2}$$

where P denotes the percentage of ulcerated area on the cornea, A_U denotes the number of pixels in the predicted corneal ulcer area, and A_C denotes the number of pixels in the predicted cornea area.

These methodologies offer multiple benefits for measuring symptoms in terms of several aspects:

1. Consistency: Utilizing software for grading the severity of corneal ulcers yields more consistent outcomes compared with manual grading. This is because the program consistently provides the same outcome on the same image, whereas manual measurement may result in different values each time or for different individuals.
2. Robustness: Variations in image size or cornea size have minimal impact on severity grading when using software.
3. Resolution: Including the calculation of the percentage of the ulcerated area on the cornea can provide better resolution than relying solely on discrete severity levels.

4. Experiments and Results

4.1. Dataset

For our experiments, we utilized ocular staining images from the SUSTech-SYSU public dataset [2]. This dataset was separated into two parts for our work, as shown in Figure 6. First, 712 ocular staining images with cornea-labeled images were used in the cornea segmentation algorithm. Image augmentation was applied to prevent overfitting by adjusting the saturation, brightness, and blurring as well as by vertically flipping the input images. Second, within the 712 ocular staining images, there were 354 images in the dataset labeled as containing corneal ulcer and showing flaky corneal ulcer symptoms. These images were utilized for corneal ulcer segmentation in the experiment. Data augmentation was implemented with the following parameters: vertical shifting, horizontal shifting, distortion, zooming, vertical flipping, horizontal flipping, brightness, rotation, and noise addition.

To use these data in the experiment, the original images were resized to 512×512 for training the models and 1600×1200 for the other operations, which included the image processing and HCT process for cornea detection, the severity grading process, and all evaluation processes. Additionally, the dataset was randomly divided into three folds for validation during training and evaluation.
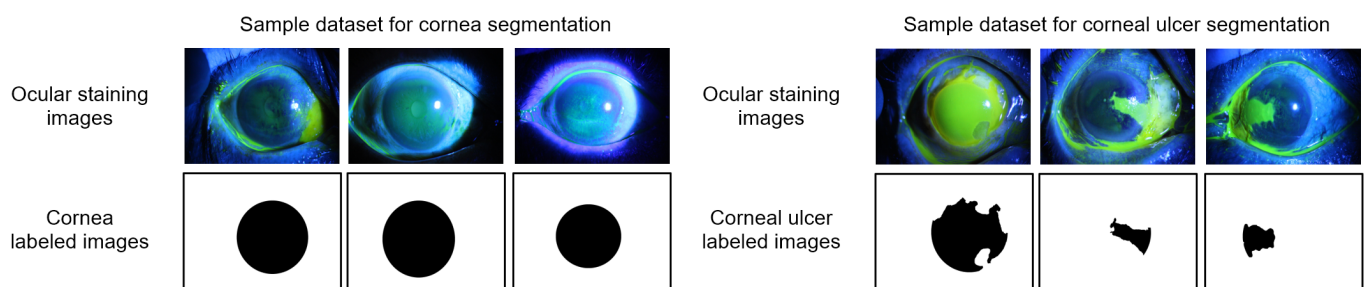
Figure 6. Sample dataset for training and testing cornea segmentation models and corneal ulcer segmentation models.

4.2. Evaluation Metrics

To evaluate the algorithm's performance, we calculated the Accuracy, Dice Similarity Coefficient (DSC), Sensitivity (Sen), and Intersection over Union (IoU) for pixelwise comparison of the cornea and corneal ulcer segmentation results with the ground truth data. In addition, the Mean Absolute Percentage Error (MAPE) was calculated to compare the

predicted size of the cornea to the actual cornea size in pixels, while the Mean Absolute Error (MAE) was calculated to compare the predicted percentage of the ulcerated cornea area to the percentage of the actual sizes. For TG grading results, Accuracy was calculated as the ratio of correct predictions for corneal ulcer severity grading to all predictions. The evaluation metrics were calculated as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{3}$$

$$Sen = \frac{TP}{TP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|A_{Ci} - \hat{A}_{Ci}|}{A_{Ci}} \times 100 \tag{6}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |P_i - \hat{P}_i| \tag{7}$$

$$Segmentation\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Grading\ Accuracy = \frac{\hat{S}}{S} \tag{9}$$

where TP denotes the number of cornea/corneal-ulcer pixels which are classified correctly, TN denotes the number of non-cornea/non-corneal-ulcer pixels which are classified correctly, FP denotes the number of non-cornea/non-corneal-ulcer pixels which are classified as cornea/corneal-ulcer, FN denotes the number of cornea/corneal-ulcer pixels which are classified as non-cornea/non-corneal-ulcer, A_{Ci} and $\hat{A}_{Ci}$ respectively denote the actual cornea size of sample i and the corresponding predicted cornea size, N denotes the number of samples in the experiment, P_i and $\hat{P}_i$ respectively denote the actual percentage of ulcerated area on cornea for sample i and the corresponding predicted percentage, and $\hat{S}$ and S respectively denote the number of correct severity grading results and the total number of samples.

4.3. Implementation Details

This experiment was implemented on Lenovo Legion 5 (model 15IAH7H) laptop, manufactured in China, with an I7-12700H processor, 16 GB of RAM, and an Nvidia RTX-3060 GPU with 6 GB of RAM. The environment for training and testing all models was based on Python 3.8.13 and Tensorflow 2.5.0. The learning rate and batch size in this work were set to 0.0001 and 1, respectively. For evaluating resource usage and training time, the experiment was conducted on Google Colab using cloud computing resources with an A100 GPU.

4.4. Cornea Segmentation Results

For the first part of this experiment, we built SCED networks for segmenting the cornea area by using corneal fluorescein-stained images as inputs. Five SCED networks with different configurations were tested, as shown in Table 1. For the encoder network, the input size was set to $512 \times 512 \times 3$. The feature map for the convolutional layer in the encoder network was set to 64, followed by the Rectified Linear Unit (ReLU) activation function and max pooling with a stride of 2×2. The kernel size of the convolutional layer varied from 7 to 11 and the depth of the CNN layers varied from 3 to 5. Dropout layers were included to prevent overfitting. For the decoder network, the upsampling size was 2×2, while the other parameters were defined the same as in the encoder network. The decoder's output layer consisted of one feature map with a sigmoid activation function.

The model depth and other parameters were referenced from the work by Juš Lozej et al. [9]. In this experiment, aiming to obtain a smooth shape of the predicted cornea area, we increased the kernel size to expand the receptive field. The differences among the models' architectures are as follows: SCED1-3 used a kernel size of 9 for all models and varying model depths of 3, 4, and 5, respectively; SCED4 increased the kernel size to 11 with a model depth of 3; and SCED5 employed different kernel sizes in each layer. In the encoder network, the kernel size for depth layers 1 and 2 were configured as 11, layers 3 and 4 were configured as 9, and layer 5 was configured as 7, maintaining this setup in the corresponding layers of the decoder network. The memory (VRAM) and training time per epoch required for training each model are also presented in Table 1.

To benchmark cornea segmentation performance, we implemented two U-Net models. The first U-Net model followed the original design proposed by Ronneberger et al. [32], while the second used transfer learning by combining the pretrained ResNet-50 encoder [42] with U-Net's decoder structure. The details of U-Net and U-Net–ResNet-50 are provided in Table 1.

Table 1. Cornea segmentation model architectures and resources required for training.

Model Name	Feature Maps	Kernel Size	Depth	Trainable Parameters	Memory Usage [GB]	Training Time per Epoch [s/Epochs]
SCED1	64	9	3	2,338,561	8.7	63
SCED2	64	9	4	3,332,241	8.8	67
SCED3	64	9	5	3,997,761	8.9	69
SCED4	64	11	3	3,493,121	8.8	74
SCED5	64	11, 11, 9, 9, 7	4	4,029,889	8.9	77
U-Net	64, 128, 256, 512, 1024	3	4	36,605,317	21.0	96
U-Net-ResNet-50	64, 128, 256, 512, 1024, 2048	1, 3, 7 *	5 **	73,378,625	11.3	110

* The kernel size of the initial convolutional layer was 7; for blocks 2 to 5, the convolutional layers used kernel sizes of 1-3-1 for each block. ** U-Net–ResNet-50 included 54 convolutional layers distributed across five depth levels.

The training and validation accuracy and loss curves for the cornea segmentation models, illustrated in Figure 7, provide crucial insights into the models' learning process and performance. These graphs display the accuracy and loss metrics for both training and validation sets throughout the training epochs. The data presented in these graphs represent the first fold of a three-fold cross-validation process for all models. The plots reveal that all SCED models reach convergence within 30 to 40 epochs. Among the SCED model variants, the SCED3 model stands out by converging the fastest and achieving the highest accuracy on the validation set. Conversely, U-Net and U-Net–ResNet-50 show signs of overfitting, as evidenced by the discrepancy between training and validation performance. This can be observed in the accuracy and loss curves, where the training set accuracy continues to improve while the validation loss progressively increases from the early epochs. In this study, model selection is based on the criterion of minimal validation loss observed during the training process.

As mentioned in Section 3, the raw outputs from SCED networks are postprocessed by image processing techniques and the HCT. In the experiment, we name the SCED networks with this postprocessing the Simple CNN Encoder–Decoder Model with Hough Circle Transform (SCED-HT). First, we utilized morphological erosion to separate excess pixels from the main cornea predicted area using a circular kernel with a radius of 4. Then, the connected component technique was applied to remove noise areas smaller than 400 pixels. After obtaining the desired area, the cornea region was expanded back to the original size using morphological dilation with the previous circular kernel. Subsequently, we applied HCT to locate the entire circle under the specified parameters. The resolution of the HCT accumulator was 1600×1200 pixels, the minimum distance for the center of each detected circle was set to 40 pixels, and the radius of the detected circles was set to 120–1600 pixels.
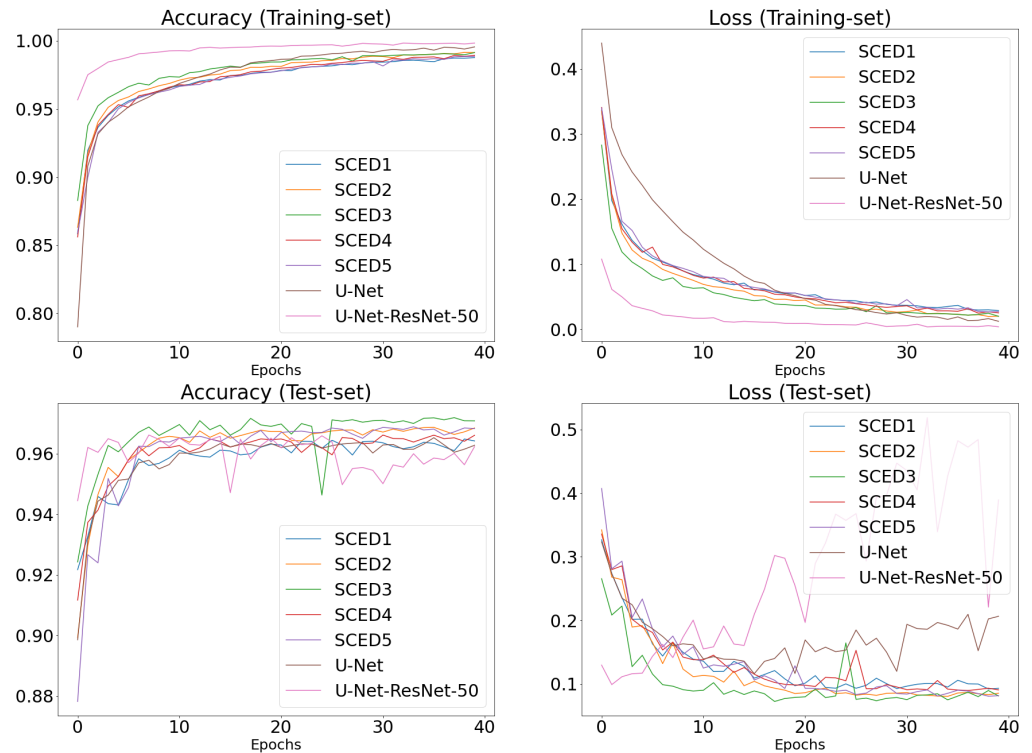
Figure 7. Training and validation accuracy and loss curves for cornea segmentation models.

All circles obtained from this process were evaluated using the objective function $F_\alpha(C)$ mentioned in Equation (1). The circle that maximizes $F_\alpha(C)$ was selected to represent the cornea area. In the experiment, we determined the best hyperparameter α by varying the values of α between 4.6 to 5.8 and calculating the IoU of the detected cornea areas for each α value. We first identified the optimal value of α using 120 additional ocular staining images that were not included in the SUSTech-SYSU dataset. An α value of 5.2 achieved the highest IoU for cornea segmentation. These α values were also verified with 712 images from the SUSTech-SYSU dataset, and provided the same results. The relation between the IoU and α is shown in Figure 8. The optimal α value of 5.2 was used for the entire experiment.

To evaluate the performance of the objective function $F_\alpha(C)$ for circle selection to represent the cornea regions, we calculated the percentage of the selected circles based on their IoU rankings among the candidates. From the graph in Figure 9, the circle which maximized $F_\alpha(C)$ is 67.79% of the circles which obtained the highest IoU from all candidates; 84.53% of the selected circles were among the two highest-ranked candidates, and 92.27% were within the top three highest-ranked candidates.

To assess the segmentation performance, the SCED and SCED-HT models were evaluated based on accuracy, IoU, and MAPE of cornea size. Table 2 displays the results from all models in this experiment. The values following MAPE represent the standard deviation of MAPE, indicating the variability of the error in predicted cornea area around the mean in percentage. From the results, the SCED3 model has the highest accuracy and IoU compared to other methods, with 96.97% and 90.21%, respectively. The SCED2 model has the lowest MAPE at 7.28%. Overall, SCED models perform slightly better than SCED-HT models with the same architecture. The samples of cornea segmentation results for our work are illustrated in Figure 10.
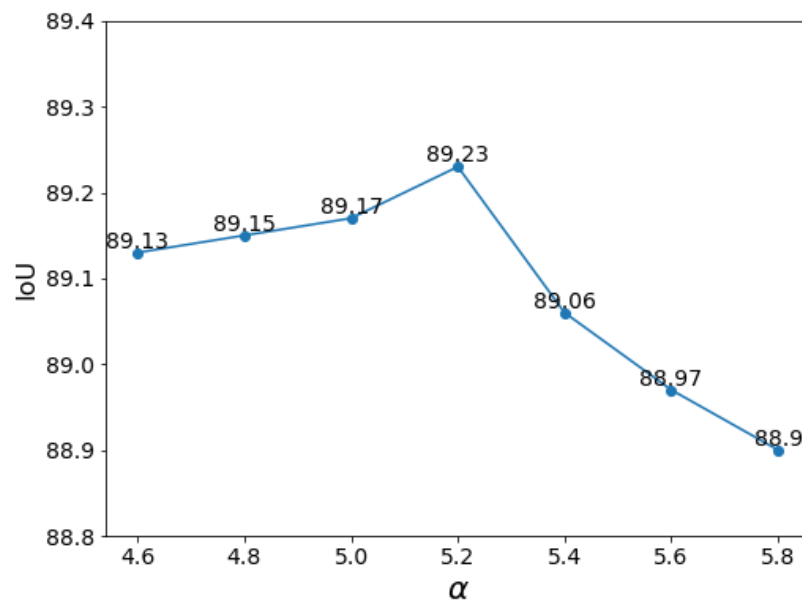
Figure 8. Cornea segmentation IoU from α values in the proposed objective function (Equation (1)).
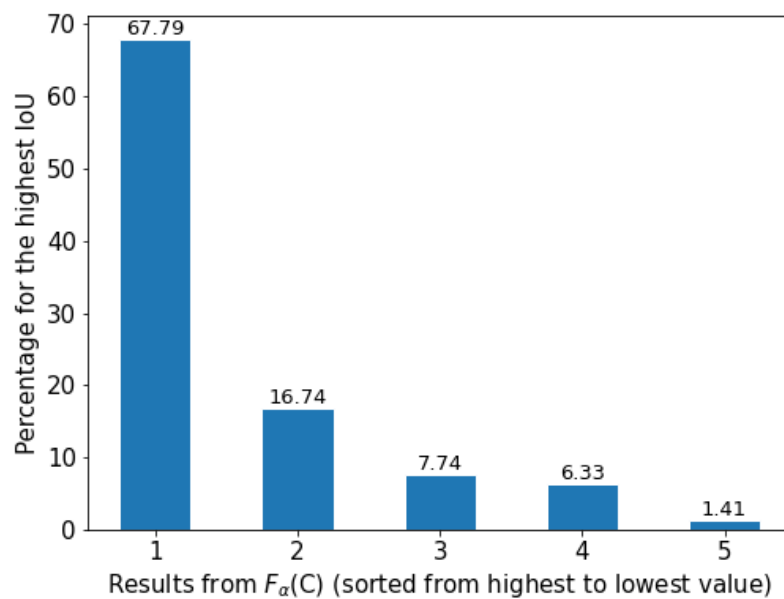

Figure 9. The percentage for the highest IoU value of the top five candidate circles calculated from the objective function (Equation (1)).

Table 2. Evaluation results for the individual cornea segmentation models.

Method	Accuracy	Micro IoU	MAPE (Cornea Size)
SCED1	96.32	88.29	8.02 ± 13.22
SCED2	96.74	89.50	7.28 ± 11.25
SCED3	96.97	90.21	7.45 ± 12.48
SCED4	96.39	88.50	8.14 ± 13.07
SCED5	96.62	89.18	7.55 ± 11.19
SCED1-HT	96.40	88.37	9.11 ± 14.11
SCED2-HT	96.36	88.18	9.06 ± 13.54
SCED3-HT	96.49	88.56	8.58 ± 13.28
SCED4-HT	96.39	88.31	9.09 ± 15.44
SCED5-HT	96.48	88.56	8.90 ± 12.67
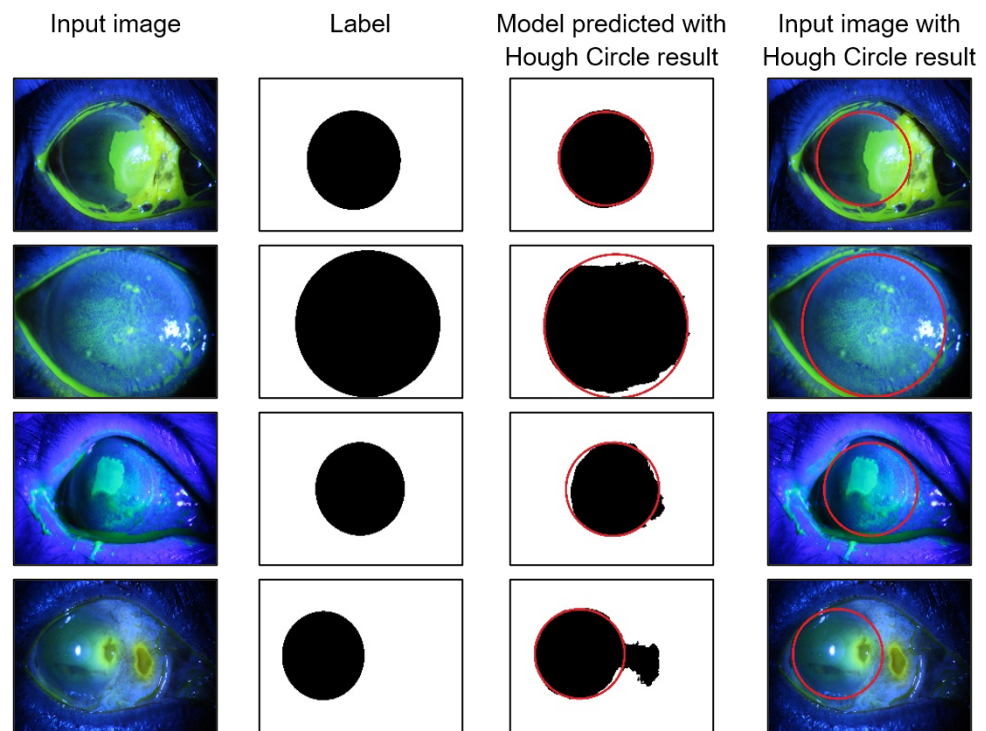
Figure 10. Sample results from the cornea segmentation algorithm.

Finally, we combined all segmentation results from all SCED networks using a voting ensemble. The results from the ensemble models are shown in Table 3. The EM-5-SCED is an ensemble model of SCED1 to SCED5, while the EM-5-SCED-HT is an ensemble model of SCED1-HT to SCED5-HT. The results from the ensemble models outperform their individual models. The EM-5-SCED has the best performance for cornea segmentation in this experiment, achieving 97.08% accuracy, 90.58% IoU in micro-average, and 6.93% MAPE for cornea size. For the EM-5-SCED-HT results, the accuracy and IoU are comparable to EM-5-SCED but cannot yield a lower MAPE value, as EM-5-SCED-HT employs the HCT to identify the cornea area based on the segmentation results from EM-5-SCED. When EM-5-SCED generates inaccurate cornea shapes but still closely approximates cornea sizes, employing the HCT to determine the cornea region can result in incorrect areas, leading to errors in cornea size.

To evaluate the efficacy of our proposed method, we conducted a comprehensive comparative analysis using both image processing and deep learning approaches for cornea segmentation. The first comparison focused on the image processing method developed by Isam Abu Qasmieh et al. [11], who proposed techniques for cornea and corneal ulcer segmentation. We implemented the process as described in their study to extract the cornea area for direct comparison with our method. The image processing method utilizes the Otsu thresholding algorithm [43] and morphological operators to determine the eye border pixels. Then, the upper and lower eye border pixels are fitted with Gielis curves [44]. Finally, the cornea region is obtained from the enclosed circle formed by the upper and lower eye border curves. The second approach involved the deep learning-based SLIT-Net model developed by Jessica Loo et al. [45]. This model, based on the Mask R-CNN architecture [46], is designed for ocular staining image segmentation, including the limbus (cornea region). In the experiment, we utilized the SLIT-Net model originally provided by the authors, which was trained with 133 slit-lamp images from the University of Michigan Kellogg Eye Center (Ann Arbor, MI, USA) and Aravind Eye Care System (Madurai, India) without additional training on the SUSTech-SYSU dataset to generate cornea segmentation results for evaluation. Both methods were tested using the same dataset employed in our study to ensure a consistent comparison. Additionally, we included the U-Net and U-Net–

ResNet-50 models in the analysis to establish a comprehensive baseline for performance comparison. The results reveal that the EM-5-SCED and EM-5-SCED-HT variants of our proposed method both achieve superior cornea segmentation performance compared to all other examined methods. This enhanced performance is evident across multiple metrics, including Accuracy, IoU, and MAPE in cornea size estimation, as presented in Table 3.

Table 3. Comparison of the cornea segmentation performance across various methods.

Method	Accuracy	Micro IoU	MAPE (Cornea Size)
Image Processing Method [11]	88.53	66.89	25.26 ± 34.34
SLIT-Net [45]	94.10	82.02	14.64 ± 20.54
U-Net	95.99	87.24	9.28 ± 16.17
U-Net-ResNet-50	95.95	86.85	9.74 ± 16.30
EM-5-SCED	97.08	90.58	6.93 ± 11.19
EM-5-SCED-HT	96.70	89.23	8.14 ± 11.71

Figure 11 illustrates the cornea segmentation results of the eight methods: Image Processing Method [11], SLIT-Net [45], U-Net, U-Net–ResNet-50, SCED3, SCED3-HT, EM-5-SCED, and EM-5-SCED-HT. These samples demonstrate that when the cornea boundary is clearly visible in the ocular staining image, all methods accurately segment the cornea area, as evidenced in samples (a), (b), and (c). However, when the cornea boundary is unclear, as in samples (e), (f), (g), and (h), the segmentation performance varies among the models. The U-Net–ResNet-50 model effectively segments clearly defined areas, but exhibits reduced performance in regions with complex and ambiguous image features. In contrast, the high-resolution U-Net model provides excessive detail, resulting in significant oversegmentation. The SLIT-Net and SCED models yield results with comparable shapes, although SLIT-Net demonstrates lower performance. Moreover, the SLIT-Net model significantly underperforms compared to the U-Net model. This discrepancy may be due to the differences in the training dataset. In this experiment, the U-Net model was trained and validated on the SUSTech-SYSU dataset, whereas the SLIT-Net model was trained on a different dataset and subsequently validated on the SUSTech-SYSU dataset for comparison. The SCED and EM-5-SCED models demonstrate superior performance in identifying the overall cornea area, which is suitable for subsequent circle detection tasks, despite some segmentation errors. The segmentation results observed in these sample images align with the performance comparisons presented in Table 3. Furthermore, we compared the segmentation results from the SCED model with the SCED-HT method, which converts the segmented area to a circular shape. When the segmentation results from SCED accurately represent the cornea areas or resemble a circular shape, as in samples (a), (b), and (c), the outputs of SCED and SCED-HT are similar. However, when the segmented results are approximately circular but include small excess areas, as in samples (d) and (e), SCED-HT can remove excess areas and identify more accurate cornea regions compared to SCED. In cases such as samples (f) and (g), where the segmented areas are irregularly shaped, using the segmentation results from SCED directly may yield better outcomes than applying the HCT to determine the boundary of the cornea region.

Finally, we compared the cornea segmentation results from the proposed method with other methods, as shown in Table 4. The details of each method are as follows:

1. CANNY and RANSAC: In this method, the original ocular staining images were first converted into grayscale, then Canny Edge Detection [47] was applied to identify the cornea edges. Next, RANdom SAmple Consensus (RANSAC) [48] was used to detect circles from the edge images.

2. CANNY and HCT: The same procedure as in the first method, except with the HCT used for circle detection. The circle with the highest accumulator value was selected as the output.

3. EM-5-SCED and RADON: First, the cornea segments were determined from the ocular staining images by using the EM-5-SCED model and extracting the contour of the

cornea segments. Then, Radon Transform-based circle detection was applied to detect the circles, as proposed by Okman et al. [49].

4. EM-5-SCED and RANSAC: The same procedure as the third method, except with RANSAC used for circle detection.

5. PROPOSED without $F_\alpha(C)$: Our proposed approach utilizes HCT to detect circles from the contour of the cornea segments. To demonstrate the effect of our proposed circle selection process, in this method the circle with the highest accumulator value is selected as the output instead of using the objective function $F_\alpha(C)$.

6. PROPOSED with $F_\alpha(C)$: This method represents the full version of the proposed cornea segmentation process, incorporating the EM-5-SCED model, HCT, and objective function $F_\alpha(C)$.
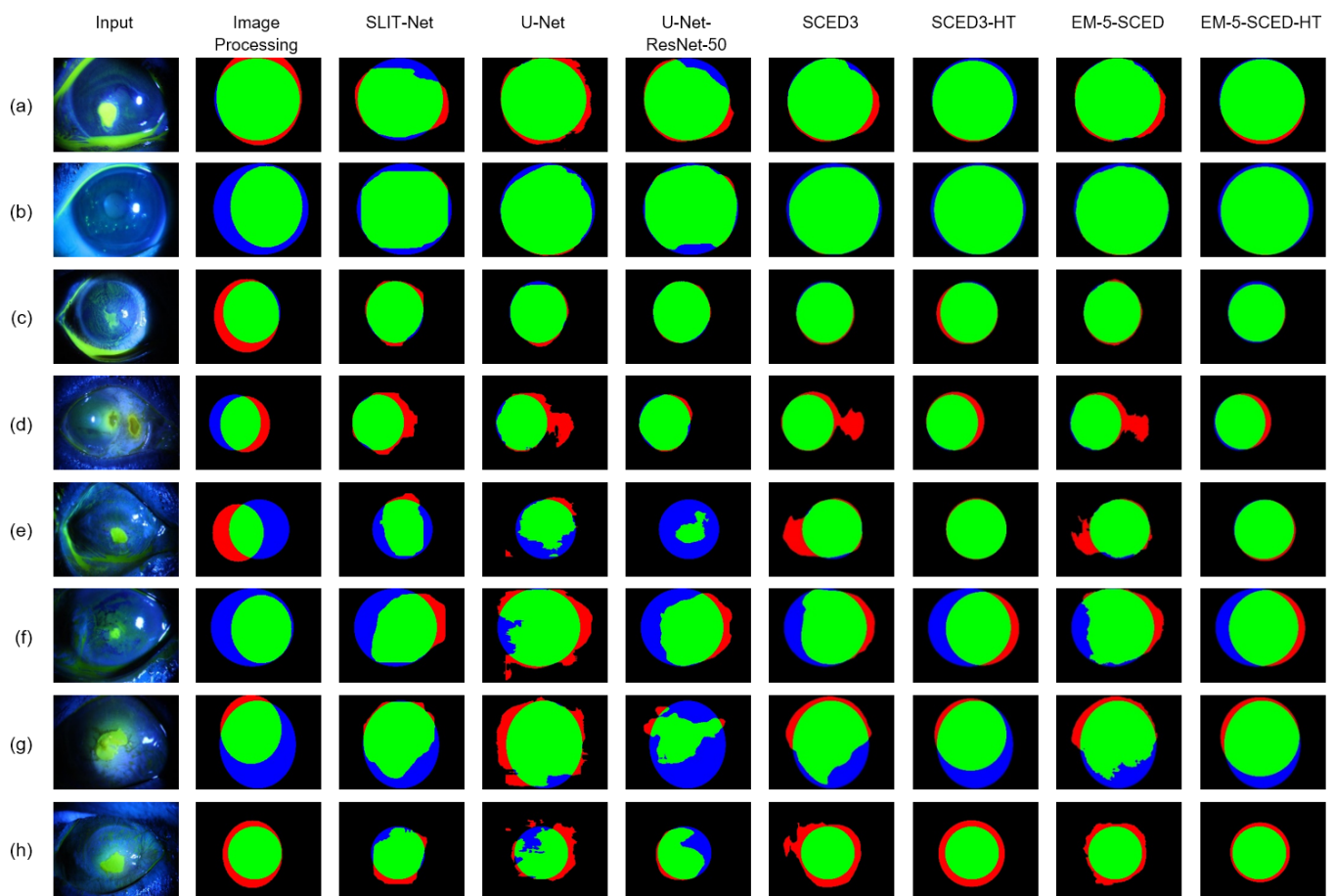
Figure 11. Comparison of cornea segmentation results. (**a**–**c**) are sample images with clearly visible cornea areas. (**d**–**h**) are sample images without clearly visible cornea areas. (Green = True Positive, Blue = False Negative, Red = False Positive)

Table 4. Comparison of various circle detection methods for the cornea segmentation process.

| Method	Accuracy	Micro IoU	MAPE (Cornea Size)
CANNY and RANSAC	79.76	57.45	77.19 ± 64.78
CANNY and HCT	80.03	58.23	76.09 ± 59.48
EM-5-SCED and RADON	95.13	84.30	11.62 ± 24.20
EM-5-SCED and RANSAC	96.48	88.75	9.64 ± 14.41
PROPOSED (EM-5-SCED and HCT) without $F_\alpha(C)$	96.21	87.77	9.73 ± 13.81
PROPOSED (EM-5-SCED and HCT) with $F_\alpha(C)$	96.70	89.23	8.14 ± 11.71

In the canny edge detection process in Methods 1 and 2, the grayscale image is first blurred using a Gaussian filter. Then, the image gradients are calculated for each pixel, and only pixels with local maxima of the gradient magnitude along the gradient direction are considered as edge candidates. Double thresholding is applied to classify candidate pixels into strong and weak edges using high and low threshold values, respectively. The final edge image is obtained from strong edges and weak edges that are connected to the strong edges. In this experiment, we varied the standard deviations of the Gaussian filter across 5, 10, 15, 20, and 25. Additionally, we varied the high and low threshold values across 6, 12, 18, 24, 30, and 36, pairing each lower value with its equivalent and all higher values to create a range of threshold combinations. We selected the optimal combination of Gaussian filter standard deviation and high and low threshold values that achieved the highest average precision in cornea edge detection. In this experiment, we identified the optimal parameters for the canny edge detection process as a Gaussian filter standard deviation of 10 with both the low and high thresholds set to 12.

For RANSAC-based circle detection, let P represent a set of pixels used for circle fitting, which includes either the edge pixels in Method 1 or the contour of the cornea segments in Method 4. The algorithm starts by randomly selecting three pixel positions from P and determining the circle C that best fits these points. The distance between every pixel in P and the circle C is calculated. The inlier pixels, which are those with distances less than the threshold, are counted and stored. In this experiment, the inlier threshold is set to 2. This process of random sampling, circle fitting, and inlier counting is repeated for a predetermined number of iterations T, which is calculated as follows [50]:

$$T = \frac{\log(1 - p)}{\log(1 - w^n)} \tag{10}$$

where p denotes the desired probability of success, w denotes the percentage of inliers in the dataset, and n denotes the number of points to define the model (for circle detection, $n = 3$). After T iterations are completed, the circle with the highest inlier count is returned as the final result. In this experiment, we set the desired probability of success p to 0.95. For Method 1, the percentage of inliers w was 1.87%, which leads to a number of iterations totaling 458,117 rounds. For Method 4, where the percentage of inliers w was 13.24%, we used 1289 iterations.

The results in Table 4 reveal that the proposed method provides more accurate results in cornea detection across all metrics compared to conventional methods. Examples of the circle detection process and results for each method are shown in Figure 12. The first two conventional methods, which apply canny edge detection directly to the original ocular staining images, provide low performance. This is because edge detection algorithms tend to detect many edges other than the cornea, especially the eyelid, which has a curve resembling a circular shape. These issues are illustrated in Figure 12, Method 1 and 2. Consequently, the RANSAC and HCT algorithms are unable to accurately detect the cornea area. The experimental results indicate that presegmentation of the corneal area prior to circle detection significantly improves cornea detection performance compared to direct application of circle detection algorithms on the original images. Additionally, the implementation of $F_\alpha(C)$ for circle selection with the HCT algorithm in the proposed method demonstrates a marked improvement in cornea detection performance. This refined approach outperforms the conventional method of selecting the circle based on the highest accumulated value, further validating the effectiveness of the proposed technique.
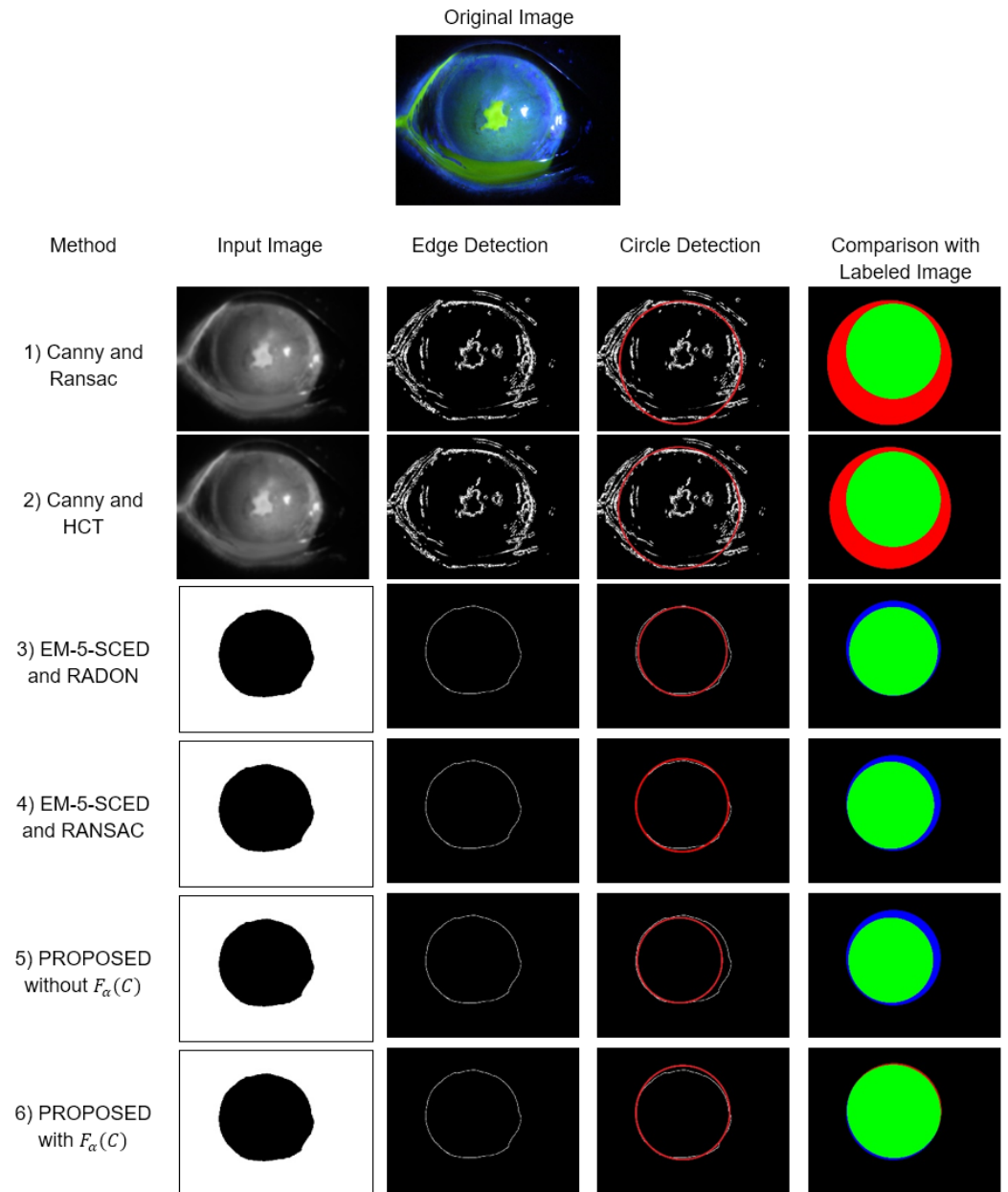
Figure 12. Sample of various circle detection methods for cornea segmentation. (Green = True Positive, Blue = False Negative, Red = False Positive)

4.5. Corneal Ulcer Segmentation Results

In the second part of the experiment, we built corneal ulcer segmentation models based on the original U-Net architecture [32]. The four model configurations used in the experiment are shown in Table 5. The input image size was $512 \times 512 \times 3$, the kernel sizes of the convolutional layers varied between 3 and 5, and ReLU was used as the activation function of the convolutional layers. The depth of the encoder and decoder networks in all models was equal to four layers. For the encoder networks, the number of feature maps in the first layer was set to 32 or 64 and then doubled in the next set of convolutional layers. A max pooling layer with a stride of 2×2 was applied to the encoders of every model. In the decoder networks, the number of feature maps in each convolutional layer was the reverse sequence of the number of feature maps in the encoder networks. The upsampling process utilized a Conv2DTranspose layer with a stride of 2×2. This setting was applied to the decoders of every model. A concatenation layer was included in the decoder networks to

combine the skip-connected feature maps from the encoder networks. The final layer was characterized by one feature map and a sigmoid activation function.

Table 5. Corneal ulcer segmentation model architecture.

Model	Feature Map	Kernel Size	Depth	Trainable Parameters
UNET1	32, 64, 128, 256, 512	5	4	21,706,949
UNET2	64, 128, 256, 512, 1024	5	4	86,811,013
UNET3	32, 64, 128, 256, 512	3	4	9,154,245
UNET4	64, 128, 256, 512, 1024	3	4	36,605,317

We compared three methods for corneal ulcer segmentation. The first method, named UNET-R-I, uses a conventional U-Net model with raw cornea mask multiplication in the inference phase. The model for segmenting the corneal ulcer area is based on the U-Net configuration in Table 5. The model outputs are multiplied with the cornea segmentation mask from EM-5-SCED. This step is applied to filter out the predicted corneal ulcer area that falls outside the boundary of the predicted cornea area. The remaining corneal ulcer region is the result of this process. This method is used as a baseline for comparison to our proposed method.

The second method is our proposed U-Net model with HCT cornea mask multiplication in the learning phase, named UNET-H-L. This model uses the same U-Net configuration in Table 5 as the baseline model. The key differences between our proposed method and the baseline approach are: (1) instead of applying a raw cornea mask, we apply the circular cornea mask obtained from EM-5-SCED-HT, and (2) instead of applying a mask in the inference phase, the cornea mask is applied in both the learning and inference phases.

The last method, named UNET-R-L, utilizes the same models as the proposed method in Figure 4, but replaces the circular cornea mask input from EM-5-SCED-HT with the raw predicted cornea mask from EM-5-SCED. The raw predicted cornea mask is considered a soft restriction in the learning phase, as it contains continuous values between 0 and 1, whereas the circular cornea mask is considered a hard restriction in the learning phase due to its binary values of 0 or 1. This method aims to observe the effect of using different masks during the learning phase.

Figure 13 presents graphs showing the training and validation accuracy and loss curves for the corneal ulcer segmentation models in the experiment. These graphs illustrate the accuracy and loss values for both training and validation processes across all training epochs. The data represent the first fold of a three-fold cross-validation process for six sample models: UNET2-R-I, UNET4-R-I, UNET2-R-L, UNET4-R-L, UNET2-H-L, and UNET4-H-L. The graphs reveal that the UNET2-R-L, UNET4-R-L, UNET2-H-L, and UNET4-H-L models, all of which incorporate the cornea mask in the learning phase, converge faster than the conventional UNET2-R-I and UNET4-R-I models. This improved convergence occurs because these models focus on learning from the fluorescein-stained areas within the cornea region. Additionally, the validation set accuracy is significantly higher for these models due to the elimination of excess areas from the corneal ulcer segmentation through the use of the cornea mask. The selection criterion for the final model is the achievement of the least validation loss over the course of training.

The performance of all methods is evaluated by comparing the predicted corneal ulcer area with the labeled images in terms of Accuracy, DSC, Sen, and IoU. Additionally, the MAE is used to compare the percentage of ulcerated area on cornea between the predicted and actual data. The values following MAE represent the standard deviation of the MAE, indicating the variability of the error around the mean in the predicted percentage of ulcerated area on the cornea. The results from this experiment are shown in Table 6. For segmentation performance, UNET2-R-I performs the best among all baseline models (UNET-R-I group), with an Accuracy of 99.32%, DSC of 89.22%, Sen of 86.91%, IoU of 80.54%, and MAE of 2.56%. The best model of the UNET-R-L group is UNET2-R-L, which segments corneal ulcers with an Accuracy of 99.27%, DSC of 88.64%, Sen of 88.53%, IoU

of 79.60%, and MAE of 2.60%. In the proposed method (UNET-H-L group), UNET4-H-L provides the best performance, achieving an Accuracy of 99.35%, DSC of 89.87%, Sen of 89.17%, IoU of 81.60%, and MAE of 2.51%.
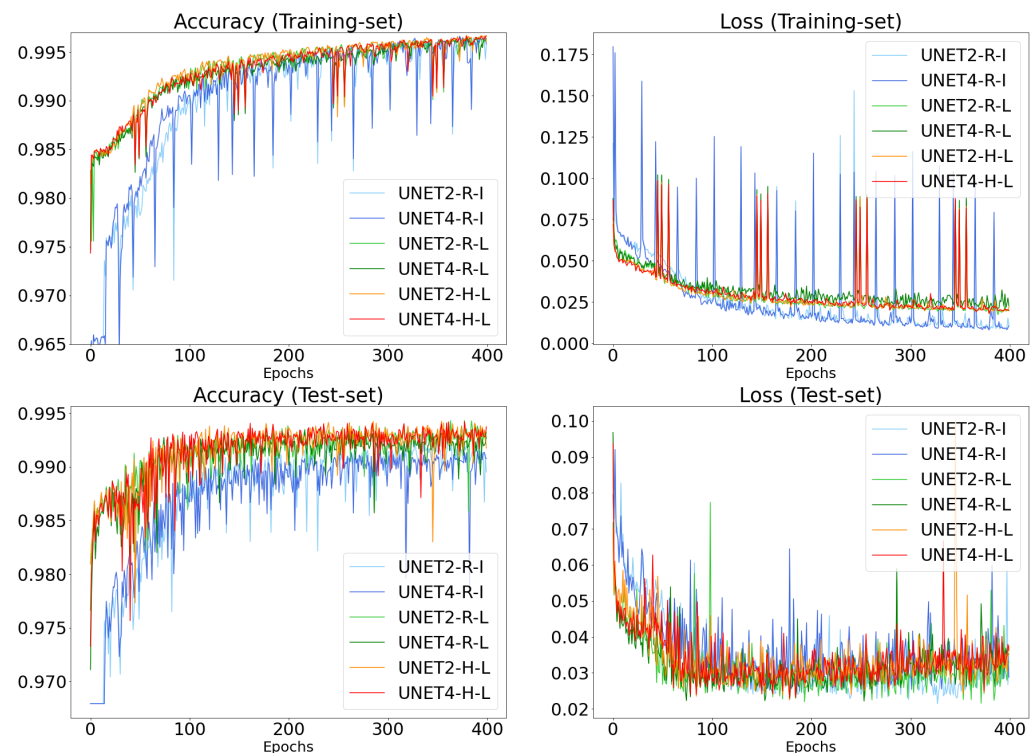
Figure 13. Training and validation accuracy and loss curves for the corneal ulcer segmentation models.

Comparing the results from the three main methods with similar U-Net architectures, UNET-H-L provides superior results in all metrics except for the MAE of UNET2-R-I and UNET2-R-L, which are slightly better than UNET2-H-L. The UNET-R-I and UNET-R-L methods demonstrate comparable performance in corneal ulcer segmentation tasks. This is because both methods use similarly shaped cornea masks to eliminate the fluorescein-stained areas outside the segmented cornea region. As a result, the outcomes do not differ significantly.

In terms of memory (VRAM), the usage for each model depends on the trainable parameters, as shown in Tables 5 and 6. Models using the UNET2 configuration have the highest number of parameters and require the longest training time per epoch across all model groups. Regarding training time per epoch, the UNET-R-L and UNET-H-L groups are quite similar. Meanwhile, models in both of these groups take longer times to train per epoch compared to models in the UNET-R-I group with the same configuration. This is because an additional image (the cornea mask) needs to be loaded and processed in the model learning phase.

In the context of varying model architectures, utilizing a kernel size of 5 generally yields better results than a kernel size of 3 across all metrics within the same model architecture for all methods except for UNET4-H-L, which uses a kernel size of 3. This method achieves better Accuracy, DSC, IoU, and MAE values than UNET2-H-L, which uses a kernel size of 5, by 0.03%, 0.36%, 0.58%, and 0.14%, respectively. For feature map size, models using larger feature map sizes tend to outperform models using smaller feature map sizes in all metrics except for the Sen of UNET4-R-I, which is lower than UNET3-R-I (refer to Tables 5 and 6). Moreover, utilizing a larger feature map with a kernel size of 5 results in UNET2-R-I and UNET2-R-L being the most efficient models in the UNET-R-I and UNET-R-L groups, respectively. However, the proposed UNET4-H-L method, which applies a kernel size of 3 and a large feature map size, provides better outcomes than

UNET2-H-L, which applies a kernel size of 5 and the same feature map size. As a result, UNET4-H-L stands out as the best method in this phase.

Table 6. Comparison of all individual corneal ulcer segmentation models.

Method	Accuracy	Micro DSC	Micro Sen	Micro IoU	MAE	Memory Usage	Training Time per Epoch
UNET1-R-I	99.25	88.21	86.37	78.91	2.74 ± 5.33	20.8 GB	38 s/epoch
UNET2-R-I	99.32	89.22	86.91	80.54	2.56 ± 5.09	21.1 GB	42 s/epoch
UNET3-R-I	99.22	87.87	86.63	78.36	2.82 ± 5.52	14.6 GB	36 s/epoch
UNET4-R-I	99.29	88.57	85.43	79.48	2.82 ± 5.80	21.0 GB	40 s/epoch
UNET1-R-L	99.20	87.54	86.90	77.85	2.87 ± 5.90	20.8 GB	44 s/epoch
UNET2-R-L	99.27	88.64	88.53	79.60	2.60 ± 5.11	21.1 GB	49 s/epoch
UNET3-R-L	99.21	87.55	85.73	77.85	2.80 ± 5.26	14.6 GB	42 s/epoch
UNET4-R-L	99.23	87.86	86.37	78.34	2.69 ± 4.98	21.0 GB	47 s/epoch
UNET1-H-L	99.29	88.87	88.31	79.96	2.74 ± 5.37	20.8 GB	44 s/epoch
UNET2-H-L	99.32	89.51	90.15	81.02	2.65 ± 5.31	21.1 GB	50 s/epoch
UNET3-H-L	99.26	88.38	87.57	79.17	2.77 ± 5.38	14.6 GB	42 s/epoch
UNET4-H-L	99.35	89.87	89.17	81.60	2.51 ± 4.68	21.0 GB	49 s/epoch

Next, we employed a voting ensemble to combine the corneal ulcer segmentation results from four models within each method and evaluate their performance. Accordingly, EM-4-UNET-R-I was an ensemble of models from UNET1-R-I to UNET4-R-I, EM-4-UNET-R-L was an ensemble of models from UNET1-R-L to UNET4-R-L, and EM-4-UNET-H-L was an ensemble of models from UNET1-H-L to UNET4-H-L. The results of the ensemble approach in Table 7 demonstrate that the proposed method (EM-4-UNET-H-L) outperforms the conventional method (EM-4-UNET-R-I) in terms of all metrics. Accuracy increases by 0.05%, DSC increases by 1.04%, Sen increases by 3.07%, IoU increases by 1.72%, and MAE decreases by 0.19%. Moreover, using the segmented cornea area from HCT multiplication in the learning phase (EM-4-UNET-H-L) yields better corneal ulcer segmentation result than using the raw predicted cornea mask (EM-4-UNET-R-L) for all metrics. Notably, Accuracy increases by 0.07%, DSC increases by 1.14%, Sen increases by 2.22%, IoU increases by 1.88%, and MAE decreases by 0.13%. The voting ensemble technique enhances the segmentation performance of the UNET-H-L method, which can be observed by comparing the results of UNET4-H-L, with the best performance in the UNET-H-L group, to EM-4-UNET-H-L. The accuracy increases from 99.35% to 99.37%, the DSC increases from 89.87% to 90.00%, and the IoU increases from 81.60% to 81.83%. However, Sen decreases from 89.17% to 88.19%, and the MAE remains similar. Based on these results, the proposed method (EM-4-UNET-H-L) outperforms all other corneal ulcer segmentation methods in this experiment.

Table 7. Comparison of the corneal ulcer segmentation ensemble models.

Method	Accuracy	Micro DSC	Micro Sen	Micro IoU	MAE	Memory Usage	Training Time per Epoch
EM-4-UNET-R-I	99.32	88.96	85.12	80.11	2.70 ± 5.28	21.1 GB	39 s/epoch
EM-4-UNET-R-L	99.30	88.86	85.97	79.95	2.64 ± 4.78	21.1 GB	46 s/epoch
EM-4-UNET-H-L	99.37	90.00	88.19	81.83	2.51 ± 4.63	21.1 GB	46 s/epoch

Sample results from the highest-performing instance of each method are shown in Figure 14. This comparison illustrates that the proposed UNET4-H-L method segments the corneal ulcer region more accurately than UNET2-R-I and UNET2-R-L. This can be observed from the true positive areas in UNET4-H-L, which more comprehensively fill the corneal ulcer area in the eye images. Furthermore, the false negative and false positive areas from UNET-4-H-L are slightly smaller than the outcomes from UNET2-R-I and UNET-2-R-L. These consequences are also evident in the ensemble model results. Upon visual

examination, EM-4-UNET-H-L can segment the corneal ulcer area most accurately; these results are related to the experimental results in Tables 6 and 7.
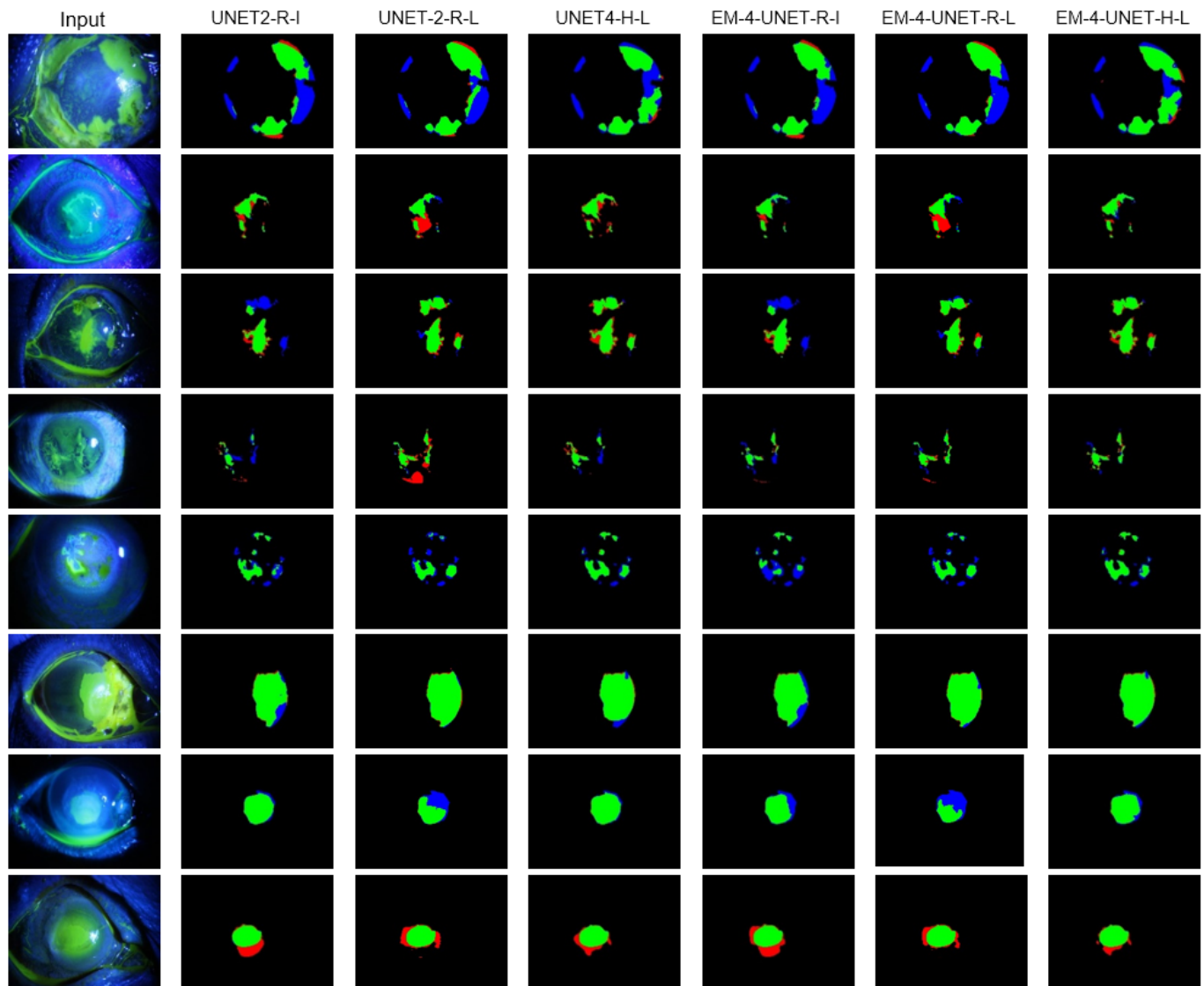
Figure 14. Comparison of corneal ulcer segmentation results. (Green = True Positive, Blue = False Negative, Red = False Positive)

We summarize the results from our corneal ulcer segmentation approach in comparison to conventional methods in Tables 8 and 9. Each method is described in detail as follows:

1. U-Net: This method uses the UNET2 model, which provided the best result of all U-Net models in this experiment, to segment the corneal ulcer area from the dataset. This method does not apply the cornea mask to filter out the segmented corneal ulcer area outside the cornea area.
2. U-Net-ResNet-50: We developed U-net–ResNet-50 by utilizing the encoder network from the ResNet-50 backbone combined with the decoder network from the U-Net architecture to segment the corneal ulcer area. This method also does not apply the cornea mask.
3. U-Net with raw cornea mask: The results of this method were derived from UNET2-R-I, which uses the corneal ulcer area from the UNET2 model and applies a raw cornea mask from EM-5-SCED during the inference phase to remove the excess corneal ulcer area outside the cornea area.

4. Ensemble U-Net models with cornea mask: This method uses the ensemble segmented corneal ulcer areas from UNET1-R-I to UNET4-R-I (EM-4-UNET-R-I), which applies a raw cornea mask from EM-5-SCED in the inference phase.

5. Proposed method: The results of the proposed method are from the ensemble segmented corneal ulcer area from UNET1-H-L to UNET4-H-L (EM-4-UNET-H-L), which applies the cornea mask from EM-5-SCED-HT in the learning phase and inference phase.

The comparison included the maximum memory usage evaluated during training and the total training time, representing the end-to-end process for training each method. For conventional methods, U-Net outperforms U-Net with the ResNet-50 backbone across all performance metrics. However, the U-Net model requires more resources in terms of training time and memory usage. Notably, the U-Net-ResNet-50 model underperforms compared to the other methods due to overfitting during model training. In comparing U-Net and U-Net with raw cornea masks, the results show that utilizing the predicted cornea mask to eliminate excess predicted corneal ulcer areas improves segmentation results across all metrics. However, this improvement comes at the cost of increased training time due to the additional step of training the cornea segmentation models. The proposed method demonstrates superior performance, outperforming all conventional methods across all metrics used in this experiment. Compared to the ensemble U-Net models with raw cornea masks, which have the same number of trainable parameters and consume the same maximum memory, the proposed method not only provides better performance but also requires less training time due to faster convergence. These results demonstrate that applying HCT cornea mask multiplication in the model learning phase can improve both performance outcomes and resource efficiency.

Table 8. Comparison of corneal ulcer segmentation performance across various methods.

Method	Accuracy	Micro DSC (%)	Micro Sen (%)	Micro IoU (%)	MAE
U-Net	99.24	88.15	87.56	78.88	2.68 ± 5.23
U-Net-ResNet-50	98.95	82.68	77.82	70.48	4.29 ± 9.89
U-Net with raw cornea mask	99.32	89.22	86.91	80.54	2.56 ± 5.09
Ensemble U-Net models with raw cornea mask	99.32	88.96	85.12	80.11	2.70 ± 5.28
Proposed method	99.37	90.00	88.19	81.83	2.51 ± 4.63

Table 9. Comparison of resources and time required to create corneal ulcer segmentation models.

Method	Trainable Parameters	Maximum Memory Usage	Total Training Time
U-Net	86,811,013	21.1 GB	230 min
U-Net-ResNet-50	73,378,625	11.4 GB	70 min
U-Net with raw cornea mask	104,002,586	21.1 GB	446 min
Ensemble U-Net models with raw cornea mask	171,469,097	21.1 GB	922 min
Proposed method	171,469,097	21.1 GB	788 min

In Figure 15, the graph represents a scatter plot between the actual and predicted percentage of ulcerated area on the cornea obtained from our best model (EM-4-UNET-H-L). From 354 images analyzed, 94.65% (335 images) fall within a margin of error of ±10%, while 86.72% (307 images) fall within a margin of error of ±5%. The coefficient of determination (r-squared) is calculated as 0.94. These results satisfy our expectations as to the accuracy and reliability of the proposed method. However, the experiment still provides some erroneous results. We highlight and analyze some of the inaccurately predicted outcomes, labeled as (a), (b), and (c), in Figures 15 and 16 to identify the causes of these inconsistencies. In these cases, significant errors in determining the percentage of the ulcerated area on the cornea arose from opposing errors in determining the cornea area and corneal ulcer area, explained as follows: in case (a), while the results from cornea segmentation are smaller than the actual cornea area, the outcome of corneal ulcer segmentation is larger than the actual corneal ulcer area; in case (b), the predicted cornea area is too large compared to the actual cornea

area, but the predicted corneal ulcer area is slightly smaller than the actual corneal ulcer area; finally, in case (c), the results from cornea segmentation provide a moderately large area compared to the actual cornea area, but the outcome from corneal ulcer segmentation is considerably smaller than the actual corneal ulcer area.
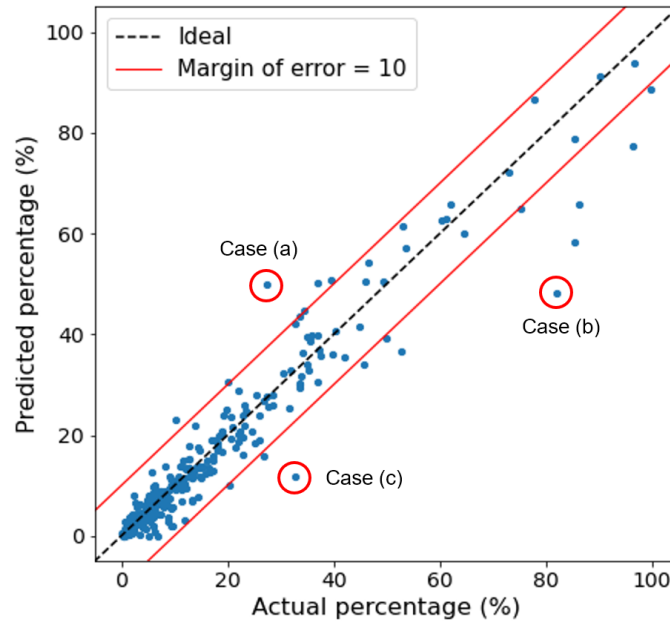
Figure 15. Scatter plot showing the predicted and actual percentage of ulcerated area on the cornea from EM-4-UNET-H-L.
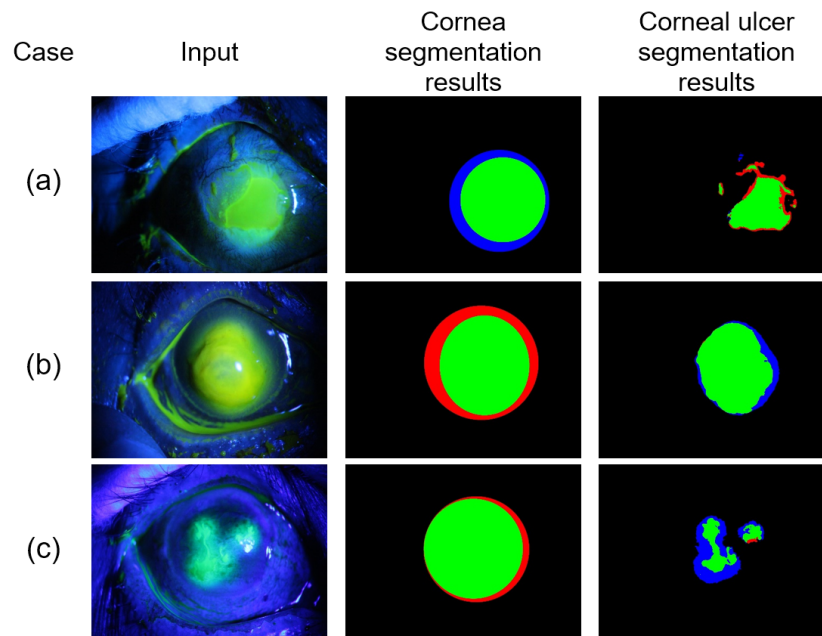

Figure 16. Case (**a**) represents a case of predicting a higher percentage of ulcerated area than the actual value displayed in Figure 15. Case (**b**,**c**) represent cases of predicting a lower percentage of ulcerated area than the actual value displayed in Figure 15. (Green = True Positive, Blue = False Negative, Red = False Positive).

4.6. Severity Grading Results

To grade the severity of corneal ulcers, the corneal area is divided into five parts based on the TG grading criteria. The central zone of the cornea is represented by a circle or an
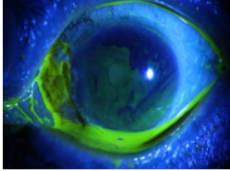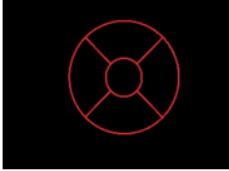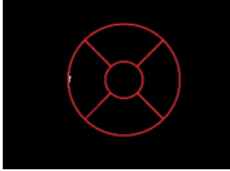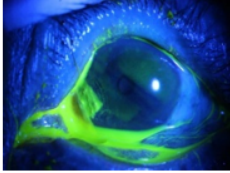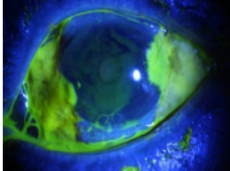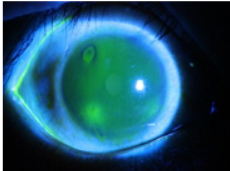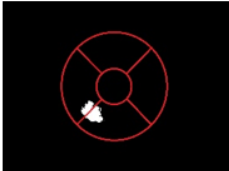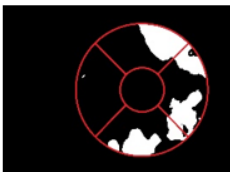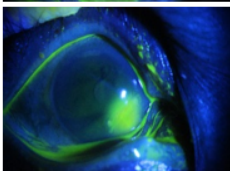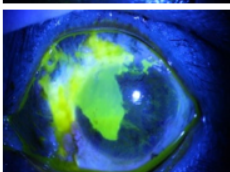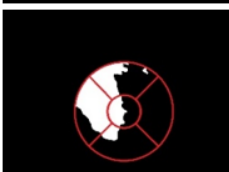
ellipse with a radius equal to one-third of the cornea's radius. The outer area around the central zone is divided equally into four quadrants, as shown in Figure 5. To evaluate the performance of severity grading, we used 390 staining images, each graded twice. The first grading result, called the actual grade, was obtained using the actual corneal area and actual corneal ulcer area as inputs. The second grading result, called the predicted grade, was obtained using the predicted corneal area from the EM-5-SCED-HT model and predicted corneal ulcer area from the EM-4-UNET-H-L model as inputs. Among the 390 test images, there were 36 images of grade0 cases, 88 images of grade1 cases, 28 images of grade2 cases, 6 images of grade3 cases, and 232 images of grade4 cases. The proposed method achieves an accuracy of 86.15% in grading the severity of corneal ulcers; the confusion matrix is illustrated in Figure 17. The system was able to correctly grade 77.78% (28 out of 36 images) of grade0 cases, 71.59% (63 out of 88 images) of grade1 cases, 60.71% (17 out of 28 images) of grade2 cases, 83.33% (5 out of 6 images) of grade3 cases, and 96.12% (223 out of 232 images) of grade4 cases.
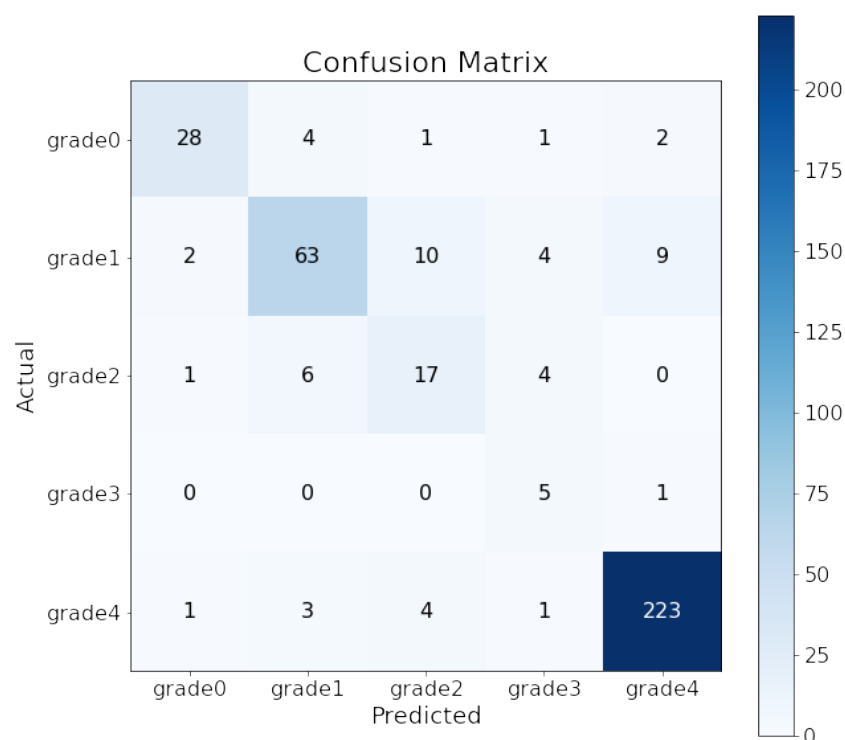
Figure 17. Confusion matrix for corneal ulcer grading.

Errors in severity grading occur due to inaccurate cornea segmentation and corneal ulcer segmentation. Most errors involve predictions that are off by one grade level. For instance, 11.11% of Grade0 classes were mistakenly predicted as Grade1, 11.36% of Grade1 classes were incorrectly classified as Grade2, and 35.71% of Grade2 classes were misclassified as either Grade1 or Grade3. Even minor segmentation errors for both cornea and corneal ulcers can lead to significant grading inaccuracies. For example, in Figure 18 case (a), the system incorrectly identified a small area in the cornea as an ulcer where none existed, leading to an erroneous severity grade of Grade1 for an image that actually contained no corneal ulcer. In case (d), an incorrect segmentation of the corneal ulcer area at the top-left of the cornea resulted in the ulcer areas covering three quadrants surrounding the central area instead of two. Moreover, when the ulcer area is close to the boundary of each section, the chance of grading errors increases significantly. For example, in case (f), there was a slight error in predicting the ulcer area at the edge of the pupil region. The system failed to detect an ulcer in the pupil area, and the segmented cornea area was slightly smaller than the actual one. These errors led to an assigned severity grade of Grade2, whereas the correct grade was Grade4. Furthermore, 10.22% of Grade1 classes were predicted as

Grade4 classes. This occurred due to erroneous predictions of the ulcer area on the cornea which extended into the central section of the corneal area, causing the system to grade the severity as higher than it actually was. Notably, the dataset used in this experiment has a limitation, as it contains an insufficient number of Grade2 and Grade3 corneal ulcer cases, which are crucial for accurately evaluating the algorithm's performance across all severity degrees of the condition.

Figure 18. Sample of the severity grading results. Case (**a**,**d**,**f**) represent cases of incorrect severity grading for corneal ulcer. Case (**b**,**c**,**e**,**g**) represent cases of correct severity grading for corneal ulcer.

5. Conclusions and Discussion

This paper proposes an automated corneal ulcer grading system using ocular staining images based on deep learning and the Hough Circle Transform. The algorithm consists of two main components, namely, the cornea segmentation and corneal ulcer segmentation modules, which together aim to assess the severity of corneal ulcers. The proposed method

utilizes a combination of deep learning techniques with cornea masks generated by the Hough Circle Transform applied in the learning phase. In our experiments, we utilized the SUSTech-SYSU dataset for training and testing both the cornea segmentation and corneal ulcer segmentation modules. The best performing method identified in this study achieved an accuracy of 86.15% in determining the severity degree of corneal ulcers.

For cornea segmentation, the proposed method employs deep learning to segment corneal areas and utilizes the Hough Circle Transform to represent model outputs in circular shapes. We compared the outputs of this approach with raw results from deep learning models. The comparison shows that the proposed method does not provide better outcomes; however, applying the results from the Hough Circle Transform as masks in the learning phase while training the corneal ulcer segmentation models can improve model performance compared to the baseline method. Additionally, employing a voting ensemble to combine corresponding predicted images from each method significantly enhances the performance of cornea segmentation and corneal ulcer segmentation over individual models.

In our comparison of corneal ulcer segmentation methods, the proposed method demonstrated superior performance, achieving improvements of 0.05% in Accuracy, 1.04% in Dice Similarity Coefficient, 3.07% in Sensitivity, 1.72% in Intersection over Union, and 0.19% in Mean Absolute Error compared to the baseline method relying on a deep learning technique with raw predicted cornea masks during the inference phase. Moreover, the implementation time was reduced by 13.88%

However, this work has several limitations. The cornea shape in some images is represented by the elliptical; therefore, using Hough Circle Transform to locate the cornea area can cause inaccurate cornea area location and adversely affect corneal ulcer segmentation performance. Furthermore, in this study we were not able to accurately segment point-like and mixed point–flaky corneal ulcers due to a lack of labeled data for these cases in the dataset. Therefore, future work should focus on applying ellipse detection to locate corneal areas and preparing labeled data for point-like and mixed point–flaky corneal ulcers in order to further improve segmentation and severity grading performance.

Author Contributions: Conceptualization, K.P.; methodology, K.P.; software, D.M.; validation, D.M.; formal analysis, D.M.; investigation, D.M.; resources, D.M. and K.P.; data curation, D.M.; writing—original draft preparation, D.M.; writing—review and editing, K.P.; visualization, D.M.; supervision, K.P.; project administration, D.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article. These data were derived from the following resources available in the public domain: https://github.com/CRazorback/The-SUSTech-SYSU-dataset-for-automatically-segmenting-and-classifying-corneal-ulcers (accessed on 20 April 2023). Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Amescua, G.; Miller, D.; Alfonso, E. What is causing the corneal ulcer? Management strategies for unresponsive corneal ulceration. *Eye* **2012**, *26*, 228–236. [CrossRef] [PubMed]
2. Deng, L.; Lyu, J.; Huang, H.; Deng, Y.; Yuan, J.; Tang, X. The SUSTech-SYSU dataset for automatically segmenting and classifying corneal ulcers. *Sci. Data* **2020**, *7*, 23. [CrossRef] [PubMed]

3. Bağbaba, A.; Şen, B.; Delen, D.; Uysal, B.S. An automated grading and diagnosis system for evaluation of dry eye syndrome. *J. Med. Syst.* **2018**, *42*, 1–14. [CrossRef] [PubMed]

4. Wang, T.; Wang, M.; Zhu, W.; Wang, L.; Chen, Z.; Peng, Y.; Shi, F.; Zhou, Y.; Yao, C.; Chen, X. Semi-msst-gan: A semi-supervised segmentation method for corneal ulcer segmentation in slit-lamp images. *Front. Neurosci.* **2022**, *15*, 793377. [CrossRef] [PubMed]

5. Deng, L.; Huang, H.; Yuan, J.; Tang, X. Superpixel based automatic segmentation of corneal ulcers from ocular staining images. In Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 19–21 November 2018; pp. 1–5.

6. Akram, A.; Debnath, R. An efficient automated corneal ulcer detection method using convolutional neural network. In Proceedings of the 2019 22nd IEEE International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 18–20 December 2019; pp. 1–6.

7. Bron, A.J.; Evans, V.E.; Smith, J.A. Grading of corneal and conjunctival staining in the context of other dry eye tests. *Cornea* **2003**, *22*, 640–650. [CrossRef]

8. Liu, Z.; Shi, Y.; Zhan, P.; Zhang, Y.; Gong, Y.; Tang, X. Automatic corneal ulcer segmentation combining Gaussian mixture modeling and Otsu method. In Proceedings of the 2019 41st IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 6298–6301.

9. Lozej, J.; Meden, B.; Struc, V.; Peer, P. End-to-end iris segmentation using u-net. In Proceedings of the 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), San Carlos, Costa Rica, 18–20 July 2018; pp. 1–6.

10. Raffei, A.F.M.; Hassan, R.; Kasim, S.; Asmuni, H.; Syifaa'Ahmad, A.; Hidayat, R.; Ahmar, A.S. Iris Segmentation. *Int. J. Eng. Technol.* **2018**, *7*, 77–83. [CrossRef]

11. Qasmieh, I.A.; Alquran, H.; Zyout, A.; Al-Issa, Y.; Mustafa, W.A.; Alsalatie, M. Automated Detection of Corneal Ulcer Using Combination Image Processing and Deep Learning. *Diagnostics* **2022**, *12*, 3204. [CrossRef]

12. Ghani, A.; See, C.H.; Sudhakaran, V.; Ahmad, J.; Abd-Alhameed, R. Accelerating retinal fundus image classification using artificial neural networks (ANNs) and reconfigurable hardware (FPGA). *Electronics* **2019**, *8*, 1522. [CrossRef]

13. Pahuja, R.; Sisodia, U.; Tiwari, A.; Sharma, S.; Nagrath, P. A Dynamic approach of eye disease classification using deep learning and machine learning model. In *Proceedings of the Data Analytics and Management: ICDAM 2021*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 1, pp. 719–736.

14. Topaloglu, I. Deep learning based convolutional neural network structured new image classification approach for eye disease identification. *Sci. Iran.* **2023**, *30*, 1731–1742. [CrossRef]

15. Sun, Q.; Deng, L.; Liu, J.; Huang, H.; Yuan, J.; Tang, X. Patch-based deep convolutional neural network for corneal ulcer area segmentation. In Proceedings of the Fetal, Infant and Ophthalmic Medical Image Analysis: International Workshop, FIFI 2017, and 4th International Workshop, OMIA 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, 14 September 2017; Proceedings 4; Springer: Berlin/Heidelberg, Germany, 2017; pp. 101–108.

16. Wang, T.; Zhu, W.; Wang, M.; Chen, Z.; Chen, X. Cu-segnet: Corneal ulcer segmentation network. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; pp. 1518–1521.

17. Kayalibay, B.; Jensen, G.; van der Smagt, P. CNN-based segmentation of medical imaging data. *arXiv* **2017**, arXiv:1701.03056.

18. Yasrab, R.; Gu, N.; Zhang, X. SCNet: A simplified encoder-decoder CNN for semantic segmentation. In Proceedings of the 2016 5th IEEE International Conference on Computer Science and Network Technology (ICCSNT), Changchun, China, 10–11 December 2016; pp. 785–789.

19. Ji, Y.; Zhang, H.; Zhang, Z.; Liu, M. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. *Inf. Sci.* **2021**, *546*, 835–857. [CrossRef]

20. Jamil, N.; Sembok, T.M.T.; Bakar, Z.A. Noise removal and enhancement of binary images using morphological operations. In Proceedings of the 2008 IEEE International Symposium on Information Technology, Kuala Lumpur, Malaysia, 26–28 August 2008; Volume 4, pp. 1–6.

21. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Pearson: Upper Saddle River, NJ, USA, 2010; pp. 649–661, ISBN 0-13-234563-3.

22. Spagnolo, F.; Frustaci, F.; Perri, S.; Corsonello, P. An efficient connected component labeling architecture for embedded systems. *J. Low Power Electron. Appl.* **2018**, *8*, 7. [CrossRef]

23. AbuBaker, A.; Qahwaji, R.; Ipson, S.; Saleh, M. One scan connected component labeling technique. In Proceedings of the 2007 IEEE International Conference on Signal Processing and Communications, Dubai, United Arab Emirates, 24–27 November 2007; pp. 1283–1286.

24. Kerbyson, D.; Atherton, T. *Circle Detection Using Hough Transform Filters*; IET Digital Library: London, UK, 1995.

25. Rizal, M.F.; Sarno, R.; Sabilla, S.I. Canny Edge and Hough Circle Transformation for Detecting Computer Answer Sheets. In Proceedings of the 2020 IEEE International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 19–20 September 2020; pp. 346–352.

26. Daud, M.M.; Diyana, W.M.; Zaki, W.; Hussain, A.; Mutalib, H.; Saad, M.H.M.; Razak, A. Automated corneal segmentation of anterior segment photographed images using centroid-based active contour model. *Procedia Comput. Sci.* **2019**, *163*, 330–337. [CrossRef]

27. Zheng, X.B.; Ling, B.W.K.; Zeng, Z.T. Evaluation of effectiveness of eye massage therapy via classification of periocular images. *Multimed. Tools Appl.* **2022**, *81*, 5743–5760. [CrossRef] [PubMed]

28. Li, P.; Liu, X.; Xiao, L.; Song, Q. Robust and accurate iris segmentation in very noisy iris images. *Image Vis. Comput.* **2010**, *28*, 246–253. [CrossRef]

29. Okokpujie, K.; Noma-Osaghae, E.; John, S.; Ajulibe, A. An improved iris segmentation technique using circular Hough transform. In *Proceedings of the IT Convergence and Security 2017: Volume 2*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 203–211.

30. Chatterjee, S.; Byun, Y.C. Voting Ensemble Approach for Enhancing Alzheimer's Disease Classification. *Sensors* **2022**, *22*, 7661. [CrossRef]

31. Karlos, S.; Kostopoulos, G.; Kotsiantis, S. A soft-voting ensemble based co-training scheme using static selection for binary classification problems. *Algorithms* **2020**, *13*, 26. [CrossRef]

32. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

33. Hu, X.; Yang, H. DRU-net: A novel U-net for biomedical image segmentation. *IET Image Process.* **2020**, *14*, 192–200. [CrossRef]

34. Rad, R.M.; Saeedi, P.; Au, J.; Havelock, J. Trophectoderm segmentation in human embryo images via inceptioned U-Net. *Med. Image Anal.* **2020**, *62*, 101612. [CrossRef]

35. Zhao, H.; Sun, N. Improved U-net model for nerve segmentation. In Proceedings of the Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, 13–15 September 2017; Revised Selected Papers, Part II 9; Springer: Berlin/Heidelberg, Germany, 2017; pp. 496–504.

36. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]

37. Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The importance of skip connections in biomedical image segmentation. In *Proceedings of the International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 179–187.

38. Begley, C.; Caffery, B.; Chalmers, R.; Situ, P.; Simpson, T.; Nelson, J.D. Review and analysis of grading scales for ocular surface staining. *Ocul. Surf.* **2019**, *17*, 208–220. [CrossRef] [PubMed]

39. Alakuş, T.B.; Baykara, M. Classification and Determination of Severity of Corneal Ulcer with Vision Transformer Based on the Analysis of Public Image Dataset of Fluorescein-Stained Corneas. *Diagnostics* **2024**, *14*, 786. [CrossRef] [PubMed]

40. Lv, L.; Peng, M.; Wang, X.; Wu, Y. Multi-scale information fusion network with label smoothing strategy for corneal ulcer classification in slit lamp images. *Front. Neurosci.* **2022**, *16*, 993234. [CrossRef] [PubMed]

41. Cinar, I.; Taspinar, Y.S.; Kursun, R.; Koklu, M. Identification of corneal ulcers with pre-trained AlexNet based on transfer learning. In Proceedings of the 2022 11th IEEE Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 7–10 June 2022; pp. 1–4.

42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

43. Ostu, N. A threshold selection method from gray-level histograms. *IEEE Trans. SMC* **1979**, *9*, 62.

44. Gielis, J. A generic geometric transformation that unifies a wide range of natural and abstract shapes. *Am. J. Bot.* **2003**, *90*, 333–338. [CrossRef]

45. Loo, J.; Kriegel, M.F.; Tuohy, M.M.; Kim, K.H.; Prajna, V.; Woodward, M.A.; Farsiu, S. Open-source automatic segmentation of ocular structures and biomarkers of microbial keratitis on slit-lamp photography images using deep learning. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 88–99. [CrossRef]

46. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

47. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698. [CrossRef]

48. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]

49. Okman, O.E.; Akar, G.B. A circle detection approach based on Radon Transform. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2119–2123.

50. Urbancic, T.; Fras, M.K.; Stopar, B.; Koler, B. The Influence of the Input Parameters Selection on the Ransac Results. *Int. J. Simul. Model.* **2014**, *13*, 159–170. [CrossRef]