



Article

Impact of Image Complexity on the Segmentation of Ophthalmic Images Using Deep Learning Methods

Ignacio A. Viedma¹, David Alonso-Caneiro^{1,2,*}, Scott A. Read¹, Jason Charng^{3,4}, Fred K. Chen^{4,5,6,7}, David A. Mackey^{3,5} and Michael J. Collins¹

¹ Contact Lens and Visual Optics Laboratory, Optometry and Vision Science, QUT, Brisbane, 4059 QLD, Australia

² School of Science, Technology and Engineering, University of Sunshine Coast, Sunshine Coast, 4502 QLD, Australia

³ Centre for Ophthalmology and Visual Science, The University of Western Australia, Perth, 6009 WA, Australia

⁴ Department of Optometry, School of Allied Health, University of Western Australia, Perth, 6009 WA, Australia

⁵ Lions Eye Institute, Nedlands, 6009 WA, Australia

⁶ Ophthalmology, Department of Surgery, University of Melbourne, East Melbourne, 3002 VIC, Australia

⁷ Royal Victorian Eye and Ear Hospital, East Melbourne, 3002 VIC, Australia

* Correspondence: dalonsocaneiro@usc.edu.au

How To Cite: Viedma, I.A.; Alonso-Caneiro, D.; Read, S.A.; et al. Impact of Image Complexity on the Segmentation of Ophthalmic Images Using Deep Learning Methods. *Journal of Bio-optics* **2025**, *1*(1), 2

Received: 10 December 2024

Revised: 10 February 2025

Accepted: 13 March 2025

Published: 1 April 2025

Abstract: Advances in medical imaging segmentation using deep learning (DL) have facilitated the development of a wide range of models based on different architectures. For example, the U-Net has become one of the most widely used architectures in the field. Due to its popularity, various modifications to the original U-Net architecture have been proposed, with the aim to improve the segmentation performance. Most studies utilizing ophthalmic images, such as optical coherence tomography (OCT) and scanning laser ophthalmoscopy (SLO), have employed U-Net methods and their variations for segmentation. Given the multitude of U-Net variations, selecting the optimum model for ophthalmic image segmentation may be challenging, and other factors that may impact the model's segmentation performance, such as image complexity and its impact on model selection, have been largely unexplored. Thus, in this study, the segmentation performance of three models, including a baseline U-Net, a popular U-Net variation (U-Net++), and a segmentation architecture with a different approach (DeepLabV3), are compared, analyzing how these different methods may vary for OCT and SLO datasets with various levels of segmentation task complexity. To analyze the effect of image complexity on segmentation performance, several metrics are extracted, including the image entropy of the datasets at the pixel level, the texture level, and the global features level. The results demonstrate a relationship between the complexity of the images in a dataset and the performance of the segmentation model used for the specific task. Data complexity may serve as a metric to inform DL model selection or aid in the early design process.

Keywords: optical coherence tomography; scanning laser ophthalmoscopy; semantic segmentation; deep learning; image complexity

1. Introduction

Ophthalmic imaging modalities, such as optical coherence tomography (OCT) and scanning laser ophthalmoscopy (SLO), have become fundamental for monitoring ocular health and diagnosing ophthalmic conditions. These technologies can capture high-resolution en face or cross-sectional images of the eye's posterior tissues (the retina and choroid) in a non-invasive manner.



Copyright: © 2025 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

These ophthalmic imaging modalities allow clinicians to assess ocular health, monitor the progression of retinal diseases, or evaluate the efficacy of treatments. To meet these clinical needs, quantitative biomarkers such as thickness metrics need to be extracted from the images. These quantitative clinical biomarkers require areas of interest within the image to be segmented with high accuracy and precision. This task can be time-consuming and prone to bias when performed manually by a human observer. Ideally, this image segmentation task can be automated using state-of-the-art deep learning (DL) computational models.

To date, DL models have been shown to achieve the repeatability and accuracy required for segmenting these types of medical images, demonstrating comparable performance to the current gold standard—expert human observers—especially when using models typically based on the encoder-decoder architecture (most commonly U-Net-based methods) [1–3]. This encoder-decoder architecture includes two main branches: the first branch downsamples the input image through a series of convolutions and pooling operations to capture the context of the image (encoder branch), generating a map with the extracted features of the image. The second branch applies the inverse operation to the generated map, upsampling the image to regain the original spatial information, thereby increasing the detail of the identified objects and generating the segmentation output (decoder branch).

When considering retinal OCT images, numerous DL examples have demonstrated highly accurate semantic segmentation performance of retinal and choroidal tissues [4–14] and anomaly detection [15–19]. As for SLO images, DL studies in the literature have focused on segmenting cone photoreceptors [20–23], retinal microaneurysms [24], or classifying glaucoma using convolutional neural networks (CNNs) [25]. In a recent literature review in the area of ophthalmic image analysis, a large portion of the segmentation studies proposed to date were noted to have utilized models with a U-Net encoder-decoder-based architecture [26].

The U-Net architecture has also become one of the most popular for semantic segmentation in other medical imaging domains. This model has demonstrated excellent performance in segmenting a range of medical images such as magnetic resonance, microscopy, dermoscopy, ultrasound, and X-ray, among others [27,28]. Due to its popularity and comparable performance to human observers, different variations of the architecture have been proposed, such as 3D U-Net [29], attention U-Net [30], recurrent U-Net [31], and U-Net++ [32], among others [27].

Given the large number of DL models, such as the various U-Net variations, it is not always clear how to select the most optimal segmentation model for a particular task. A study by Kugelman et al. [3] demonstrated that for four different OCT datasets, the benefit of U-Net variations was almost negligible when compared to the standard U-Net that served as a baseline for the segmentation task. This suggests that the standard U-Net model may be reliable for learning the features needed to perform well on the OCT segmentation task, but the use of more complex U-Net variations (deeper layers and a greater number of parameters), or DL architectures with a different approach (i.e., multiscale, increased resolution) may be needed for more complex datasets. Several studies, such as [33] and [34], have analyzed the impact that the complexity of dataset images can have on the training of DL segmentation architectures. These studies have demonstrated a relationship between the performance of segmentation models and the complexity of the datasets, which may help in choosing the best model for a specific dataset and segmentation task. In the context of medical imaging analysis, such as OCT and SLO segmentation, the link between model performance and data complexity has not been explored.

In this study, the segmentation performance of a baseline U-Net model is compared to another U-Net-based model (U-Net++), and a different DL architecture that focuses on extracting multiscale features (DeepLabV3) [35]. The models are tested on two different ophthalmic imaging modalities (OCT and SLO images) with images from both healthy and diseased eyes. This allows an investigation of how the various methods could aid in segmenting datasets of different complexity, including retinal and choroidal tissue segmentation in OCT images and anomaly segmentation in SLO images. To further explore the aspect of image complexity, an additional analysis is performed on the datasets to map the complexity of the images within the dataset and further explore how the complexity metrics relate to the segmentation performance of the trained models.

The rest of this study is divided into 5 sections, with Section 2 describing each of the datasets used. Section 3 presents each of the methods used in this study and the defined experiments. Section 4 describes the results and discusses both the segmentation and complexity analysis. Finally, in Section 5, the conclusions of the study are presented.

2. Dataset

For the purposes of this study, two different posterior eye imaging modalities were used to assess the impact of dataset complexity on the DL segmentation performance. These two different imaging modalities include: OCT imaging that provides cross-sectional images of the posterior segment of the eye providing details of the

retinal and choroidal tissues, and en-face “front-on” SLO images of the posterior segment of the eye. The details of each dataset are provided below. No preprocessing was applied to either dataset, as most DL methods are designed to work with raw and unprocessed images.

2.1. OCT Dataset

Two different OCT image datasets were used to explore the relationship between image complexity, segmentation performance and network architecture. The first OCT dataset from healthy children, which will be referred in this study as the Paediatric OCT dataset, has been presented in a previous study [36]. It contains OCT images from 100 healthy children aged between 10 and 15 years (mean age, 13.1 ± 1.4 years). The dataset was captured using the Spectralis OCT instrument (Heidelberg Engineering, Heidelberg, Germany), with the automatic real-time tracking modality (ART) enabled, and frame averaging to reduce speckle noise. The participants attended four visits with a six-month period between each visit. For each participant and visit, one set of six radial B-scans of 1536×512 pixels centred at the fovea and separated by 30 degrees were acquired right eyes. Approval from the Queensland University of Technology human research ethics committee was obtained before commencement of the study, and written informed consent was provided by all participating children and their parents.

From this Paediatric OCT dataset, only a subset of the images with quality index (QI) over 20 dB was used (mean QI from all images was 33 ± 3 dB), which included a total of 2135 OCT images obtained from 98 subjects belonging to the first and last visits of the dataset (18 months apart between these two visits). These images were split into three groups: the training set consisting of 1293 images belonging to 59 subjects; the validation set consisting of 429 images belonging to 20 subjects; and the evaluation set consisting of 413 images belonging to 19 subjects. For the segmentation task, the annotations to delineate the full retinal and choroidal tissues were used. The three annotated boundaries include (from top to bottom of the image): the internal limiting membrane (ILM), the retinal pigment epithelium (RPE) and the choroid scleral interface (CSI) boundaries. The retinal tissue is located between the RPE and ILM, while the choroidal tissue lies between the RPE and the CSI. Figure 1 presents an example of an annotated OCT image belonging to the Paediatric OCT dataset, where the retinal and the choroidal area annotations used to train each of the segmentation models are shown. It is worth noting that within each image, the segmentation of these boundaries represents a task of varying complexity, which may affect the tested models. It is expected that the choroid segmentation is a more complex task due to the lower contrast and less well-defined CSI boundary while the retinal boundaries have a more homogenous appearance across and within the dataset, representing a less complex task.

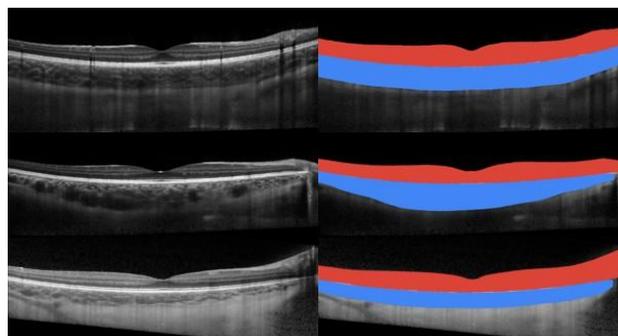


Figure 1. Example of the Paediatric OCT dataset, showing the original OCT images (**left** side) and their corresponding mask (**right** side). The retinal region is delineated in red while the choroidal tissue is in blue.

The second OCT dataset, referred to as Stargardt OCT dataset, consisted of retinal images of patients with Stargardt disease at various levels of progression [37]. This dataset was selected to compare how the retinal segmentation performance can be affected when more variability in the retinal tissues (due to the presence of pathology) is presented, thus adding complexity to the segmentation task in comparison to the use of the more homogenous data (i.e., shape, contrast, boundaries and level of noise) from the Paediatric dataset belonging to healthy eyes. To capture this Stargardt OCT dataset, 22 patients with an established diagnosis of Stargardt disease, had a series of SD-OCT scans captured using the Spectralis OCT+HRA device (Heidelberg Engineering, Heidelberg, Germany). The scanning protocol consisted of 61 raster lines covering an area of 30×25 degrees (8.8 mm horizontally \times 7.2 mm vertically) of the macula. The dataset consists of a total of 177 high-resolution volumetric scans (1536×512 pixels) captured from both eyes of each patient. The capture process consisted of non-regularly spaced visits, with a different number of visits per patient. Typically scans from the first and last visit were used, to add more variability into the data. Approval to identify and use SD-OCT images from

patients with genetically confirmed Stargardt disease for developing segmentation methods was obtained from the Human Ethics Office of Research Enterprise, The University of Western Australia (RA/4/1/8932 and RA/4/1/7916) and the Human Research Ethics Committee, Sir Charles Gairdner Hospital (2001-053).

For this Stargardt OCT dataset, a subset of images was selected with those subjects who presented more visual variability between the first and last visits (disease progression), obtaining a final dataset composed of a total of 2430 images divided in three groups: 1457 images from 8 subjects for training, 487 images from 2 subjects for validation, and 486 from 2 subjects to test the trained models and report the results. Figure 2 presents different examples of the Stargardt OCT dataset, showing the different stages of the disease and the variation in the retinal tissue. For this dataset, only the boundaries that delineate the full retinal tissue were available (ILM and RPE), so the analysis and performance comparison was only performed on the retinal tissue.

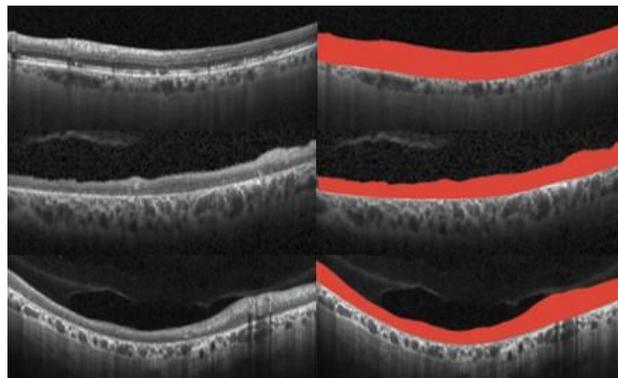


Figure 2. Example images of the Stargardt OCT dataset, at different stages of progression (less advanced at the top, to more advanced at the bottom). The original OCT image (left side) and their corresponding mask (right side) are shown. The retinal region is delineated in red.

2.2. SLO Dataset

For the SLO dataset, fundus autofluorescence images captured with two different excitation wavelengths: infrared (787 nm) (SLO-IRAF) and short (488 nm) (SLO-AF) wavelength were used. The images were centred on the fovea (30×30 degrees) and captured using the Heidelberg Spectralis OCT+HRA system (Heidelberg Engineering, Heidelberg, Germany) in both eyes of patients with a clinical diagnosis of retinitis pigmentosa. Clinical assessments were conducted by a retinal specialist (FKC) and eyes were dilated (tropicamide 1% and phenylephrine 2.5%) prior to image acquisition. High-resolution images (1536×1536 pixels) were acquired with the ART modality enabled. Image acquisition was performed by trained retinal imaging technicians. Patients with eccentric fixation were asked to shift their gaze to ensure that centre of the fovea aligned with the centre of the field of view. This study adhered to the tenets of the Declaration of Helsinki. Ethics approval was obtained from the human ethics committee of the University of Western Australia (2021/ET000251) and the Human Research Ethics Committee, Sir Charles Gairdner Hospital (2001-053). Written informed consent was obtained from all participants for their data to be used for research purposes.

The SLO autofluorescence imaging modality provides information on RPE integrity and can be regarded as an indirect marker of retinal health. For this dataset, the analysis was focused on the segmentation performance for defining the extent of retina affected by retinitis pigmentosa by delineating the boundary between hypo-fluorescent and hyper-fluorescent regions. The two different image wavelengths provide complementary information on the health of the RPE and show different features. For each image, using the manual annotations of the boundary between hyper-fluorescent and hypo-fluorescent retinal tissue, a segmentation mask for each annotated image was generated. Two sets of images were defined from the SLO images, with the SLO-IRAF dataset of infrared autofluorescence images and the SLO-AF dataset of short wavelength autofluorescence images. For each SLO image dataset, a total of 1100 images from 110 patients were used for this study, which were divided in 660 images for training, 220 for validation, and 220 for test. Figure 3 provides a graphical example of two SLO-AF and two SLO-IRAF images with their respective segmentation masks.

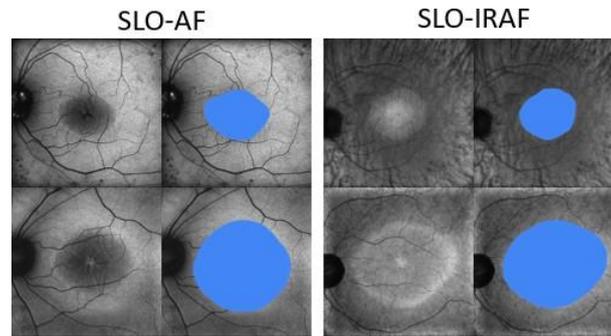


Figure 3. Example images from the SLO-AF and SLO-IRAF datasets with the presence of retinitis pigmentosa with different shapes and sizes, and their corresponding mask annotated in blue.

A total of four different datasets were used to compare the segmentation performance: “Paediatric OCT dataset”, “Stargardt OCT dataset”, “SLO-IRAF dataset”, and “SLO-AF dataset”. For all datasets, the data was divided to ensure different subjects were allocated in each set, thus avoiding mixing the images from subjects used for training, testing, and evaluating the models, assessing only images from new ‘unseen’ subjects.

3. Methods & Experiments

3.1. U-Net

Presented by Roenneberger et al. in [38], the U-Net architecture was developed based on the use of fully convolutional networks (FCN) [39], with the main difference being that U-Net concatenates the feature maps from each successive convolutional layer in the encoder branch to each layer in the upsampling step of the decoder at the same level, sharing context information to higher resolution layers and forming a symmetric U-shaped architecture. The wide use of the U-Net architecture for medical imaging segmentation is due to its good segmentation performance even while training with a limited number of images. This characteristic is a particularly appealing feature for medical image segmentation applications, in which datasets are often hard to capture and label, with many datasets only containing a limited number of annotated images.

The ‘standard architecture’ of the U-Net can be described in two parts: first, the encoder branch, in which, each convolutional block is composed of two successive 3×3 convolutions followed by a ReLU activation and a max-pooling layer. These blocks are arranged to be repeated several times and generate the feature map that will be used as the input for the decoder branch. For the second part of the U-Net architecture, the decoder up samples the feature map (encoder output) using the same quantity of arranged blocks, consisting of a 2×2 upsampler operator, followed by the concatenation of the feature map from the corresponding decoder layer, to which two successive 3×3 convolution and ReLU activation is applied. A final 1×1 convolution is applied at the end of the decoder branch to generate the segmented image with the required number of channels. Figure 4 illustrates an example of the U-Net architecture.

A U-Net architecture was trained to perform the defined segmentation experiments with both OCT and SLO imaging datasets, this network acts as a baseline reference model, which can be compared to the other implementations. Added to this, for segmentation performance comparison, different backbones for the U-Net architecture were tested. Namely ResNet50 [40], InceptionResNetV2 [41], and Xception [42], this was done to analyse how the variation in the backbone depths (i.e., network complexity and learning capacity) may impact the model performance.

3.2. U-Net++

In 2018, the authors in [32,43] proposed the U-Net++ architecture as a more powerful variation of the U-Net for medical image segmentation. In their study, the authors identified a limitation of the original U-Net architecture, where the skip connections are considered as being ‘too restrictive’, directly concatenating the output of the up-sampling convolution layers to the feature maps from its corresponding encoder branch layer. This combination of the high-level feature maps from the decoder with the low-level feature maps extracted from the encoder, may often combine semantically dissimilar features [32], which may affect the segmentation performance.

In order to overcome this limitation, the authors proposed the use of nested, dense skip connections. By using a dense network of skip connections inspired from DenseNet [44], the authors were able to propagate more semantic information between the encoder and decoder branches, to obtain a more accurate segmentation

performance. Figure 4 describes an example of the U-Net++ architecture, where the skip connections receive all the feature maps from all previous steps at the same level, reducing the loss of semantic information between the encoder and decoder branches, and which operation is defined as:

$$x^{ij} = \begin{cases} C(x^{i-1,j}), & j = 0 \\ C([\mathbf{x}^{i,k}]_{k=0}^{j-1}, U(x^{i+1,j-1})), & j > 0 \end{cases} \quad (1)$$

where x represents the feature map generated at the i and j index of the decoder branch level across the skip connections. $C(\cdot)$ represents the convolution and activation operation, $U(\cdot)$ the up-sampling operation and $[\]$ represents the concatenation step.

In this study, the U-Net++ architecture was used to analyse how this widely used modified structure of a baseline U-Net [27] could influence the segmentation performance. For comparison, the same backbone variations from using the U-Net were applied to the U-Net++.

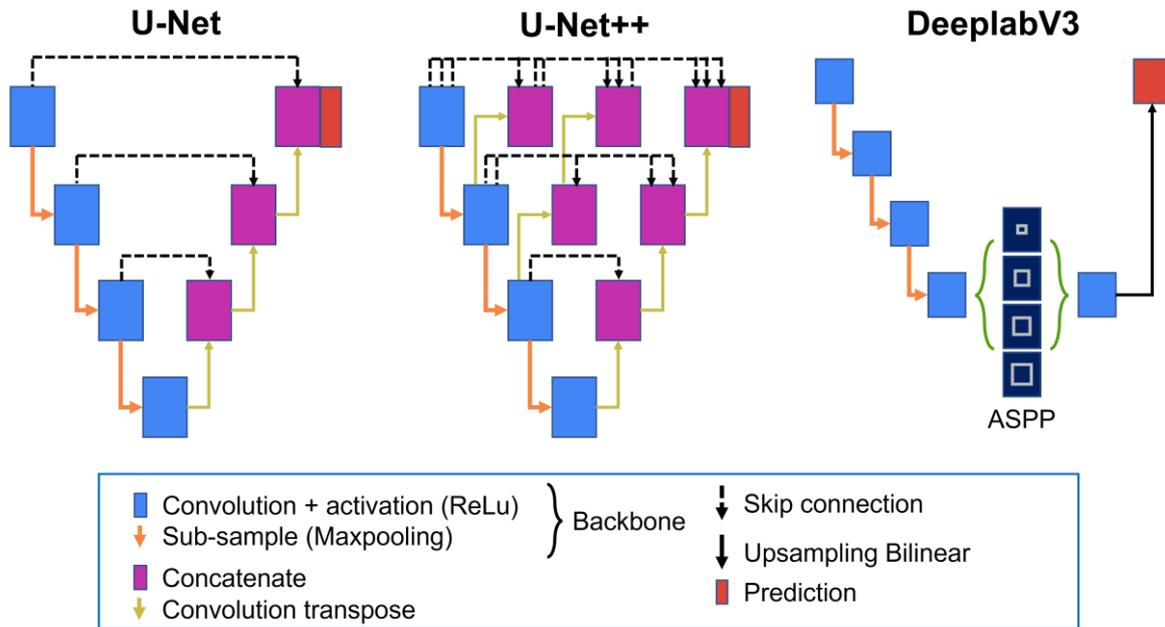


Figure 4. Graphical summary of the three models used in this study, including the U-Net, a U-Net variation (U-Net++) and different architecture (DeepLabV3) showing the differences between them. Where both U-Net based architectures share the information of the features extracted at different levels of the backbone through transposed convolutions and concatenation of the feature maps, while the DeepLabV3 architecture shares multi-scale information through the atrous spatial pyramid pooling (ASPP), to then generate the prediction through a bilinear up sampling layer. Three sub-sampling layers per backbone of the models are shown for simplicity, but each model varies in depth according to the used backbone.

3.3. DeepLabV3

When proposing the DeepLab architecture in [45], the authors considered a challenge present in the application of DL for the semantic segmentation task related to the reduction of features' resolution caused by the consecutive convolution and pooling operations, and the existence of objects at multiple scales. To overcome this challenge, the authors proposed the use of “atrous convolutions” (also known as dilated convolution) [46], which has been applied in different DL studies [47–49]. These atrous convolutions allows full resolution and more dense feature maps to be obtained, maintaining the detailed spatial information of the objects to be segmented. Considering an input feature map x , for each location i on the generated output y and a filter w , the atrous convolution is applied as follows:

$$y[i] = \sum_k x[i + r \cdot k]w[k] \quad (2)$$

where r corresponds to the stride rate used to sample the input x , which is equivalent to convolving it using up-sampled filters produced by inserting $r - 1$ zeros between two consecutive filter values.

An updated version of the DeepLab architecture was presented in [50], where the authors proposed the use of atrous spatial pyramid pooling (ASPP). Based on the success of the spatial pyramid pooling method of the R-CNN architecture [51], where using feature maps with filters or pooling operations at multiple rates and multiple scales can accurately and efficiently classify objects at multiple scales, the authors were able to improve the segmentation performance of their previous DeepLab architecture.

DeepLabV3 was presented in [35], where the authors discussed an important practical issue when applying 3×3 atrous convolutions in their DeepLab architecture, which fails to capture long range information when atrous convolution rate is close to the feature map size, reducing the 3×3 filter to a 1×1 , since only the centre filter weight is effective [35]. To overcome this problem, the authors trained different models using cascade and ASPP modules. DeepLabV3 uses both these modules, with a cascaded module which gradually doubles the atrous rates while a ASPP module augmented with image-level features probes the features with filters at multiple sampling rates and effective field of views. Figure 4 illustrates an example of the DeepLabV3 architecture, comparing it with both the U-Net based ones.

The DeepLabV3 with a ResNet50 backbone architecture was used to analyse how this semantic segmentation state-of-the-art architecture, different from the U-Net, performed in comparison to U-Net based architectures for OCT and SLO images.

3.4. Data Training

To train and test each of these architectures, the OCT images were cropped from their original size of 1536×512 pixels to three non-overlapping patches of 512×512 pixels each and those were fed to the network. The training process run for 100 epochs, calculating a Dice overlap loss, and using the Adam algorithm ($\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-7}$) to minimise it. After the training process, the models with the best validation Dice accuracy were chosen to perform the evaluation process. The same training parameters were used for the segmentation of the square SLO images, which were resized to 512×512 pixels.

Each of the segmentation architectures were implemented using the publicly available Segmentation Models Pytorch library [52] on an Intel i7-9800X 3.80 GHz processor and Nvidia TITAN Xp GPU.

3.5. Performance Evaluation

To assess the segmentation performance of each architecture and across datasets, the commonly used Dice coefficient metric was employed and compared in each experiment. All the values represent the mean of three independent network runs, to ensure appropriate assessment of the network repeatability and stability.

Image Complexity

In order to analyse how the differences between the features in healthy and diseased OCT, and SLO images can affect the segmentation performance of the methods, an image complexity analysis was performed. For this analysis, three different metrics were calculated, which included the entropy of each dataset at pixel, texture, and global levels. These metrics allow the information content at different levels of abstraction to be captured. The entropy metrics serve as a measure of the variability and complexity of the dataset. For all three entropy metrics, an increase in values indicates an increase in information and as a result the segmentation model may need to learn more ‘complex’ or a larger amount of features.

The first entropy corresponds to the standard Shannon entropy, which gives information about the entropy of a grayscale image at pixel level, defined as:

$$H = - \sum_{i=0}^{n-1} p_i \log p_i \quad (3)$$

where n is the number of grey levels and p is the probability of a pixel having a specific grey level i . This entropy has been previously used as a complexity measure for semantic segmentation in autonomous driving [33], showing that the pixel entropy of an image can be related to the first layers of the segmentation model which interprets pixel distributions.

The grey level co-occurrence matrix (GLCM) gives a histogram of co-occurrences of grayscale values at a given interval. This entropy measure helps to describe the complexity of texture features in an image. Defined as:

$$H_g = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} p(i, j) \log p(i, j) \quad (4)$$

where n is the number of grey levels and $p(i, j)$ is the probability of two pixels having intensities i and j , separated by the given interval. Previously used in [33], texture entropy is used to measure the image complexity at a higher feature level, which can be related to the difficulty that segmentation models will have learning in the middle or deeper layers, which are used to learn higher level features.

The last entropy measure calculated is the Delentropy, which is computed similar to the Shannon entropy but based on a probability density function called deldensity. This allows the entropy to be calculated for the spatial image structure and pixel co-occurrence, capturing global image features thanks to the deldensity function, which is based on gradient vectors. This global entropy metric has been used for complexity measures in medical imaging [34], and autonomous driving [33], showing that global image complexity measures can be related to the performance of segmentation models at later layers.

These three types of entropy were calculated for each image belonging to each dataset to be able to relate how their complexity may influence the segmentation results in the experiments. The values for the three different metrics have not been normalized, so direct comparisons across the entropy metrics do not provide meaningful information. Instead, comparisons can be made across datasets based on the relative entropy or complexity changes.

4. Results & Discussion

4.1. Segmentation Performance across Models

The results for the various dataset and architecture tests are described in Table 1. This table presents the outcomes of the segmentation results (Dice coefficients) obtained when applying the various models for both the OCT and SLO image datasets. The values in Table 1 presents the mean and standard deviation of Dice score from three independent runs of each network, which allows appropriate assessment of the networks' repeatability and stability performance.

Table 1. Summary of the results, including the model characteristics (number of layer [depth]/parameters in millions), and mean (standard deviation) Dice coefficient results after three repetitions training and testing each of the used models with the Paediatric OCT dataset for retina (PD-RT) and choroid (PD-CH) segmentation, the Stargardt OCT dataset for retina (SD-RT) segmentation, and with the SLO-AF and SLO-IRAF datasets for retinitis pigmentosa segmentation.

Model	UNet			UNet++			DeepLabV3	
Backbone	Baseline	ResNet50	InceptionResV2	Xception	ResNet50	InceptionResV2	Xception	ResNet50
Depth	162/0.5 M	328/34.5 M	1003/62 M	278/28.7 M	393/50 M	1068/71 M	343/34 M	297/41 M
OCT Segmentation performance								
PD-RT	0.995 (0.001)	0.987 (0.001)	0.988 (0.001)	0.988 (0.001)	0.988 (0.001)	0.988 (0.001)	0.988 (0.001)	0.988 (0.001)
SD-RT	0.966 (0.005)	0.963 (0.001)	0.964 (0.001)	0.962 (0.001)	0.962 (0.001)	0.962 (0.002)	0.963 (0.002)	0.960 (0.002)
PD-CH	0.980 (0.001)	0.977 (0.001)						
SLO Segmentation performance								
SLO-AF	0.852 (0.018)	0.931 (0.002)	0.948 (0.004)	0.932 (0.008)	0.926 (0.003)	0.942 (0.008)	0.939 (0.004)	0.943 (0.003)
SLO-IRAF	0.925 (0.001)	0.954 (0.002)	0.958 (0.003)	0.956 (0.002)	0.955 (0.001)	0.960 (0.002)	0.953 (0.002)	0.959 (0.001)

For the OCT results, the best segmentation performance for the retinal tissue in both OCT datasets (Paediatric and Stargardt) and the choroidal tissue for the Paediatric OCT dataset correspond to the baseline U-Net, obtaining a Dice coefficient of 0.99 and 0.96 when segmenting healthy and diseased retina respectively, and 0.98 when segmenting healthy choroid. This is to be expected since the retinal segmentation for the Stargardt OCT dataset would be more complex than the Paediatric, thus resulting in a slightly lower Dice coefficient for retinal segmentation. Similarly, the choroid segmentation task would be more complex than the healthy Paediatric retina segmentation, thus the lower averaged Dice metric. It is worth noting that the variation across the tested methods for the OCT image datasets is small, with absolute differences in Dice values being less than 0.01. Thus, for the OCT dataset it seems clear that the baseline U-Net, which has a shallower depth than the other networks, may be

sufficient to provide good segmentation performance for both healthy and disease examples. Deeper and more complex models do not seem to be needed, since that requires an increase in the size of the segmentation model and more computational resources but had limited positive impact on performance. This finding agrees with the results of Kugelman et al. [3] showing only small performance differences across various U-Net models for a range of OCT segmentation datasets.

Considering the results of the SLO images, the segmentation of the baseline U-Net provided the worst performance overall, with a 0.85 Dice for the SLO-AF and 0.92 for the SLO-IRAF image capture modalities. In the case of the SLO images, the results presented a larger variability across the different tested architectures than the OCT images, with absolute differences in Dice values ranging by 0.09 and 0.03 for the SLO-AF and SLO-IRAF, respectively. The best performance was obtained when training the U-Net with the InceptionResnetV2 backbone for the SLO-AF images (Dice of 0.94) and the best performance with SLO-IRAF images obtained when using the same backbone in a U-Net++ architecture (Dice of 0.96). Additionally, the results obtained with DeepLabV3 showed minor differences when compared to the best performance with the SLO images. Thus, in contrast to the OCT dataset, different results are shown when assessing the SLO dataset, with more variability in the segmentation performance across the different architectures. In this case, the lower performance was obtained for both AF and IRAF scan modalities when using the baseline U-Net, which was the architecture with the best results for segmentation of the OCT images while the InceptionResNetV2 backbone, which is the network with the deeper architecture, provides the best performance. This suggests that, to segment the SLO dataset, a more complex model may be required to provide good segmentation results.

Due to the larger variability in the SLO results, a qualitative graphical analysis was performed with the SLO images segmentation results for various models. Figure 5 illustrates the segmentation outputs of each model with the best encoder for the SLO-AF images. Although the figure presents only a few representative examples from the segmentation with SLO-AF images, the difference in performance when using each of the segmentation models is evident. The baseline U-Net, which was the model that gave best performance for the segmentation of retina and choroidal tissue on the OCT images, showed the worst performance on the SLO images, able only to segment small sections of the retinitis pigmentosa anomaly within the en-face images. After changing the backbone of the U-Net architecture, a noticeable improvement in the segmentation output was observed. However, inaccuracies were still noticeable, especially for some of the large anomalous regions. These segmentation inaccuracies can also be noticed in the U-Net++ segmentation performance and DeepLabV3 but at a lower level. It is also important to mention that for smaller anomalies, the segmentation seems to be more precise, reaching almost the same shape as the manual annotations.

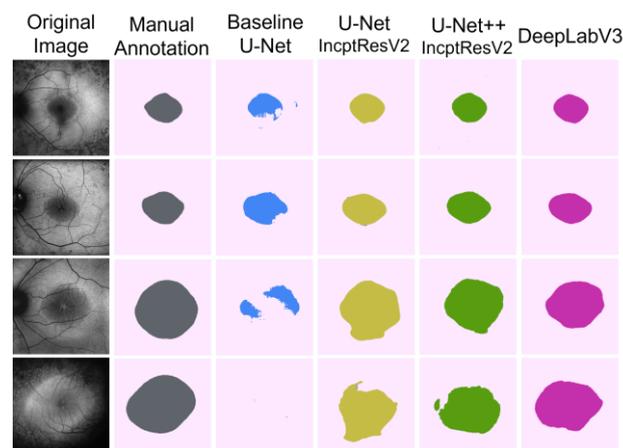


Figure 5. Graphical examples of the segmentation results obtained with the SLO-AF images across various DL models.

Figure 6 describes the segmentation outputs of each model with the best performance for the SLO-IRAF images. Similar to the quantitative findings, the segmentation output of each segmentation model using the SLO-IRAF images was better than SLO-AF. The performance is shown to be more consistent when compared with the manual annotations, with the exception of the baseline U-Net, which also shows the worst performance for this type of image modality.

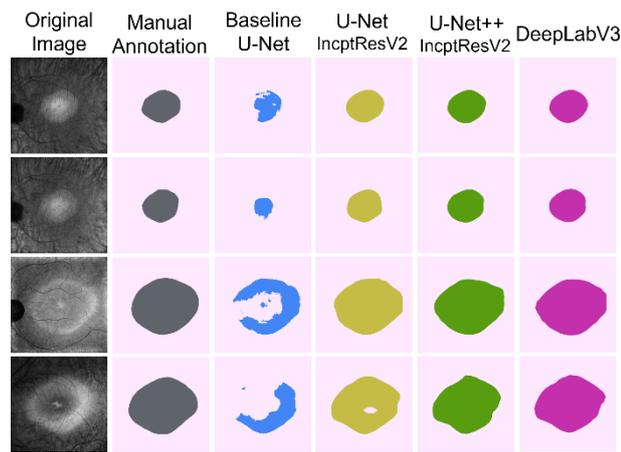


Figure 6. Graphical examples of the segmentation results obtained with the SLO-IRAF images across various DL models.

4.2. Data Complexity Analysis

The following provides an overview of the analysis of the image complexity of the datasets after calculating the three different entropy metrics based on different image features, to understand if the performance of the tested models relates to the image complexity. Table 2 describes the average entropy obtained from all the images within each of the OCT and SLO datasets. It is important to note that, across the three entropy metrics, the magnitudes cannot be compared as they reflect different measurements, but comparisons of values across datasets should provide an indication of the change in the information content for different imaging modalities.

Table 2. Mean (and standard deviation) of the calculated entropy of each dataset.

Dataset	Shannon	GCLM	Delentropy
Paediatric	6.06 (0.42)	3.92 (0.19)	4.53 (0.48)
Stargardt	6.27 (0.29)	3.42 (0.39)	4.90 (0.36)
SLO-AF	7.28 (0.22)	3.84 (0.63)	5.70 (0.35)
SLO-IRAF	7.01 (0.25)	3.96 (0.52)	5.83 (0.27)

From these results, a number of observations can be made: (i) for the two OCT datasets, across the three entropy metrics, small differences are observed, with the Stargardt OCT dataset presenting higher values (increased complexity) in two (pixel and global entropy) out of the three metrics when compared to the Paediatric OCT dataset. The relatively low differences in entropy between the two OCT dataset, may indicate small differences between the datasets, which also had similar segmentation results. The higher entropy metrics may indicate a greater challenge to learn image features at top (pixel level) and bottom (global level) layers of the model when using the Stargardt OCT dataset, affecting the performance of the model. (ii) When analysing the texture entropy, the Paediatric OCT dataset presents the higher mean score of the OCT images, showing more variability on texture features. The Stargardt OCT dataset presents a higher speckle noise in the images (grainy appearance), which reduces the contrast between the tissues, thus making it more difficult to accurately segment. And (iii) when looking across the two imaging modalities (SLO vs OCT) the entropy analysis reveals that the SLO images shows higher complexity than the OCT images (with lower variability on the texture entropy), showing the highest entropy at the three feature levels. These differences in complexity of the OCT and SLO images may also relate to the variation of segmentation performance, where higher Dice results are achieved when using more complex models for the more complex SLO datasets.

Figure 7 shows the entropy distribution obtained in each dataset (fitted to a normal distribution). The entropy distribution of the datasets can provide insight into the variability in the measured complexities, which can be later related to the segmentation models performances. From the pixel entropy distribution (Shannon), it is observable that SLO images have the lower spread (lower standard deviation), with higher entropy values, which can be related to a more difficult scenario for segmentation models when learning this type of features (more complex data). On the other hand, the pixel entropy distribution of the OCT images, specifically from the Paediatric OCT dataset, shows a larger spread of the values, demonstrating more variability, which can be related to the segmentation model being able to obtain a better generalization, aiding their performance.

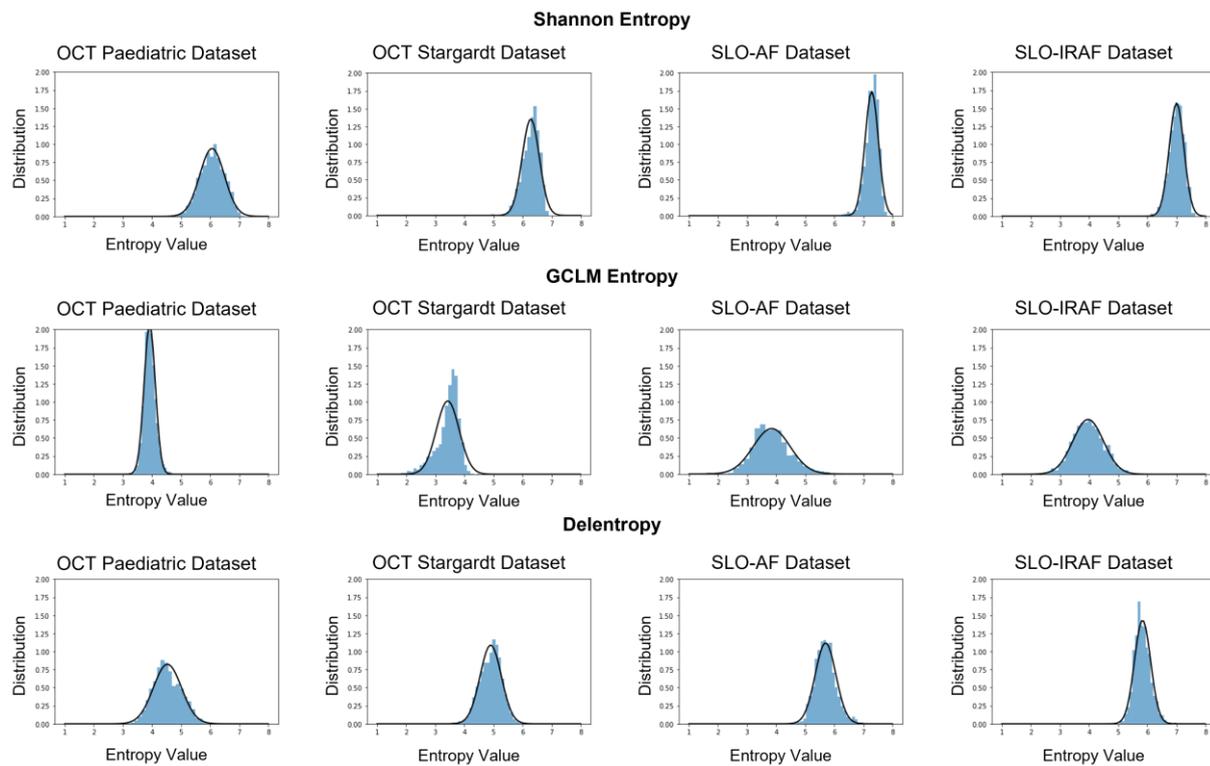


Figure 7. Shannon entropy (pixel level), GCLM entropy (texture level), and Delentropy (global level) distribution of the images inside the datasets.

A different case is observed when analysing the GCLM (texture) entropy in Figure 7. Although a lower spread is shown for the Paediatric OCT dataset, similar texture entropy was obtained among the OCT and SLO datasets, which can be attributed to the similar texture features that each of the images have due to the type, condition, or instrument on which they were captured, which are often affected by the presence of speckle noise, affecting the same way to the performance of the segmentation models. Similar behaviour to the pixel entropy distribution can be observed when analysing the global entropy distributions (Delentropy), showing smaller spread and higher complexity with the SLO images, and a better case scenario with the OCT images, with more variability and lower complexity.

In order to visualize the relationship between image complexity and segmentation performance, Figure 8 shows a scatterplot of the mean Dice coefficient metrics and the entropy metrics for each used dataset. Each subplot corresponds to a different network (Baseline U-Net vs. U-Net++), and the graphs emphasise some of the previous findings: (i) the gap in performance between the OCT and SLO datasets for baseline U-Net, with performance reducing substantially for greater Shannon (pixel level) entropy and Delentropy (global level) (left plot), which is greatly reduced when using the U-Net++ model with the InceptionResnetv2 backbone (right plot), (ii) a greater difference across OCT and SLO datasets is shown for the Shannon (pixel level) entropy and Delentropy (global level), with the highest entropy values, and lower Dice coefficient belonging to the SLO datasets, and (iii) different behaviour for the GCLM (texture) entropy, with lower differences across datasets.

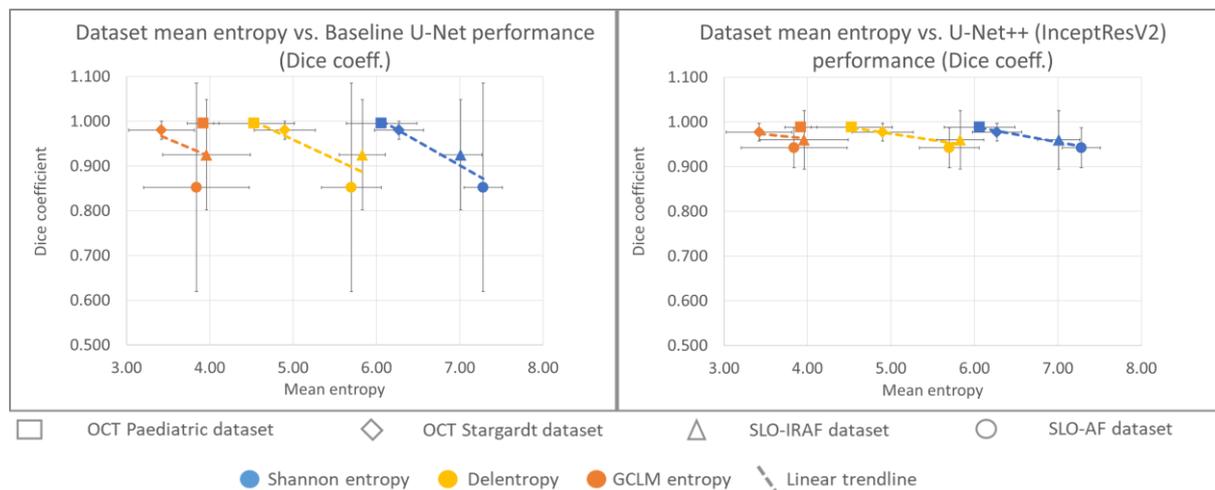


Figure 8. Graphical representation for the baseline U-Net (left plot) and U-Net++ with InceptionResNetV2 backbone (right plot) performance (Dice coefficient) in relation with the entropy metrics (represented with different colours) for each of the four tested datasets (each represented with a different symbol). Error bars describe the standard deviation.

5. Conclusions

Since its inception, deep learning models have become widely used for medical imaging segmentation tasks, with a wide range of applications such as segmentation of cell nuclei, cancer tissue segmentation, cardiac structures and vessels, and tumour detection, among others [27]. Hence, a rapid development of models such as U-Net-based architectures have occurred, with the use of more complex and sophisticated models implemented in medical imaging. In the field of ophthalmic imaging segmentation, most of the studies have proposed the use of U-Net models and their variations, obtaining promising results. Despite this, selecting the optimal model for a particular task within the larger range of models available is not simple. Besides the DL architecture used for the segmentation task, it is important to consider other factors that may affect the obtained results, such as the characteristics of the dataset, which may present images with more complex features that need to be “learned” by the applied segmentation method.

In this study, three DL segmentation models were compared, analysing how they perform when segmenting the retinal and choroidal tissues from OCT images, and in the segmentation of changes associated with retinitis pigmentosa in SLO images. The results from this study demonstrate that not every model behaves the same way when trained for different segmentation tasks or imaging modalities. For the retinal and choroidal tissue segmentation using OCT images, the baseline U-Net provided a good segmentation performance for both healthy eyes and for eyes with Stargardt disease. This finding supports previous studies [3] suggesting that the use of deeper (more complex model) may not be needed for OCT segmentation since the performance is similar to a baseline U-Net architecture. On the other hand, for the SLO dataset, this same model produced the lowest segmentation performance, whereas using the U-Net++ model with a backbone (InceptionResNetV2) provided better results, showing that a deeper model was needed for this dataset, which is visually more complex.

To explore the concept of dataset complexity, an additional image complexity analysis was performed, calculating the image entropy at three different levels of abstraction for each of the datasets. This analysis showed that at pixel and global level, the images in the SLO datasets presented more complex features than OCT images, which are related to the difficulty involved in the learning of top and bottom layers of the model for the segmentation task. Thus, it is likely that the baseline U-Net model, which presents a shorter depth, is not deep enough to learn the more complex features of the SLO images, obtaining the lowest performance. It is worth noting that SLO and OCT are just two of the commonly used ophthalmic imaging modalities. Further exploration of this concept of the data/DL-model complexity link could be valuable, either by expanding the method to other data formats (enface OCT analysis), by investigating alternative imaging modalities such as OCT-A or ultrasound, or by exploring inter/intra device/manufacture effects.

Although this analysis should be extrapolated to more datasets, the results obtained from this study allow us to better understand how image complexity measures such as Shannon entropy and Delentropy could be used as a proxy for model selection and its relation to segmentation performance when using OCT and SLO images. Given the large number of DL models available, image complexity metrics may provide an indication for the need for more complex segmentation models. Thus, the application of these metrics could be useful in the field and worth

exploring further. This may provide more information, in future research studies and better strategies for deciding which segmentation model to use with a given dataset for a specific₇₃₅ segmentation approach.

Author Contributions

I.A.V. conceptualization, methodology, software, writing—original draft preparation, D.A.-C. conceptualization, supervision, methodology, writing—reviewing and editing, S.A.R. investigation, supervision, funding, data collection, data curation, writing—reviewing and editing, J.C. investigation, validation, writing—reviewing and editing, F.K.C. data collection, funding, data curation, writing—reviewing and editing, D.A.M. data collection, funding, data curation, writing—reviewing and editing, M.J.C. investigation, supervision, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

DA-C receives funding from the National Health and Medical Research Council [Ideas Grant NHMRC 1186915]. FKC receives funding from the National Health and Medical Research Council [Centre of Research Excellence Grant GNT1116360 and Fellowship MRF1142962] and WA Health Medical and Research Infrastructure Fund. DA-C and JC received funding from the Ophthalmic Research Institute of Australia (ORIA).

Institutional Review Board Statement

Paediatric OCT dataset: Approval from the Queensland University of Technology human research ethics committee was obtained before commencement of the study. Stargardt OCT dataset: Approval to identify and use SD-OCT images from patients with genetically confirmed Stargardt disease for developing segmentation methods was obtained from the Human Ethics Office of Research Enterprise, The University of Western Australia (RA/4/1/8932 and RA/4/1/7916) and the Human Research Ethics Committee, Sir Charles Gairdner Hospital (2001-053). SLO Dataset: This study adhered to the tenets of the Declaration of Helsinki. Ethics approval was obtained from the human ethics committee of the University of Western Australia (2021/ET000251) and the Human Research Ethics Committee, Sir Charles Gairdner Hospital (2001-053).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The dataset consisting of medical images is not publicly available at the time of writing due to institutional restrictions.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348.
2. Ulku, I.; Akagunduz, E. A survey on deep learning-based architectures for semantic segmentation on 2D images. *Appl. Artif. Intell.* **2021**, *36*, e2032924.
3. Kugelman, J.; Allman, J.; Read, S.A.; et al. A comparison of deep learning U-Net architectures for posterior segment oct retinal layer segmentation. *Sci. Rep.* **2022**, *12*, 14888.
4. Venhuizen, F.G.; van Ginneken, B.; Liefers, B.; et al. Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. *Biomed. Opt. Express* **2017**, *8*, 3292–3316.
5. Hassan, T.; Usman, A.; Akram, M.U.; et al. Deep learning based automated extraction of intra-retinal layers for analyzing retinal abnormalities. In Proceedings of the 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), Ostrava, Czech Republic, 17–20 September 2018.
6. Kiaee, F.; Fahimi, H.; Rabbani, H. Intra-retinal layer segmentation of optical coherence tomography using 3D fully convolutional networks. In Proceedings of the International Conference on Image Processing, ICIP, Athens, Greece, 7–10 October 2018; pp. 2795–2799.

7. Liu, Y.; Ren, G.; Yang, G.; et al. Fully convolutional network and graph-based method for co-segmentation of retinal layer on macular OCT images. In Proceedings of the 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 3081–3085.
8. Pekala, M.; Joshi, N.; Liu, T.A.; et al. Deep learning based retinal OCT segmentation. *Comput. Biol. Med.* **2019**, *114*, 103445.
9. Ruan, Y.; Xue, J.; Li, T.; et al. Multi-phase level set algorithm based on fully convolutional networks (FCN-MLS) for retinal layer segmentation in SD-OCT images with central serous chorioretinopathy (CSC). *Biomed. Opt. Express* **2019**, *10*, 3987–4002.
10. Apostolopoulos, S.; De Zanet, S.; Ciller, C.; et al. Pathological OCT retinal layer segmentation using branch residual U-shape networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017; pp. 294–301.
11. He, Y.; Carass, A.; Yun, Y.; et al. Towards topological correct segmentation of macular OCT from cascaded FCNs. *Lect. Notes Comput. Sci.* **2017**, *10554*, 202–209.
12. Roy, A.; Conjeti, S.; Karri, S.; et al. RelayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed. Opt. Express* **2017**, *8*, 3627–3642.
13. Hamwood, J.; Alonso-Caneiro, D.; Read, S.; et al. Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers. *Biomed. Opt. Express* **2018**, *9*, 3049–3066.
14. Kugelman, J.; Alonso-Caneiro, D.; Read, S.A.; et al. Automatic choroidal segmentation in OCT images using supervised deep learning methods. *Sci. Rep.* **2019**, *9*, 13298.
15. Gao, K.; Niu, S.; Ji, Z.; et al. Double-branched and area-constraint fully convolutional networks for automated serous retinal detachment segmentation in SD-OCT images. *Comput. Methods Programs Biomed.* **2019**, *176*, 69–80.
16. Li, M.-X.; Yu, S.-Q.; Zhang, W.; et al. Segmentation of retinal fluid based on deep learning: Application of three-dimensional fully convolutional neural networks in optical coherence tomography images. *Int. J. Ophthalmol.* **2019**, *12*, 1012.
17. Lee, C.S.; Tying, A.J.; Deruyter, N.P.; et al. Deep learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed. Opt. Express* **2017**, *8*, 3440–3448.
18. Tennakoon, R.; Gostar, A.K.; Hoseinnezhad, R.; et al. Retinal fluid segmentation in OCT images using adversarial loss based convolutional neural networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1436–1440.
19. Kang, S.H.; Park, H.S.; Jang, J.; et al. Deep neural networks for the detection and segmentation of the retinal fluid in OCT images. In Proceedings of the Retinal OCT Fluid Challenge (RETOUCH), Quebec City, Canada, 10–14 September 2017; pp. 9–14.
20. Cunefare, D.; Fang, L.; Cooper, R.F.; et al. Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. *Sci. Rep.* **2017**, *7*, 6620.
21. Chen, Y.; He, Y.; Wang, J.; et al. Automated cone photoreceptor cell identification in confocal adaptive optics scanning laser ophthalmoscope images based on object detection. *J. Innov. Opt. Health Sci.* **2022**, *15*, 2250001.
22. Chen, Y.; He, Y.; Wang, J.; et al. Deeplab and bias field correction based automatic cone photoreceptor cell identification with adaptive optics scanning laser ophthalmoscope images. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 2034125.
23. Hamwood, J.; Alonso-Caneiro, D.; Sampson, D.M.; et al. Automatic detection of cone photoreceptors with fully convolutional networks. *Transl. Vis. Sci. Technol.* **2019**, *8*, 10.
24. Zhang, Q.; Sampani, K.; Xu, M.; et al. AOslo-net: A deep learning-based method for automatic segmentation of retinal microaneurysms from adaptive optics scanning laser ophthalmoscopy images. *Transl. Vis. Sci. Technol.* **2022**, *11*, 7.
25. Sułot, D.; Alonso-Caneiro, D.; Ksieniewicz, P.; et al. Glaucoma classification based on scanning laser ophthalmoscopic images using a deep learning ensemble method. *PLoS ONE* **2021**, *16*, e0252339.
26. Viedma, I.A.; Alonso-Caneiro, D.; Read, S.A.; et al. Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey. *Neurocomputing* **2022**, *507*, 247–264.
27. Siddique, N.; Paheding, S.; Elkin, C.P.; et al. U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057.
28. Du, G.; Cao, X.; Liang, J.; et al. Medical image segmentation based on U-Net: A review. *J. Imaging Sci. Technol.* **2020**, *64*, 1–12.
29. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; pp. 424–432.
30. Oktay, O.; Schlemper, J.; Folgoc, L.L.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
31. Alom, M.Z.; Hasan, M.; Yakopcic, C.; et al. Recurrent residual convolutional neural network based on U-Net (r2U-Net)

- for medical image segmentation. *arXiv* **2018**, arXiv:1802.06955.
32. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; et al. Unet++: A nested U-Net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Granada, Spain, 20 September 2018; pp. 3–11.
 33. Rahane, A.A.; Subramanian, A. Measures of complexity for large scale image datasets. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC), Fukuoka, Japan, 19–21 February 2020; pp. 282–287.
 34. Khan, T.M.; Naqvi, S.S.; Meijering, E. Leveraging image complexity in macro-level neural network design for medical image segmentation. *Sci. Rep.* **2022**, *12*, 22286.
 35. Chen, L.-C.; Papandreou, G.; Schroff, F.; et al. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
 36. Read, S.A.; Collins, M.J.; Vincent, S.J.; et al. Choroidal thickness in myopic and nonmyopic children assessed with enhanced depth imaging optical coherence tomography. *Investig. Ophthalmol. Vis. Sci.* **2013**, *54*, 7578–7586.
 37. Kugelmann, J.; Alonso-Caneiro, D.; Chen, Y.; et al. Retinal boundary segmentation in stargardt disease optical coherence tomography images using automated deep learning. *Transl. Vis. Sci. Technol.* **2020**, *9*, 12.
 38. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
 39. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
 40. He, K.; Zhang, X.; Ren, S.; et al. 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.
 41. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; et al. Inception-v4, inception-resnet and the impact of residual connections on learning. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*, 4278–4284.
 42. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
 43. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867.
 44. Huang, G.; Liu, Z.; Van Der Maaten, L.; et al. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
 45. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; et al. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
 46. Holschneider, M.; Kronland-Martinet, R.; Morlet, J.; et al. A real-time algorithm for signal analysis with the help of the wavelet transform. In Proceedings of the Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, 14–18 December 1987; Springer: Berlin/Heidelberg, Germany, 1990; pp. 286–297.
 47. Sermanet, P.; Eigen, D.; Zhang, X.; et al. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv* **2013**, arXiv:1312.6229.
 48. Papandreou, G.; Kokkinos, I.; Savalle, P.-A. Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 390–399.
 49. Giusti, A.; Cires, D.C.; Masci, J.; et al. Fast image scanning with deep max-pooling convolutional neural networks. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2013; pp. 4034–4038.
 50. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848.
 51. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916.
 52. Iakubovskii, P. Segmentation Models Pytorch. 2019. Available online: https://github.com/qubvel/segmentation_models.pytorch (accessed on 27th of March 2025).