IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 4, 2020, accepted February 22, 2020, date of publication March 2, 2020, date of current version March 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977355

Effect of Altered OCT Image Quality on Deep Learning Boundary Segmentation

JASON KUGELMAN[1], DAVID ALONSO-CANEIRO[1,2], SCOTT A. READ[1], STEPHEN J. VINCENT[1], FRED K. CHEN[2,3], AND MICHAEL J. COLLINS[1]

[1]Contact Lens and Visual Optics Laboratory, School of Optometry and Vision Science, Queensland University of Technology, Brisbane, QLD 4000, Australia
[2]Centre for Ophthalmology and Visual Science (incorporating Lions Eye Institute), The University of Western Australia, Perth, WA 6009, Australia
[3]Department of Ophthalmology, Royal Perth Hospital, Perth, WA 6000, Australia

Corresponding author: David Alonso-Caneiro (d.alonsocaneiro@qut.edu.au)

This work was supported in part by the Rebecca L. Cooper 2018 Project Grant (DAC), in part by the Perth Children's Hospital-Telethon Grant (FKC, DAC), in part by the NHMRC/MRFF Career Development Fellowship (FKC) under Grant APP1142962, and in part by the MD Foundation Australia (FKC).

ABSTRACT Deep learning methods provide a platform to segment boundaries within the retina and choroid in OCT images of the posterior eye, with the ultimate goal of having a robust model that works well across a wide range of different datasets. However, since most studies of deep learning methods use datasets exhibiting similar image quality for both training and evaluation, the effect of varied image quality on such methods is not normally explored in the context of OCT image segmentation. An understanding of the effects of image quality factors is vital to determine the robustness of the methods and their ability to be applied in clinical practice where images exhibiting a range of different qualities are encountered. This study examined a range of factors that can affect standard OCT image quality and determined how and why the performance of an existing neural network based segmentation method can subsequently degrade as a result. Three image quality factors (noise, contrast reduction, and gamma correction) all had a negative impact upon performance, while more robust performance was maintained in the presence of both JPEG and JPEG2000 image compression. Improving the method's robustness to each of these degradations is also demonstrated with marked performance improvements identified by applying a fine-tuning approach to the network. This study improves our understanding of the effect of OCT image degradation on neural network performance, the effect that fine-tuning with poor-image quality data has on the network and highlights the benefit and importance of training resilient models using data augmentation.

INDEX TERMS Image segmentation, machine learning, optical coherence tomography.

I. INTRODUCTION

Changes in the posterior ocular tissues (e.g. the retina and choroid) occur over the course of the day [1], [2], throughout normal ocular development in childhood [3] and over a lifetime [4], [5], and in association with the development of myopia [3], [6], [7] and numerous posterior eye diseases [8]–[11]. The quantification of these changes is fundamental for both clinical and research tasks. In particular, the segmentation of the chorio-retinal layers and subsequent measurement of layer thickness is often required for comparison to age-matched data or previously measured patient data, for the detection and management of disease and documenting the eye's normal aging changes. In recent years, the introduction of optical coherence tomography (OCT) has allowed for non-invasive, high-resolution images of the retina and choroid to be captured for such analyses. However, the manual image segmentation and analysis of large data sets of OCT images by human experts is a time consuming task, and has necessitated the development of automated analysis methods.

Previously, several methods for the automatic segmentation of chorio-retinal layers have been introduced. Initial methods used standard image processing techniques, which predominantly relied upon handcrafted algorithms following a predefined set of rules. Such algorithms involved: active contours [12], graph theory [13]–[15], energy minimization [16], diffusion maps [17], Chan-Vese models [18],

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

Markov models [19], and kernel regression filtering [20]. A potential downside of many of these approaches is their lack of robustness and generalization, particularly in the presence of noise and artefacts.

To address some shortcomings of these approaches, machine learning (ML) based methods have become common for the automatic segmentation of OCT images. These have included the use of random forests [21], [22] and support vector machines [23]. However, deep learning with neural networks, in particular, is frequently used for OCT image segmentation. A number of different methods have been utilized including patch-based classification [24]–[27], semantic segmentation [28]–[34], adversarial learning [35], and transfer learning [36]. Additionally, some methods [37], [38] have used volumetric input data, consisting of multiple image slices instead of a standard single image. The performance of these methods is heavily dependent on the quality of the image data used for training. In particular, it is important that the training data is representative of the testing data to ensure a correct and fair evaluation of machine learning method performance.

Several prior studies, not involving ocular OCT image analysis, have examined the effect of various image quality degradations and their effect on the performance of machine learning models [39]–[44]. Dodge and Karam [41] investigated the effect of image quality distortions (blur, contrast, noise, JPEG compression) on the performance of existing neural networks. Using a subset of the ImageNet 2012 dataset, they showed that the networks were susceptible to quality distortions, particularly blur and noise. Similarly, Roy *et al.* [42] experimented with a number of types of degradation including three types of noise (Gaussian, colored Gaussian, salt and pepper), two types of blur (motion, Gaussian) as well as JPEG compression. Using support vector machines, da Costa *et al.* [43] analyzed the effects of different types of noise (Gaussian, Poisson, salt & pepper) on the Corel and Caltech101-600 datasets. Likewise, Franceschi *et al.* [44] examined the robustness of various classifiers to both uniform and Gaussian noise. Overall, these studies found that such image quality degradations had a negative impact on the performance of machine learning models.

Some work has also examined the effect of image quality on the performance of non-machine learning retinal segmentation algorithms for ocular OCT images. Somfai *et al.* [45] reported that the performance of their retinal segmentation algorithm, incorporating standard image processing techniques, degraded in the presence of OCT image artefacts such as defocus, depolarization and decentration. In another study analyzing the effect of scanning distance on OCT image analysis, Varga *et al.* [46] showed that using a sub-optimal scanning distance increased retinal boundary detection error. Speckle noise and artefacts also affect OCT image quality and the subsequent efficiency and accuracy of segmentation algorithms [14], [47]. Balasubramanian *et al.* [48] analyzed the effect of image quality on tissue thickness measurements obtained directly from three different SD-OCT instruments and noted that the presence of noise in low quality images may negatively affect the performance of a segmentation algorithm.

To compensate for poor image quality, a few studies in the context of deep learning for OCT image segmentation perform preprocessing on the images for de-noising or contrast enhancement purposes. For instance, image filtering [14] or image registration and scan averaging [47], can be used to reduce speckle noise in OCT images. However, most deep learning segmentation studies do not perform any pre-processing leaving many of these methods susceptible to reduced performance in the presence of poor quality data. Additionally, the effect of image compression on performance has not been explored previously.

Thus, there is no existing analysis of the impact of image quality factors on OCT boundary segmentation with deep learning methods and, as a result, the effect of the relevant image degradations is poorly understood. An examination of performance in the presence of degraded image quality is important to evaluate the generalizability of such deep learning methods and their ability to be applied as useful segmentation tools in clinical practice. By using a previously proposed network and methodology, which has demonstrated good performance on OCT chorio-retinal layer segmentation [28], this study aims to focus exclusively on analyzing the effect that OCT image quality degradation has on segmentation performance. The contributions of this paper are as follows:

- We provide a comprehensive simulation of the effects of degraded OCT image quality and present a detailed analysis for the segmentation performance of testing on such data, for a range of image quality factors and parameters.
- Using fine-tuning, we demonstrate the feasibility of improving the method's robustness in the presence of variable image quality data. Thus, providing a demonstration of training regimes to provide robust deep learning OCT segmentation methods.
- The results of network changes provide a greater understanding of the underlying neural network behavior when exposed to OCT image degradations.

II. METHODS

A. DATA

A dataset consisting of SD-OCT images from healthy eyes was used for this study and has been described in detail elsewhere [49]. The data consists of images from 101 healthy participants collected at four visits over an 18-month period. At each visit, two sets of six cross-sectional foveal-centered radial chorio-retinal scans were acquired. Each image has a pixel dimension of 1536 $\times$ 496 (width $\times$ height) which is approximately equivalent to a real physical dimension of 8.8 $\times$ 1.9 mm (vertical scale of 3.9 μm per pixel and a horizontal scale of 5.7 μm per pixel). The images were acquired using the Heidelberg Spectralis (Heidelberg

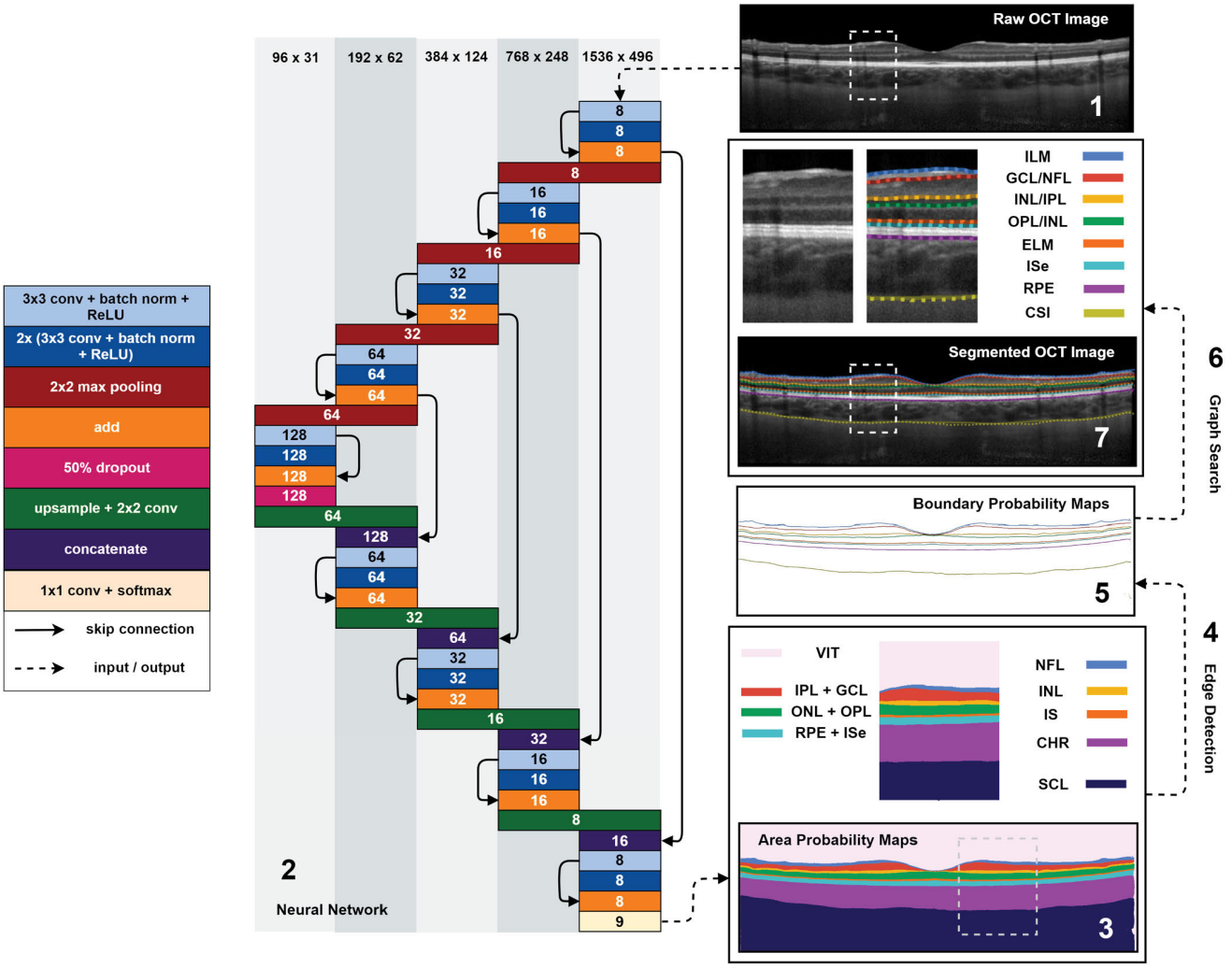


FIGURE 1. Overview of the general method and the neural network model, including an example of the use of the trained model output for segmentation of OCT chorio-retinal layers. Bold numbers (1-7) and dashed arrows indicate the order that the steps are performed. 1: Raw OCT image (grayscale 1536 × 496) input. 2: Neural network architecture. 3: Area probability maps output from network. Grey dashed box corresponds to cropped region of interest. Each area is drawn in a different color. 4: Edge detection performed using Sobel filter. 5: Probability maps constructed for each boundary. 6: Shortest path graph search performed on each boundary probability map. 7: Image segmented and error calculated. Grey dashed boxes correspond to cropped region of interest. For visualization, dashed lines (predictions) and solid lines (ground truths) are overlaid on the image. Each boundary is drawn in a different color.

Engineering, Heidelberg, Germany) SD-OCT instrument with the Enhanced Depth Imaging mode enabled. Automatic real time tracking was utilized, with 30 frames averaged for each image to improve the signal to noise ratio. Due to the similarity and redundancy of data between visits and sets for each participant, only data from the first visit and from one set is used. Segmentations, labelled manually by a trained observer, are provided for the eight chorio-retinal layer boundaries including the chorio-scleral interface (CSI), the outer boundary of the retinal pigment epithelium (RPE), the inner boundary of the inner segment ellipsoid zone (ISe), the inner boundary of the external limiting membrane (ELM), the boundary between the outer plexiform layer and inner nuclear layer (OPL/INL), the boundary between the inner nuclear layer and the inner plexiform layer (INL/IPL), the boundary between the ganglion cell layer and the nerve fiber layer (GCL/NFL), and the inner boundary of the inner limiting membrane (ILM).

Throughout this work, a training dataset, a validation dataset, and a testing dataset are used, each consisting of a subset of these OCT scans (with different individual subjects across the different dataset). The training dataset contains 240 images from 40 randomly selected participants (6 scans from each) while the testing dataset contains 294 images from 49 randomly selected participants (6 scans from each) and the validation dataset contains 60 images from 10 randomly selected participants (6 scans from each). A roughly 40/10/50 split, chosen empirically, is used between the training, validation and testing sets as a trade-off between providing sufficient training data, calculating representative testing

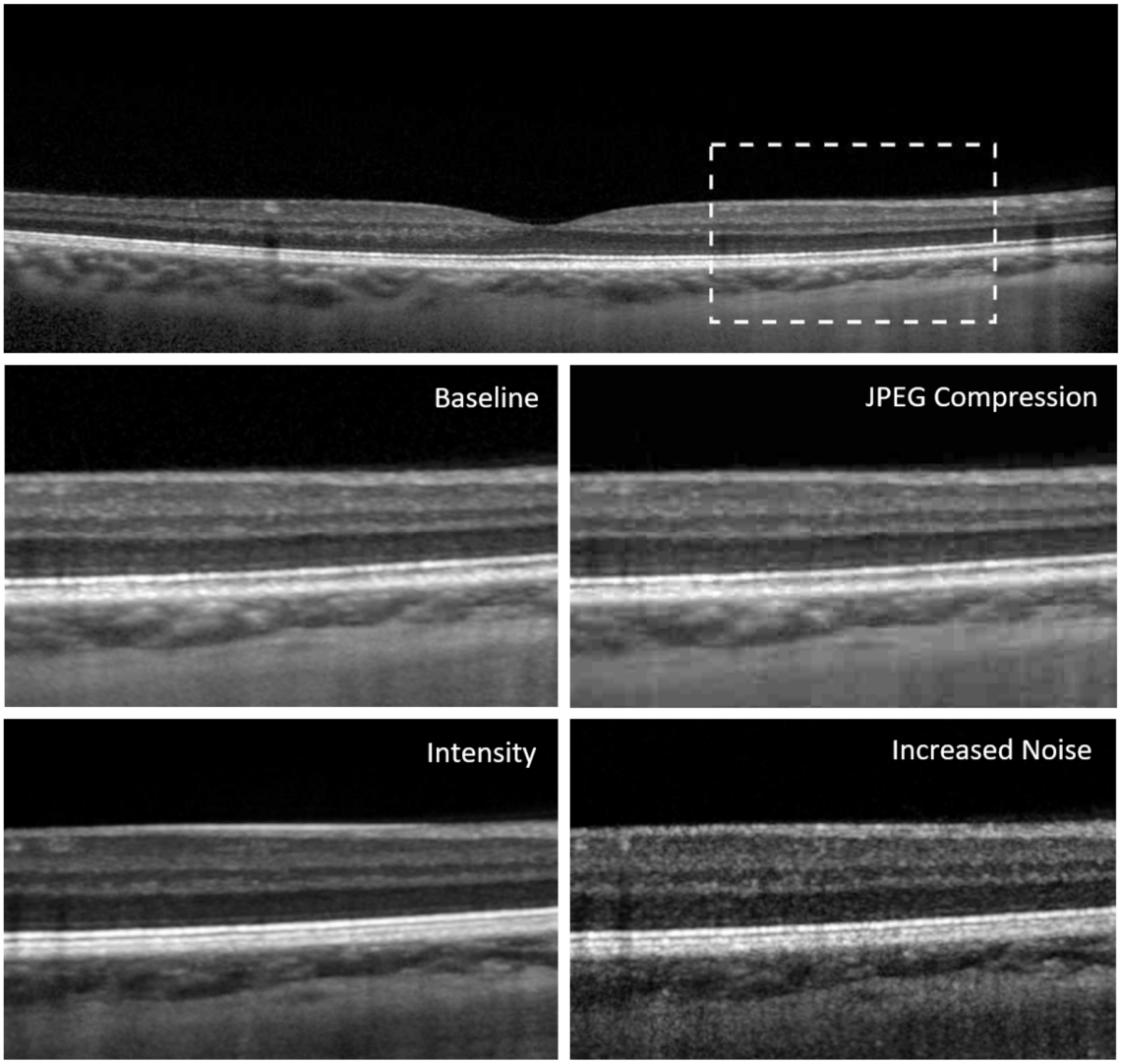


FIGURE 2. Example of image quality factors evident in real OCT data. The white dashed box in the top image indicates the common region of interest corresponding to the four subplots. The "Increased Noise" subplot demonstrates the effect of scan averaging with a higher level of noise present compared to the baseline as a result of fewer averaged scans. The "Intensity" subplot depicts the effect of image acquisition conditions resulting in a change in the intensity/contrast of the baseline image. The "JPEG compression" subplot shows the baseline image exported from the instrument using the lossy JPEG file format as opposed to a lossless bitmap.

results and performing accurate model selection. An individual participant's scans were not included in more than one of the datasets and with six different radially oriented scans per participant, there is an equal distribution of scan orientations within each dataset. All images are grayscale with an intensity range of 0-255 (inclusive). The manufacturer's mean image quality for the training, validation and testing datasets were 32.3 dB, 32.1 dB and 32.5 dB respectively. Overall, the image quality was not statistically significantly different between the three datasets (all $p>0.05$).

B. SEGMENTATION METHOD

The segmentation method employed in this work consists of a two-stage process with the objective of delineating eight chorio-retinal layers. The first stage involves training a fully-convolutional neural network for semantic segmentation. Here, the input is an OCT image and the outputs/labels are area probability maps. These areas are defined between the respective layer boundaries and the edges of the image. By applying the Sobel filter, edge detection is performed on each area probability map to extract the corresponding boundary probability map. Using this acquired map, a shortest path graph search is then performed to delineate the position of each boundary. The graph search component of the method has been described in detail previously [25]. By comparing the predicted boundary position to the actual manually segmented boundary position provided in the data, the overall performance on an image is evaluated by measuring the mean absolute error (MAE) (in pixels) for each layer

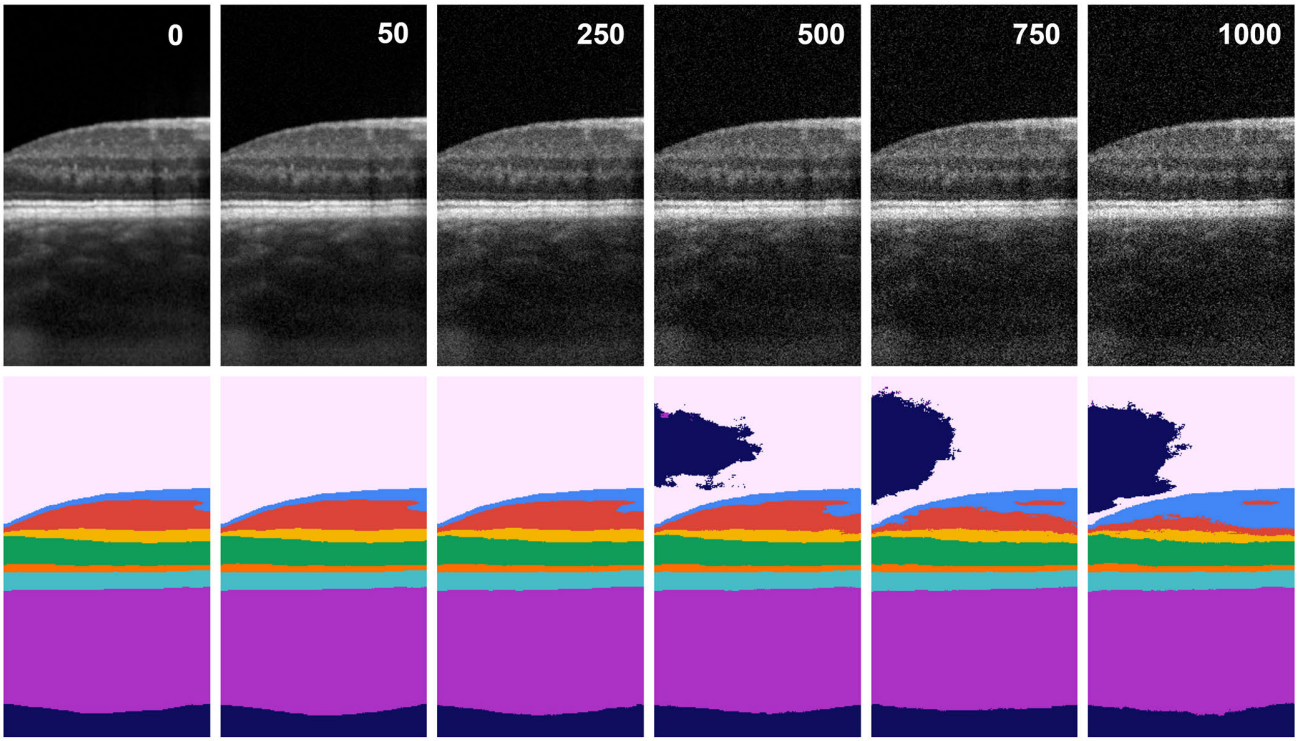


FIGURE 3. Example of performance degradation with the original networks (no-fine-tuning) as a result of Gaussian noise (variances of 0, 50, 250, 500, 750, 1000) applied to a section of an OCT image. Here the raw images correspond to a variance of 0. Top: augmented OCT images. Bottom: corresponding predictions.

boundary. For the whole testing dataset, the overall median absolute error (median mean absolute error) is used to assess performance. Here, the median is used to prevent outliers from skewing the results. An illustration of the second stage process to segment an OCT image using a trained network is shown in Fig. 1.

C. NEURAL NETWORK MODEL

The neural network architecture examined here is similar to that presented previously for OCT chorio-retinal layer segmentation [28] which has shown promising results for this problem. This network is inspired by the U-net architecture [50] for medical image segmentation exhibiting an encoder-decoder structure with skip connections between the contracting and expanding path. However, unlike the U-net, residual learning [51] is also incorporated. The network consists of four layers with convolutional blocks containing eight filters in the first layer. The number of filters is doubled at each subsequent pooling layer to allow a greater level of feature encoding. In an effort to improve training performance and assist convergence, batch normalization [52] is used within each convolutional block while 50% dropout [53] is added to the bottleneck (between the encoder and decoder) of the network to prevent overfitting. The details of the overall structure of the network are illustrated in Fig. 1.

The network was trained for 100 epochs using the training dataset (240 images) while validation was performed using the validation dataset (60 images). A batch size of three was chosen as a trade-off between computational resources and training speed. Training was performed using the Adam optimizer [54], and all parameters were set at their recommended default values ($\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \times 10^{-8}$). This optimizer was chosen based on its documented performance for training deep neural networks [54]. The loss function utilized here consists of the unweighted sum of binary cross-entropy loss and dice loss. This is similar to the weighted sum approach used in previous work [28]. For model selection purposes, the epoch with greatest dice overlap percentage on the validation dataset was selected. The programming environment used for this study consisted of Keras 2.2.4 [55] with Tensorflow [56] backend in Python 3.6.4. To examine any variability associated with random initialization, four networks were trained identically with the exception of their initial random weights.

D. IMAGE QUALITY FACTORS

In this work, we aimed to examine how various image quality factors affect the performance of a neural network trained purely using raw images. It is worth noting that these ''raw images'' represent an equivalent ''optimal'' image quality as these were acquired using frame-averaging to reduce noise, data was exported using a lossless image format and no ocular pathologies were present. Using the segmentation method and data outlined above, each trained neural network is

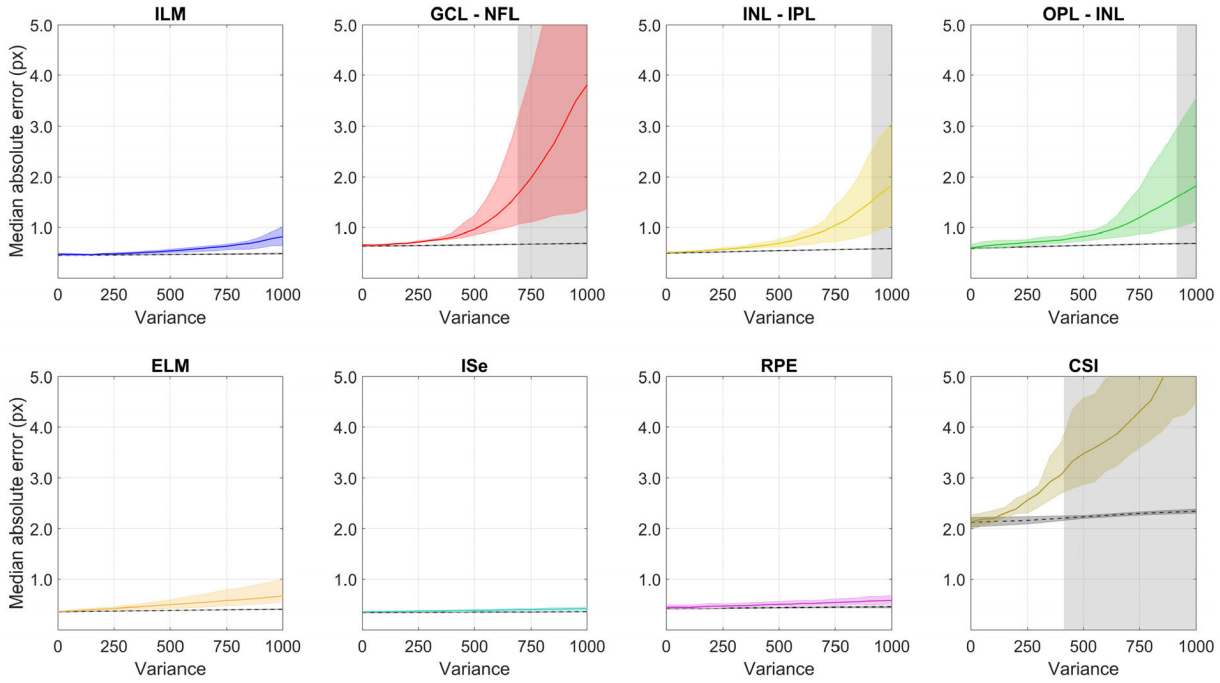


FIGURE 4. Median mean absolute error (pixels) for each boundary of interest as Gaussian noise is applied to the test images. The solid colored line in each plot represents the mean of the four original networks (no fine-tuning) while the surrounding shaded region depicts the range of median errors exhibited by these networks. Likewise, the dashed grey line and corresponding shaded region represents the mean and range respectively of the networks fine-tuned using Gaussian noise augmented data. Here, the raw images (baseline performance) correspond to a variance of 0. The grey shaded rectangular region indicates where the mean error (for the original networks) becomes practically significant (+1 pixel error compared to baseline).

used to evaluate the raw testing images in addition to these same images augmented with four different image quality factors. Here, the effect of image compression (JPEG and JPEG2000), noise, and image intensity (contrast reduction and gamma correction) are considered. Fig. 2 provides an example of such image quality factors evident in real OCT data. Note that the networks used for evaluations are trained on just the raw training data (''optimal'' quality with no degradation).

The output file format of OCT instruments and related software may vary. For example, JPEG is one of the data export options for the instrument used in this study. It is therefore important to evaluate the effect on performance for the case where a lossy image format, such as JPEG, may be present. To do this, image compression may be applied to the lossless testing images and evaluated using the network trained on lossless images. In this work, two image compression algorithms are utilized: JPEG and JPEG2000. The standard JPEG compression algorithm is applied to an image using an image quality parameter with lower values corresponding to a higher level of compression but decreased image quality. Each of the four identically trained networks were used to evaluate the testing dataset with each image augmented by applying JPEG compression. In this work, image quality values from 2 to 50 (inclusive) in steps of 2 are considered for experimentation. Image quality values above 50 are not considered as there is little to no performance degradation present in this range. JPEG2000 compression is applied to an image using a compression ratio parameter. This parameter determines the ratio of the original file size to the compressed size with higher values corresponding to a higher level of compression but decreased image quality. The compression ratios used in the experiments throughout include values of the form 2^x with x taking all integer values in the range 1 to 7 (inclusive). Additionally, values of the form $128 + 32x$ are also considered with x taking all integer values in the range 1 and 12 (inclusive). Values above 512 are not considered as degradation is already well advanced at this point. The other parameters for the JPEG2000 algorithm are held constant throughout including the number of quality layers and the number of reduction (wavelet decomposition) levels which are set at 1 and 4 respectively.

OCT scans acquired under sub-optimal conditions (e.g. poor image alignment and focus or incorrect scanning distance) or in the presence of various ocular pathologies can affect the image intensity and contrast. For example, the presence of a cataract can cause light scattering which can affect the intensity and contrast of whole or part of an OCT scan [57]. Due to the low level of contrast associated with some OCT scans, it is important to examine how the trained network generalizes in such cases. Here, contrast reduction is applied to the testing images by blending them with a constant intensity grayscale image. Here, a blending factor (alpha) [0-1] is used to determine the weighting of

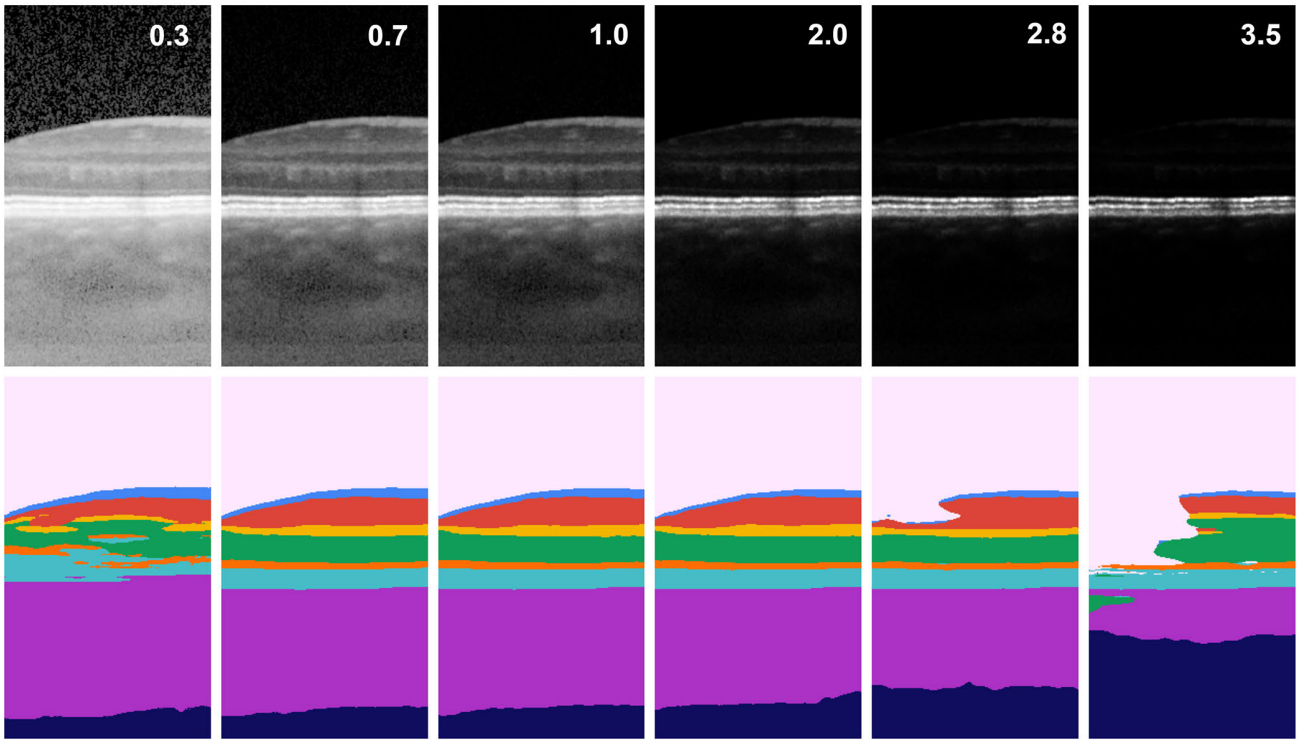


FIGURE 5. Example of performance degradation with the original networks (no fine-tuning) as a result of gamma correction (gamma values of 0.3, 0.7, 1.0, 2.0, 2.8, 3.5) applied to an OCT image. Here, the original images correspond to a gamma of 1.0. Top: augmented OCT images. Bottom: corresponding network predictions.

the grayscale image [41]. For example, a blending factor of 0.1 indicates that, for each pixel, 10% of the constant grayscale image intensity and 90% of the raw OCT image intensity is combined to produce the output. To ensure that the effect of contrast reduction is tested in isolation by avoiding any overall brightening or darkening, the intensity of the grayscale image is set to be equal to the average intensity of the raw OCT image on an image-by-image basis. Each of the four identically trained networks were used to evaluate the testing dataset with each image augmented by applying contrast reduction. For this study, blending factors/alpha values of 0.01 to 0.60 (inclusive) in steps of 0.01 are considered. Blending factors greater than 0.60 were not considered given the severe level of degradation already present at this point. With respect to image intensity, differences in the gamma of OCT images is also a factor that may be considered. To evaluate the effect of this, gamma correction is applied to the normalized [0-1] raw OCT images by raising the input image intensities to a specified power (gamma). Each of the four identically trained networks were used to evaluate the testing dataset with each image augmented by applying gamma correction. The gamma values used in the experiments here include values of 0.3 to 3.5 (inclusive) in steps of 0.1 as well as values of the form $1/x$ with x taking values from 1.1 to 3.0 (inclusive) in steps of 0.1. Gamma values outside of this range (< 0.3 and > 3.5) are not considered as degradation is already severe at these points.

Speckle noise in OCT images, a property inherent to coherent imaging, commonly degrades the image quality as well as the accuracy of subsequent analysis [58]. Additionally, poor fixation and various ocular pathologies can exacerbate this. One option to help reduce such noise is to use frame- averaging as is the case with the raw data in this study. However, this is not always possible and is impractical for dense volumetric imaging protocols given the number of scans required. Alternatively, a number of other techniques to minimize speckle noise have also been proposed. Because of this, the noise present in OCT images may vary considerably depending on the instrument, conditions, noise reduction techniques used and any ocular pathologies present. To model noise in our study here we utilize a Gaussian distribution which provides a good approximation for speckle noise after logarithmic compression of the signal [59]. To analyze the network's robustness to speckle noise in this study, noise is introduced to the raw images by adding a random value to each individual pixel. These values are sampled from a Gaussian distribution with a mean of zero and specified variance (> 0) with larger variances corresponding to ''noisier'' images. Each of the four identically trained networks were used to evaluate the testing dataset with each image augmented by applying Gaussian noise. For these experiments, variances of 50 to 1000 (inclusive) in steps of 50 are used. Larger variances (> 1000) are not considered here as significant degradation is already evident at this point.

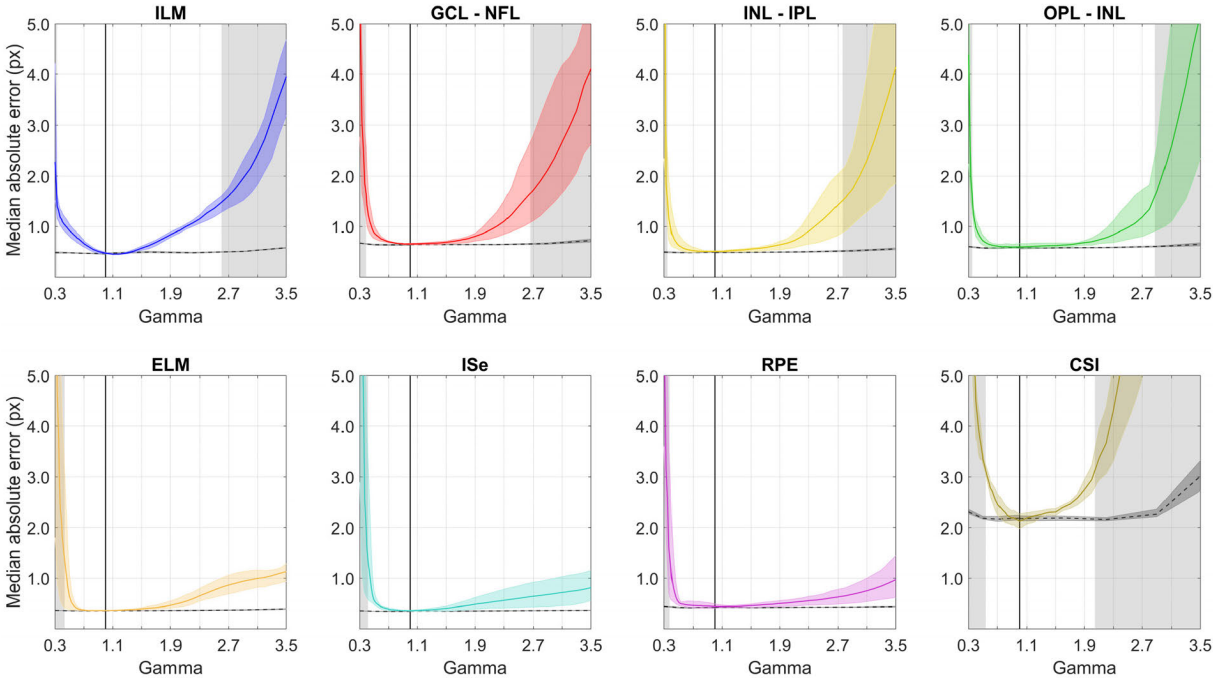


FIGURE 6. Median absolute error (pixels) for each boundary of interest as a result of gamma correction applied to the test images. The solid colored line in each plot represents the mean of the four original networks (no fine-tuning) while the surrounding shaded region depicts the range of median errors exhibited by these networks. Likewise, the dashed grey line and corresponding shaded region represents the mean and range respectively of the networks fine-tuned using data augmented with gamma correction. Here the raw images correspond to a gamma of 1.0 (marked by a solid vertical black line). The grey shaded rectangular regions indicates where the mean error (for the original networks) becomes practically significant (+1 pixel error compared to baseline).

To examine the statistical significance of changes in segmentation performance, a repeated measures ANOVA was performed to assess the effect of each image quality factor. This analysis showed a highly statistically significant degradation of segmentation performance (increased error) for all image quality factors (all $p < 0.001$). In many cases statistical significance was reached with only very low magnitude increases in segmentation error, which may be of limited practical significance for segmentation performance. Therefore, to consider the practical impact of image quality factors on segmentation performance and to allow a standardized comparison across factors and layers, an increase in boundary error of 1 pixel or more above the baseline value for that particular boundary, was chosen in this work to be a threshold representative of a practically significant reduction in segmentation performance.

E. FINE-TUNING METHOD

Increasing the robustness of the networks in the presence of these image degradations is important in practice. One possible method for doing so is to train (or fine-tune) the existing networks on degraded images. In this work by fine-tuning the existing network we aimed to: (i) increase robustness in the segmentation performance by using image-quality degraded augmentations and (ii) understand the network changes after the fine-tuning is applied.

For all experiments, each individual image in an epoch was augmented with the corresponding image quality factor with a 50% probability. Additionally, each image also had a 50% chance to be flipped horizontally, similar to the initial training of the networks. The choice of parameter value for each image quality factor was randomly selected (uniformly from a predefined range or subset) each time for every image. For fine-tuning, the Adam optimizer with default parameters was used with the exception of the learning rate which was instead set an order of magnitude below the default at 0.0001. Note that images for both training and validation are augmented. Due to randomness within training, each network was fine-tuned four times on each image quality factor.

An additional aim of fine-tuning these networks was to further analyze their behavior and gain further insight into the differing performance between layer boundaries, image quality factors and inter-network variability. To achieve this, the original networks and their fine-tuned variants were compared by calculating the mean absolute difference for each set of weights in every convolutional layer. For a fair comparison, the fine-tuned models were compared to their original version after ten epochs to avoid problems such as overfitting and to minimize bias as a result of variable training times.

A gamma correction augmentation was performed by either decreasing or increasing the gamma value. As the two cases may exhibit different behavior it was proposed to treat these separately with gamma < 1 and gamma

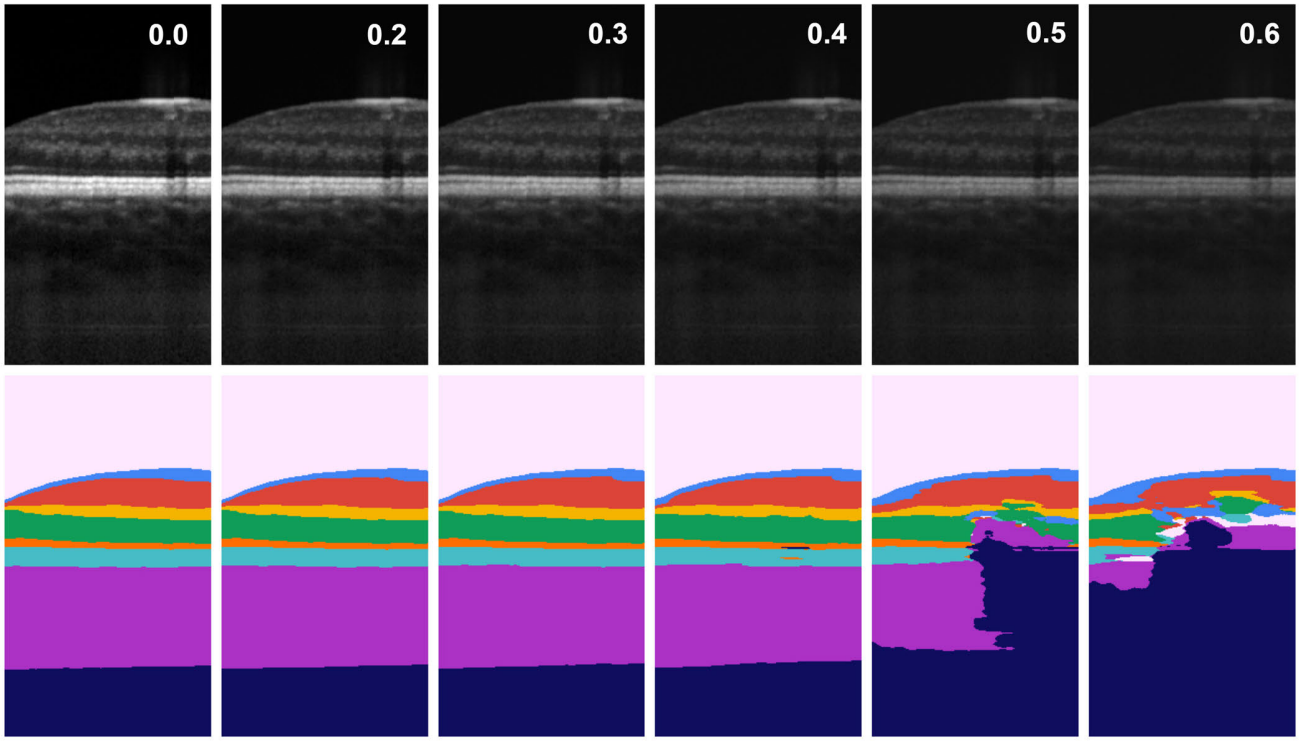


FIGURE 7. Example of performance degradation with the original networks (no-fine-tuning) as a result of contrast reduction (blending factors/alphas of 0.0, 0.2, 0.3, 0.4, 0.5, 0.6) applied to an OCT image. Top: augmented OCT images. Bottom: corresponding predictions.

>1 augmentations. For gamma <1, the value was randomly chosen each time as one of 0.3, 0.5, or 0.7. For gamma >1, each augmentation was selected to be one of 1.6, 2.2, 2.9, 3.2, or 3.5. For Gaussian noise augmentations, the variance was randomly chosen as any decimal value between 50 and 1000 (inclusive). For contrast reduction, the alpha value was randomly chosen as any decimal value between 0.1 and 0.6 (inclusive). For JPEG compression, the image quality value was randomly selected as any integer value in the range 2 to 20 (inclusive) while for JPEG2000 compression, the compression ratio was chosen randomly as one of 128, 256, 384, or 512. These particular sets and ranges of values here were chosen empirically with the goal of improving segmentation performance but do not necessarily represent the most optimal selection.

For analyzing segmentation performance improvement, each fine-tuned network was only evaluated at a subset of the points used previously. This was due to computational reasons as well as the expected performance improvements. The evaluation points were chosen to highlight the improvement in segmentation performance across the whole range but also to emphasize that good baseline performance is retained. Therefore, all fine-tuned networks were evaluated on the baseline raw images. Additionally, for gamma <1, evaluation points included gamma values of 0.3, 0.5 and 0.7 while, in the case of gamma >1, these were 1.6, 2.2, 2.9 and 3.5. Equally spaced points were used for noise (variances of 250, 500, 750 and 1000), contrast (alpha values of 0.15, 0.3, 0.45 and 0.6) and JPEG2000 (compression ratios of 128, 256, 384 and 512). For evaluating using JPEG compression, image quality values of 2, 5, 10 and 50 were selected. Unlike the weight difference comparison, model selection here is performed by choosing the epoch with the highest validation dice overlap percentage.

III. RESULTS AND DISCUSSION

A. OVERVIEW

The baseline performance (evaluation of images with non-altered image quality) demonstrated a high level of consistency across the four networks. Here, a median absolute error below 0.70 pixels for the retinal boundaries and 2.30 pixels for the choroid/scleral interface was obtained, which matches well with previously published results [24]–[26], [28].

Using both the original networks (no fine-tuning) and their fine-tuned variants, the performance was evaluated for each of the image quality factors over their corresponding parameter range. An example from a section of an OCT scan exhibiting performance degradation on one of the original networks is provided for Gaussian noise (Fig. 3), gamma correction (Fig. 5), contrast reduction (Fig. 7), JPEG (Fig. 9), and JPEG2000 (Fig. 11). Likewise, the relationship between the segmentation error (median absolute error) and variance for both the original (solid colored lines) and fine-tuned networks (gray dashed lines) is illustrated for Gaussian noise (Fig. 4), gamma correction (Fig. 6), contrast reduction (Fig. 8), JPEG (Fig. 10) and JPEG2000 (Fig. 12).

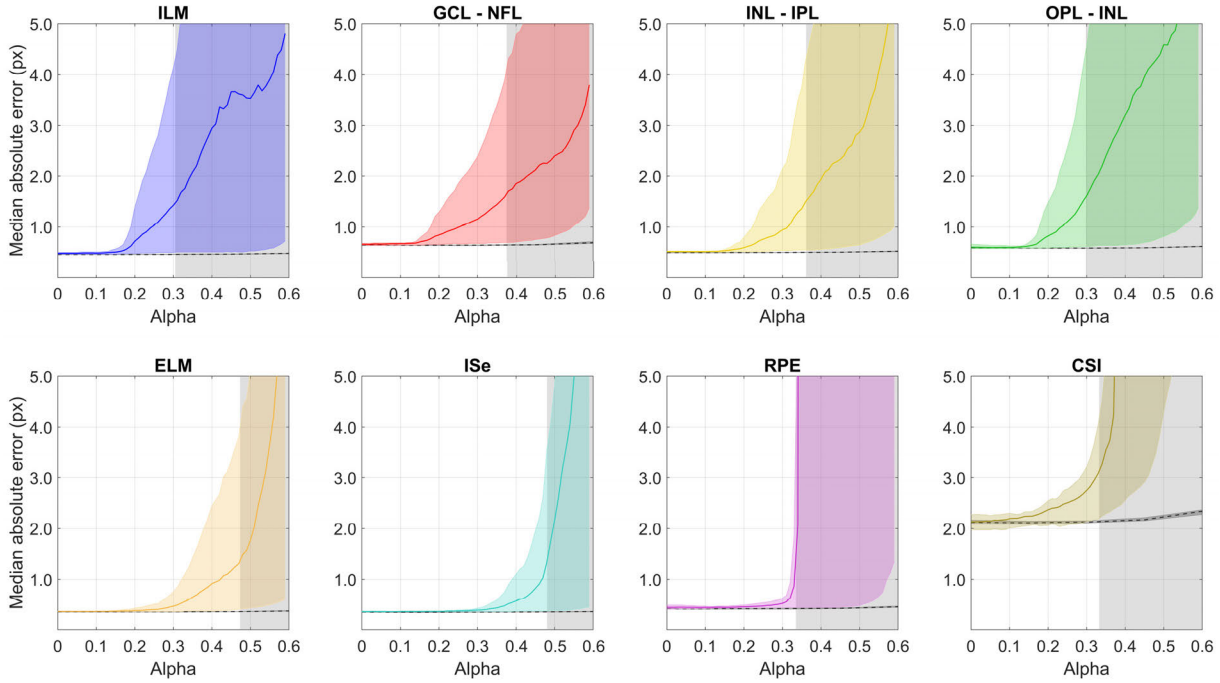


FIGURE 8. Median mean absolute error (pixels) for each boundary of interest as a result of contrast reduction applied to the test images. The solid colored line in each plot represents the mean of the four original networks (no fine-tuning) while the surrounding shaded region depicts the range of median errors exhibited by these networks. Likewise, the dashed grey line and corresponding shaded region represents the mean and range respectively of the networks fine-tuned using data augmented with contrast reduction. Here, the raw images (baseline performance) correspond to a blending factor/alpha value of 0. The grey shaded rectangular region indicates where the mean error (for the original networks) becomes practically significant (+1 pixel error compared to baseline).

B. EFFECT OF ALTERED IMAGE QUALITY

Overall, the results show that three of the image quality factors (Gaussian noise, contrast reduction and gamma correction) appear to have a considerable effect on the segmentation error of the chorio-retinal layer boundaries of interest, with more severe image degradations generally causing larger increases in error as expected. With similar baseline performance, there appears to be a range of values of smaller degradations for which the performance between networks remains similar. However, for larger values of image degradation, the magnitude of these effects generally tend to increase and vary between the four networks. Although the four networks were trained identically and exhibited similar baseline performance, each was initialized with a different set of random weight parameters. As a result, the combination of weight parameters can differ considerably between the four trained networks, which may explain the variable generalization performance in the presence of degrading factors.

In general, the performance degradation (with no fine-tuning) was not similar between boundaries or for the same boundary across different image quality factors. Indeed, the boundaries that reached practical significance varied considerably between the factors. However, the CSI boundary generally appeared to be the least robust to each of the image quality factors and was the only boundary that exhibited a practically significant reduction in performance for all five of the image quality factors. The practically significant threshold of performance degradation for the CSI was also generally reached at a lower level of image degradation than observed for other boundaries. This inferior performance is possibly due to its comparatively lower baseline performance, and often reduced contrast and increased noise in the choroidal region in the raw images.

For Gaussian noise, the performance (no fine-tuning) on four boundaries (ILM, ELM, ISe, RPE) was significantly more robust then for the other boundaries (Fig. 4). Observing Fig. 3, a possible explanation is that these boundaries are located at the transition between layers which possess a higher level of contrast relative to one another. For such layers, it appears that larger changes (greater variance) are needed in the pixel values before the transition between the two layers deteriorates sufficiently. Additionally, it appears that adding noise to the vitreous causes it to more closely resemble the sclera and as a result, the network is prone to misclassifying these regions. This is evident when observing Fig. 3 where part of the vitreous is instead classified as sclera when the variance of noise is increased.

For gamma < 1, the performance on all boundaries deteriorated sharply near the extremity (Fig. 6). This may be explained by analyzing the example in Fig. 5 which clearly shows that, with a gamma correction of 0.3, all layers increase in intensity with the contrast between them significantly reduced. It is likely that the performance degrades because the network has not learnt to distinguish between layers at such

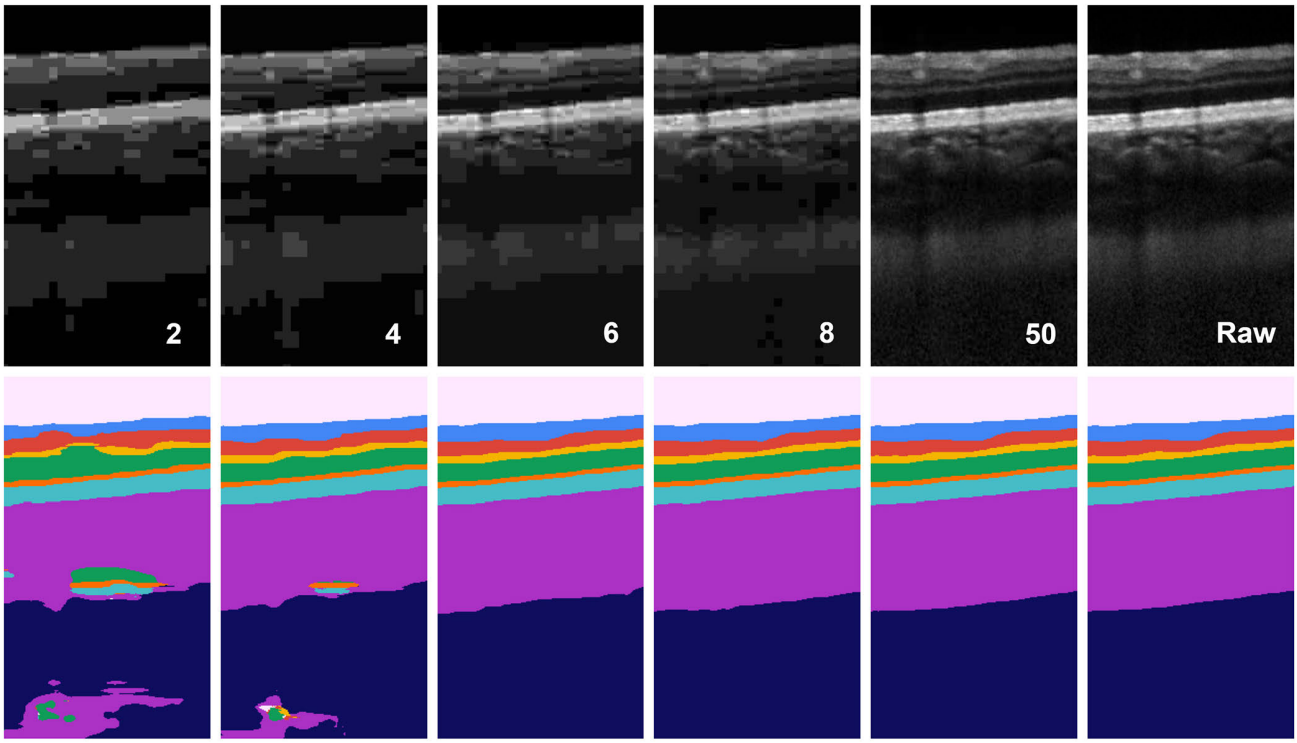


FIGURE 9. Example of performance degradation with the original networks (no fine-tuning) as a result of JPEG compression (image quality values of 2, 4, 6, 8, 50, Raw) applied to an OCT image. Top: augmented OCT images. Bottom: corresponding predictions.

small intensity differences. For gamma > 1, three boundaries (ELM, ISe, RPE) do not degrade as severely as the others (Fig. 6). This result can be understood through observation of the example in Fig. 5. With a gamma correction of 3.5, the ISe and RPE retain a very strong gradient across them while, despite fading into the background, the networks appear to have learnt to identify the IS and its top boundary (the ELM) relative to the RPE and ISe. For the other boundaries the networks struggle as they have clearly not learnt to identify the layers with such small intensity differences between them. In fact, with a gamma correction of 3.5, the intensity of the upper retinal layers decreases to the point where parts become indistinguishable from the vitreous and are therefore classified as such. For contrast reduction, all boundaries appear to degrade significantly as illustrated in Fig. 8. Like gamma correction, a possible explanation is that the networks have become hyper-tuned to identify the layers at particular levels of intensity difference and have not needed to learn the layer transitions in these conditions (Fig. 7). However, the variability between networks was noticeably greater than that observed for gamma correction suggesting a greater sensitivity to the particular combination of model weights.

The results for JPEG compression (Fig. 10) are somewhat different to the three aforementioned factors (noise, gamma, contrast) with minimal decrease in performance across the majority of the range of tested values. Here, a small increase in error was only observed for very low image quality values (< 10). Additionally, there was little spread in the performance of the networks for most image quality values. Likewise, JPEG2000 compression also demonstrated a high level of consistency between the four networks (Fig. 12) albeit with somewhat different behavior than that for JPEG. Within the tested ranges, the degradation of JPEG2000 occurred earlier and more gradually than that of standard JPEG compression where degradation instead occurred in a more pronounced fashion only near the lower extremity. These differences are apparent despite the two algorithms sharing a similar name and purpose. However, the differing behavior makes sense when analyzing the example degraded image sections for JPEG (Fig. 9) and JPEG2000 (Fig. 11). Here, clear differences are apparent with the JPEG compressed images appearing ''blocky'' or ''discretized'' while JPEG2000 causes a ''blurring'' or ''smoothing'' effect. Overall though, important features such as the boundary transitions and the relative intensity of the adjacent layers appear to be well preserved by both of these image compression algorithms which potentially explains their greater robustness compared to the other three factors.

C. EFFECT OF FINE-TUNING

It is evident that OCT image quality is an important consideration when designing intra-chorio-retinal layer segmentation methods using neural networks. One possible approach to create a network that is robust and resilient to these factors, is to train the network using images with degraded quality incorporated into the training dataset. Such data may be

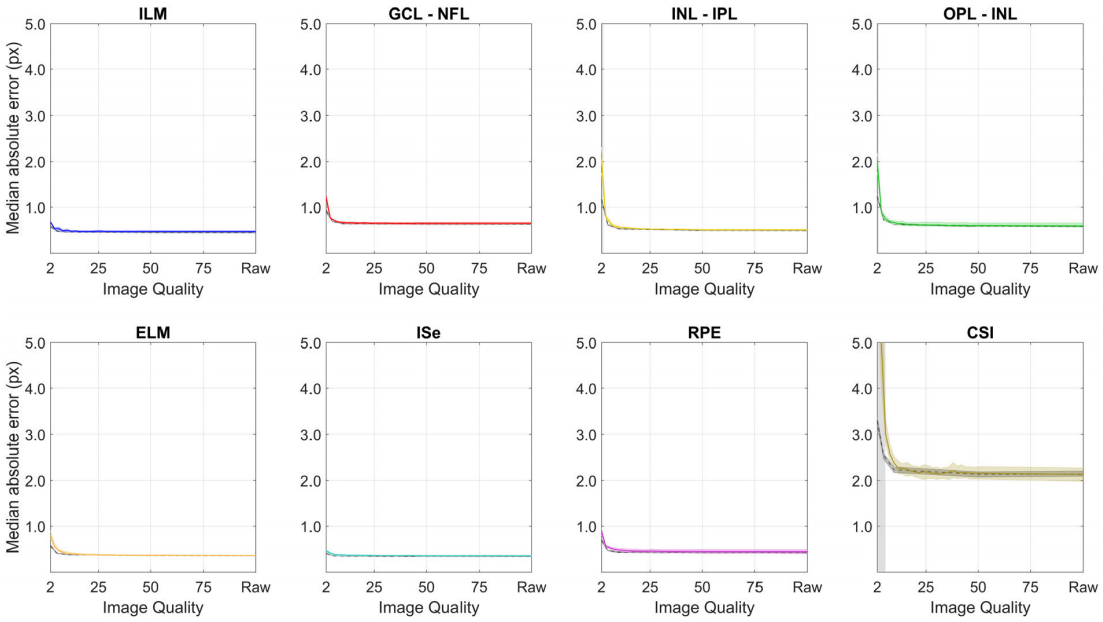


FIGURE 10. Median mean absolute error (pixels) for each boundary of interest as a result of JPEG compression applied to the test images. The solid colored line in each plot represents the mean of the four original networks (no fine-tuning) while the surrounding shaded region depicts the range of median errors exhibited by these networks. Likewise, the dashed grey line and corresponding shaded region represents the mean and range respectively of the networks fine-tuned using data augmented with JPEG compression. Here, the raw images (baseline performance) are marked at the rightmost end of the scale as 'Raw'. For the JPEG compression algorithm, it should be noted that an image quality of 100 does not correspond to an uncompressed raw image. The grey shaded rectangular region indicates where the mean error (for the original networks) becomes practically significant (+1 pixel error compared to baseline).

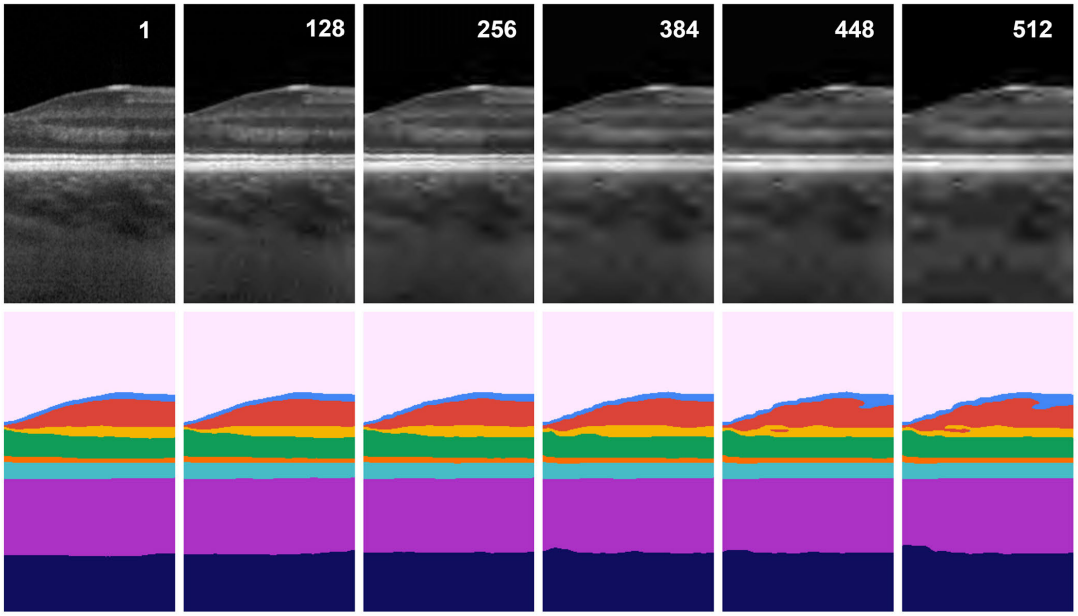


FIGURE 11. Example of performance degradation with the original networks (no-fine-tuning) as a result of JPEG2000 compression applied to an OCT image (compression ratios of 1, 128, 256, 384, 448, 512). Top: augmented OCT images. Bottom: corresponding predictions.

additional real data acquired using different acquisition parameters, instruments or techniques. Alternatively, this may be artificial data constructed using a range of representative augmentations as was the case in this study. By fine-tuning the existing networks, we have demonstrated the feasibility of such a method since significant improvements to the segmentation performance in the presence of each of the five image quality factors was observed. In most cases, the performance across the entire parameter range closely matched the baseline performance. Importantly, the baseline

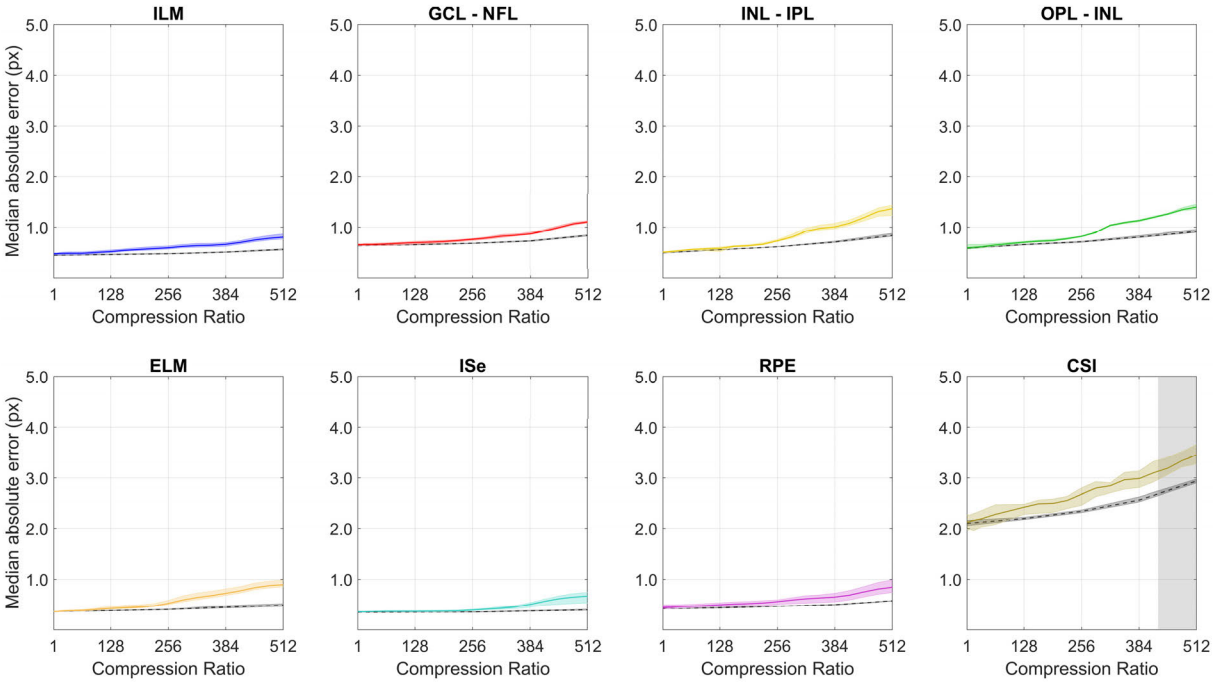


FIGURE 12. Median mean absolute error (pixels) for each boundary of interest as a result of JPEG2000 compression applied to the test images. The solid colored line in each plot represents the mean of the four original networks (no fine-tuning) while the surrounding shaded region depicts the range of median errors exhibited by these networks. Likewise, the dashed grey line and corresponding shaded region represents the mean and range respectively of the networks fine-tuned using data augmented with JPEG2000 compression. Here, the raw images (baseline performance) correspond to a compression ratio of 1. The grey shaded rectangular region indicates where the mean error (for the original networks) becomes practically significant (+1 pixel error compared to baseline).

performance did not deteriorate as a result of fine-tuning in any of our experiments. It can be noted that there were a few cases where the performance, despite improving, still failed to reach the baseline level. This may be as a result of unrecoverable information loss (e.g. from image compression) or simply that the problem was significantly harder to learn in some of these extreme cases.

The differing behavior across networks, boundaries and image quality factors is notable. Fig. 13 demonstrates that the change in weights between the original and fine-tuned networks varies between image quality factors suggesting that the process for learning each augmentation differs. However, for all image quality factors, the more prominent changes in the encoder appear to occur in the earlier layers and less so towards the bottleneck. This suggests that changes occurring in the fine-tuning process are with respect to low level features. This is logical as each of the degradations applied to the images result in visual changes at a low level and do not modify the overall structure, positioning, scale or shape of the chorio-retinal layers. A previous study noted similar results for contrast as a low level visual feature when analyzing the weight differences of fine-tuned networks [60]. Unlike the encoder, the changes within the decoder did not clearly appear to correspond in a symmetric way. There was a similar trend observed across all image quality factors with the largest differences for each level of the decoder generally occurring within the second convolutional layer. This may be because these particular layers receive input from two places: the output of the previous layer as well the output of the corresponding encoder layer and therefore undergo a greater level of tuning. The exception to this rule was the very final convolutional layer at the output which exhibited much larger changes than almost all other layers in the networks. The inter-network variability for each image quality factor (Fig. 14) illustrates that a level of inconsistency between networks may also be present for each image quality factor as was observed in the results. Here, the ranges for each factor and for each convolutional layer also differ considerably from one another suggesting that these inconsistencies may also vary between factors. As highlighted previously, the results showed this was the case with notable dissimilarities in the inter-network variability across image quality factors and boundaries. Overall though, due to the highly complex nature of deep neural networks, we note the difficulty in determining specific conclusions, correlations and explanations from the network weights on a per boundary, per network, or per factor basis.

D. FUTURE WORK

This study analyzed five representative image quality factors that are very relevant to OCT imaging of the eye. Thus providing a framework for robust image quality model training. However, there are other image quality factors that may be

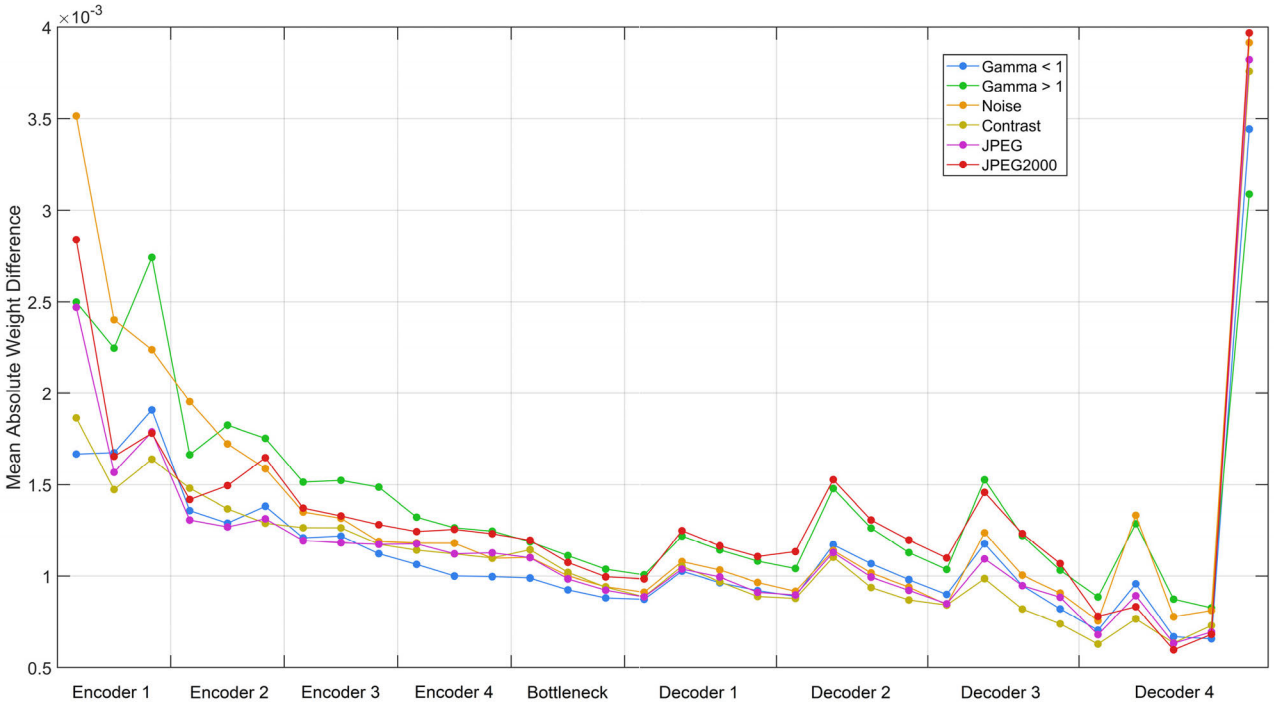


FIGURE 13. Mean absolute weight difference for each convolutional layer between the original networks and their fine-tuned versions on each image quality factor. For each factor the results for each of four runs across each of the four networks are all averaged together. Each dot corresponds to a convolutional layer with these grouped into their corresponding level in the network.

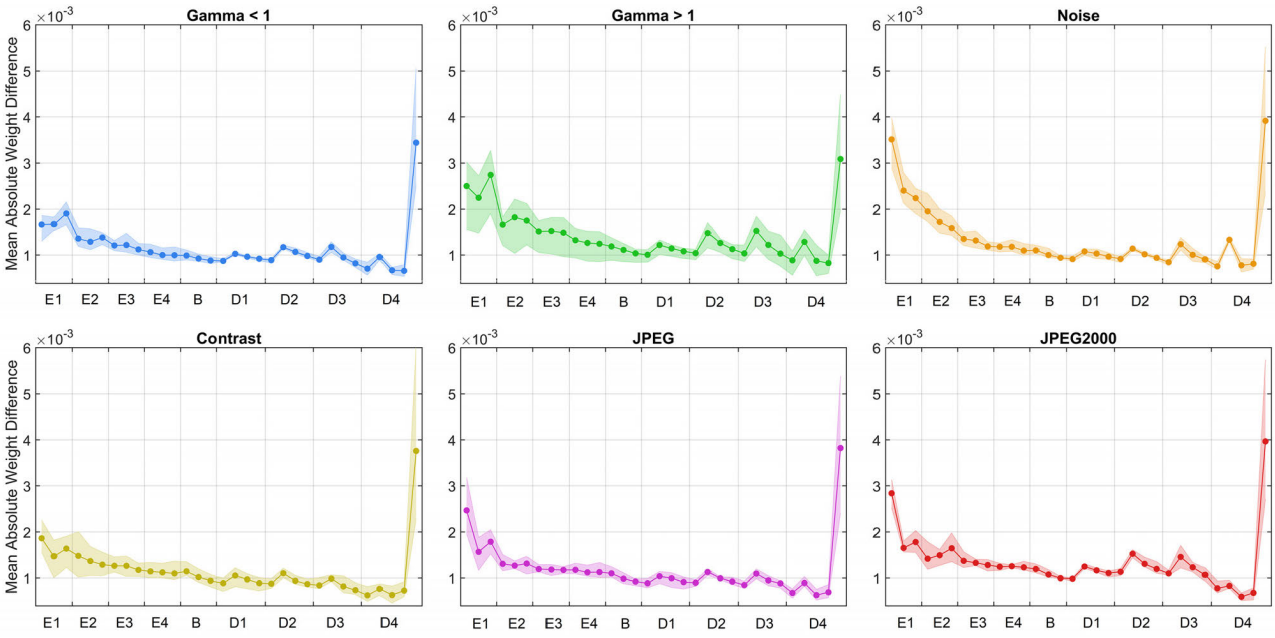


FIGURE 14. Inter-network range of mean absolute weight differences for each convolutional layer between the original networks and their fine-tuned versions on each image quality factor. Here, the shaded region indicates the range of the means of the four networks. Note that the four runs for each network are averaged together. Each dot corresponds to a convolutional layer with these grouped into their corresponding level in the network. E# corresponds to encoder layer number #, B corresponds to the bottleneck layer, and D# corresponds to decoder layer number #.

considered in future for different image modalities or imaging of different human tissues. In our current analyses, we examined each image quality factor in isolation, in order to observe the individual impact of each factor. A limitation of our work is that combinations of such image quality factors were not considered. In particular, increasing noise will inherently reduce image contrast and this is not taken into account within our study. Future work may also explore the effect

of such image quality factors on a patch-based segmentation approach and provide a comparison to the semantic segmentation approach analyzed here. Another option is to consider the possibility of exploring network models which are inherently more resilient to noise. Future analyses which compare these effects for a range of different OCT instruments would also be useful but this is beyond the scope of this particular study.

IV. CONCLUSION

This is the first study to provide an in depth analysis on the effect of image quality factors on the performance of a semantic segmentation approach to delineate chorio-retinal layer boundaries using neural networks. Data were augmented to simulate different types of image degradation commonly observed in OCT images. For noise, contrast reduction and gamma correction, the results suggest that the method is susceptible to these degradations with an increase in boundary errors observed. On the other hand, the method is somewhat more robust in the presence of both JPEG and JPEG2000 image compression. In several cases, noticeable variability was observed between identically trained networks (with different initial weights) indicating that the behavior is sensitive to the particular combination of weight parameters. A method to improve network robustness was also developed and demonstrated, with significant performance improvements observed when fine-tuning the original networks with augmented data. The results and analysis presented here provide insights into and direction for creating and testing more robust machine learning methods for OCT image analysis.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge support from the NVIDIA Corporation for the donation of GPUs used in this work. Computational resources and services used in this work were provided in part by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

REFERENCES

[1] R. Chakraborty, S. A. Read, and M. J. Collins, "Diurnal variations in axial length, choroidal thickness, intraocular pressure, and ocular biometrics," *Investigative Opthalmol. Vis. Sci.*, vol. 52, no. 8, pp. 5121–5129, Jul. 2011.

[2] S. A. Read, M. J. Collins, and D. Alonso-Caneiro, "Diurnal variation of retinal thickness with spectral domain OCT," *Optometry Vis. Sci.*, vol. 89, no. 5, pp. 611–619, May 2012.

[3] S. A. Read, M. J. Collins, S. J. Vincent, and D. Alonso-Caneiro, "Choroidal thickness in myopic and nonmyopic children assessed with enhanced depth imaging optical coherence tomography," *Investigative Opthalmol. Vis. Sci.*, vol. 54, no. 12, pp. 7578–7586, Nov. 2013.

[4] S. Grover, R. K. Murthy, V. S. Brar, and K. V. Chalam, "Normative data for macular thickness by high-definition spectral-domain optical coherence tomography (spectralis)," *Amer. J. Ophthalmol.*, vol. 148, no. 2, pp. 266–271, Aug. 2009.

[5] R. Margolis and R. F. Spaide, "A pilot study of enhanced depth imaging optical coherence tomography of the choroid in normal eyes," *Amer. J. Ophthalmol.*, vol. 147, no. 5, pp. 811–815, May 2009.

[6] S. A. Read, D. Alonso-Caneiro, S. J. Vincent, and M. J. Collins, "Longitudinal changes in choroidal thickness and eye growth in childhood," *Investigative Opthalmol. Vis. Sci.*, vol. 56, no. 5, pp. 3103–3112, May 2015.

[7] E. Harb, L. Hyman, M. Fazzari, J. Gwiazda, and W. Marsh-Tootle, "Factors associated with macular thickness in the COMET myopic cohort," *Optometry Vis. Sci.*, vol. 89, no. 5, pp. 620–631, May 2012.

[8] A. Sakamoto, M. Hangai, M. Nukada, H. Nakanishi, S. Mori, Y. Kotera, R. Inoue, and N. Yoshimura, "Three-dimensional imaging of the macular retinal nerve fiber layer in glaucoma with spectral-domain optical coherence tomography," *Investigative Opthalmol. Vis. Sci.*, vol. 51, no. 10, pp. 5062–5070, Oct. 2010.

[9] A. Wood, "Retinal and choroidal thickness in early age-related macular degeneration," *Amer. J. Ophthalmol.*, vol. 152, no. 6, pp. 1030.e2–1038.e2, 2011.

[10] I. I. Bussel, G. Wollstein, and J. S. Schuman, "OCT for glaucoma diagnosis, screening and detection of glaucoma progression," *Brit. J. Ophthalmol.*, vol. 98, no. Suppl 2, pp. ii15–ii19, Dec. 2013.

[11] F. J. L. Medina, C. I. Callén, G. Rebolleda, F. J. Muñoz-Negrete, M. J. I. Callén, and F. G. del Valle, "Use of nonmydriatic spectral-domain optical coherence tomography for diagnosing diabetic macular edema," *Amer. J. Ophthalmol.*, vol. 153, no. 3, pp. 536.e1–543.e1, Mar. 2012.

[12] A. Yazdanpanah, G. Hamarneh, B. R. Smith, and M. V. Sarunic, "Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 484–496, Feb. 2011.

[13] K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images—A graph-theoretic approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 119–134, Jan. 2006.

[14] Z. Sun, H. Chen, F. Shi, L. Wang, W. Zhu, D. Xiang, C. Yan, L. Li, and X. Chen, "An automated framework for 3D serous pigment epithelium detachment segmentation in SD-OCT images," *Sci. Rep.*, vol. 6, no. 1, p. 21739, Feb. 2016.

[15] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Express*, vol. 18, no. 18, pp. 19413–19428, Aug. 2010.

[16] M. A. Mayer, J. Hornegger, C. Y. Mardin, and R. P. Tornow, "Retinal nerve fiber layer segmentation on FD-OCT scans of normal subjects and glaucoma patients," *Biomed. Opt. Express*, vol. 1, no. 5, pp. 1358–1383, Nov. 2010.

[17] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map," *Med. Image Anal.*, vol. 17, no. 8, pp. 907–928, Dec. 2013.

[18] S. Niu, L. de Sisternes, Q. Chen, T. Leng, and D. L. Rubin, "Automated geographic atrophy segmentation for SD-OCT images using region-based C-V model via local similarity factor," *Biomed. Opt. Express*, vol. 7, no. 2, pp. 581–600, Jan. 2016.

[19] D. Koozekanani, K. Boyer, and C. Roberts, "Retinal thickness measurements from optical coherence tomography using a Markov boundary model," *IEEE Trans. Med. Imag.*, vol. 20, no. 9, pp. 900–916, Sep. 2001.

[20] S. J. Chiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Express*, vol. 6, no. 4, pp. 1172–1194, 2015.

[21] X. Liu, T. Fu, Z. Pan, D. Liu, W. Hu, J. Liu, and K. Zhang, "Automated layer segmentation of retinal optical coherence tomography images using a deep feature enhanced structured random forests classifier," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1404–1416, Jul. 2018.

[22] S. Karri, D. Chakraborthi, and J. Chatterjee, "Learning layer-specific edges for segmenting retinal layers with large deformations," *Biomed. Opt. Express*, vol. 7, no. 7, pp. 2888–2901, 2016.

[23] P. P. Srinivasan, S. J. Heflin, J. A. Izatt, V. Y. Arshavsky, and S. Farsiu, "Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology," *Biomed. Opt. Express*, vol. 5, no. 2, pp. 348–365, 2014.

[24] J. Hamwood, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers," *Biomed. Opt. Express*, vol. 9, no. 7, pp. 3049–3066, 2018.

[25] J. Kugelman, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search," *Biomed. Opt. Express*, vol. 9, no. 11, pp. 5759–5777, 2018.

[26] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express*, vol. 8, no. 5, pp. 2732–2744, Apr. 2017.
[27] D. Alonso-Caneiro, J. Kugelman, J. Hamwood, S. A. Read, S. J. Vincent, F. K. Chen, and M. J. Collins, "Automatic retinal and choroidal boundary segmentation in OCT images using patch-based supervised machine learning methods," in *Proc. Asian Conf. Comput. Vis. (ACCV)*. Cham, Switzerland: Springer, 2018, pp. 215–228.
[28] J. Kugelman, D. Alonso-Caneiro, S. A. Read, J. Hamwood, S. J. Vincent, F. K. Chen, and M. J. Collins, "Automatic choroidal segmentation in OCT images using supervised deep learning methods," *Sci. Rep.*, vol. 9, no. 1, Sep. 2019, Art. no. 13298.
[29] G. N. Girish, B. Thakur, S. R. Chowdhury, A. R. Kothari, and J. Rajan, "Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 296–304, Jan. 2019.
[30] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, Jul. 2017.
[31] F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J. J. P. van Grinsven, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sánchez, "Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks," *Biomed. Opt. Express*, vol. 8, no. 7, pp. 3292–3316, Jun. 2017.
[32] M. Chen, J. Wang, I. Oguz, B. L. VanderBeek, and J. C. Gee, "Automated segmentation of the choroid in EDI-OCT images with retinal pathology using convolution neural networks," in *Fetal, Infant and Ophthalmic Medical Image Analysis*. Cham, Switzerland: Springer, 2017, pp. 177–184.
[33] Y. Xu, K. Yan, J. Kim, X. Wang, C. Li, L. Su, S. Yu, X. Xu, and D. D. Feng, "Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy," *Biomed. Opt. Express*, vol. 8, no. 9, pp. 4061–4076, Aug. 2017.
[34] D. Xiang, H. Tian, X. Yang, F. Shi, W. Zhu, H. Chen, and X. Chen, "Automatic segmentation of retinal layer in OCT images with choroidal neovascularization," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5880–5891, Dec. 2018.
[35] X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046–3061, 2019.
[36] J. Loo, L. Fang, D. Cunefare, G. J. Jaffe, and S. Farsiu, "Deep longitudinal transfer learning-based automatic segmentation of photoreceptor ellipsoid zone defects on optical coherence tomography images of macular telangiectasia type 2," *Biomed. Opt. Express*, vol. 9, no. 6, pp. 2681–2698, May 2018.
[37] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.
[38] A. Shah, L. Zhou, M. D. Abrámoff, and X. Wu, "Multiple surface segmentation using convolution neural nets: Application to retinal layer segmentation in OCT images," *Biomed. Opt. Express*, vol. 9, no. 9, pp. 4509–4526, Aug. 2018.
[39] H. Jeelani, J. Martin, F. Vasquez, M. Salerno, and D. S. Weller, "Image quality affects deep learning reconstruction of MRI," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 357–360.
[40] T. Maruyama, N. Hayashi, Y. Sato, S. Hyuga, Y. Wakayama, H. Watanabe, A. Ogura, and T. Ogura, "Comparison of medical image classification accuracy among three machine learning methods," *J. X-Ray Sci. Technol.*, vol. 26, no. 6, pp. 885–893, Dec. 2018.
[41] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Proc. 8th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6.
[42] P. Roy, S. Ghosh, S. Bhattacharya, and U. Pal, "Effects of degradations on deep neural network architectures," 2018, *arXiv:1807.10108*. [Online]. Available: http://arxiv.org/abs/1807.10108
[43] G. B. Paranhos da Costa, W. A. Contato, T. S. Nazare, J. E. S. Batista Neto, and M. Ponti, "An empirical study on the effects of different types of noise in image classification tasks," 2016, *arXiv:1609.02781*. [Online]. Available: http://arxiv.org/abs/1609.02781
[44] J.-Y. Franceschi, A. Fawzi, and O. Fawzi, "Robustness of classifiers to uniform ℓ_p and Gaussian noise," 2018, *arXiv:1802.07971*. [Online]. Available: https://arxiv.org/abs/1802.07971
[45] G. M. Somfai, H. M. Salinas, C. A. Puliafito, and D. C. Fernandez, "Evaluation of potential image acquisition pitfalls during optical coherence tomography and their influence on retinal image segmentation," *J. Biomed. Opt.*, vol. 12, no. 4, 2007, Art. no. 041209.
[46] B. E. Varga, E. Tátrai, D. C. DeBuc, and G. M. Somfai, "The effect of incorrect scanning distance on boundary detection errors and macular thickness measurements by spectral domain optical coherence tomography: A cross sectional study," *BMC Ophthalmol.*, vol. 14, no. 1, p. 148, Nov. 2014.
[47] D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Speckle reduction in optical coherence tomography imaging by affine-motion image registration," *J. Biomed. Opt.*, vol. 16, no. 11, 2011, Art. no. 116027.
[48] M. Balasubramanian, C. Bowd, G. Vizzeri, R. N. Weinreb, and L. M. Zangwill, "Effect of image quality on tissue thickness measurements obtained with spectral domain-optical coherence tomography," *Opt. Express*, vol. 17, no. 5, p. 4019, Mar. 2009.
[49] S. A. Read, D. Alonso-Caneiro, and S. J. Vincent, "Longitudinal changes in macular retinal layer thickness in pediatric populations: Myopic vs non-myopic eyes," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0180462.
[50] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[53] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[55] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras
[56] *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://tensorflow.org
[57] M. E. J. van Velthoven, "Influence of cataract on optical coherence tomography image quality and retinal thickness," *Brit. J. Ophthalmol.*, vol. 90, no. 10, pp. 1259–1262, Jul. 2006.
[58] J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *J. Biomed. Opt.*, vol. 4, no. 1, pp. 95–105, 1999.
[59] H. M. Salinas and D. C. Fernandez, "Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography," *IEEE Trans. Med. Imag.*, vol. 26, no. 6, pp. 761–771, Jun. 2007.
[60] A. Akbarinia and K. R. Gegenfurtner, "Manifestation of image contrast in deep networks," 2019, *arXiv:1902.04378*. [Online]. Available: http://arxiv.org/abs/1902.04378

JASON KUGELMAN was born in Brisbane, Queensland, Australia, in 1994. He received the B.Eng. degree (Hons.) in software engineering and the B.Sc. degree in mathematics from The University of Queensland, in 2018.

In 2018, he was a Research Assistant with the Contact Lens and Visual Optics Laboratory and since 2019, has been a Research Engineer. His research interests include image processing algorithms, machine learning, neural networks, mathematical optimization, graph algorithms, and generative modeling. In 2018, he has coauthored his first research article titled *Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search*.

DAVID ALONSO-CANEIRO received the B.Eng. degree in telecommunications and the M.Eng. degree in electronics from the University of Valencia, Valencia, Spain, in 2002 and 2004, respectively, and the Ph.D. degree from the Queensland University of Technology (QUT), in 2010.

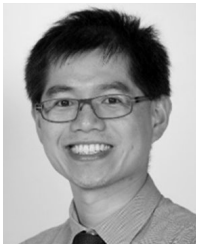
From 2011 to 2016, he was a Postdoctoral Fellow with the School of Optometry and Vision Science, QUT, where he has been a Senior Research Fellow, since 2016. He has published more than 50 peer-reviewed research articles. His research interests include ocular imaging, image analysis, and image processing, with specially interest in the application of machine learning method to medical image analysis.

SCOTT A. READ was born in Brisbane, Queensland, Australia, in 1976. He received the B.App.Sci. degree (Hons.) in optometry and the Ph.D. degree in vision science from the Queensland University of Technology (QUT), Brisbane, in 1997 and 2006, respectively.

From 2007 to 2012, he was a Postdoctoral Fellow with the School of Optometry and Vision Science, QUT. From 2012 to 2015, he was an Australian Research Council, Discovery Early Career Research Fellow (DECRA). Since 2015 he has been Associate Professor with the School of Optometry and Vision Science, QUT. He has authored two book chapters and more than 80 peer-reviewed research articles. His research interests include ocular imaging and understanding the ocular and environmental factors contributing towards refractive error development and progression. He is currently a member of the Association for Research in Vision and Ophthalmology (ARVO) and a Fellow of the American Academy of Optometry. He was awarded the Zeiss Young Investigator Award in Myopia Research, in 2015. He is also an Associate Editor and an Editorial Board Member of the *Journal of Optometry and Vision Science*.

STEPHEN J. VINCENT received the B.App.Sc. (Optom) (Hons.) and Ph.D. degrees from the Queensland University of Technology (QUT), Brisbane, QLD, Australia, in 2004 and 2011, respectively.

He is currently Associate Professor with the School of Optometry and Vision Science, QUT. His current research interests include visual optics, refractive error development, and ocular imaging. He is a Fellow of the American Academy of Optometry, the British Contact Lens Association, the Australia College of Optometry, the Scleral Lens Education Society, and the Higher Education Academy. He is an Associate Editor of the *Journal of Contact Lens and Anterior Eye*.



FRED K. CHEN the received M.B.B.S. degree (Hons.) from The University of Western Australia, in 1999, and the Ph.D. degree in ophthalmology from the University College of London, in 2011.

Since 2011, he has been leading the Ocular Tissue Engineering Laboratory, Lions Eye Institute, Perth, WA, Australia. He is currently an Academic Clinician trained in ophthalmology in Perth, and London, U.K. He is the author of five book chapters and more than 100 peer-reviewed articles. His research interests include epidemiology of blindness, disease phenotyping with multimodal retinal imaging, genotype-phenotype correlation in inherited retinal disease, induced pluripotent stem cell and retinal organoid generation for disease modeling, development of antisense oligonucleotide for treatment, and clinical trials of new treatment in retinal and macular degenerations. He is an Ex-Board Member of the Ophthalmic Research Institute of Australia. He was a recipient of the several international and national competitive research grants and the NHMRC Early Career Fellowship, from 2013 to 2016, and the Career Development Award, from 2018 to 2021. He is a Section Editor of the journal, *Clinical and Experimental Ophthalmology*.

MICHAEL J. COLLINS received the Diploma degree in applied science (optometry) from QUT, in 1977, the M.App.Sc. degree in visual psychophysics, in 1986, and the Ph.D. degree in visual optics, in 1996.

Then, he worked as a Clinical Optometrist. He was appointed to the teaching staff at QUT, in 1980. He has published more than 190 research articles. His research laboratory, the Contact Lens and Visual Optics Laboratory, commenced, in 1990, and specializes in myopia and the visual and optical characteristics of the eye and contact lenses. He is currently a Fellow of the American Academy of Optometry, the Cornea and Contact Lens Society of Australia, and the Association for Research in Vision and Ophthalmology.