RAC-CNN: multimodal deep learning based automatic detection and classification of rod and cone photoreceptors in adaptive optics scanning light ophthalmoscope images

DAVID CUNEFARE,[1,*] ALISON L. HUCKENPAHLER,[2] EMILY J. PATTERSON,[3] ALFREDO DUBRA,[4] JOSEPH CARROLL,[2,3] AND SINA FARSIU[1,5] (ID)

[1]*Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA*
[2]*Department of Cell Biology, Neurobiology, & Anatomy, Medical College of Wisconsin, Milwaukee, WI 53226, USA*
[3]*Department of Ophthalmology & Visual Sciences, Medical College of Wisconsin, Milwaukee, WI 53226, USA*
[4]*Department of Ophthalmology, Stanford University, Palo Alto, CA 94303, USA*
[5]*Department of Ophthalmology, Duke University Medical Center, Durham, NC 27710, USA*
[*]david.cunefare@duke.edu

Abstract: Quantification of the human rod and cone photoreceptor mosaic in adaptive optics scanning light ophthalmoscope (AOSLO) images is useful for the study of various retinal pathologies. Subjective and time-consuming manual grading has remained the gold standard for evaluating these images, with no well validated automatic methods for detecting individual rods having been developed. We present a novel deep learning based automatic method, called the rod and cone CNN (RAC-CNN), for detecting and classifying rods and cones in multimodal AOSLO images. We test our method on images from healthy subjects as well as subjects with achromatopsia over a range of retinal eccentricities. We show that our method is on par with human grading for detecting rods and cones.

1.   Introduction

Analysis of rod and cone photoreceptors is valuable for the study, diagnosis, and prognosis of various retinal diseases. The high resolution of adaptive optics (AO) ophthalmoscopes enables the visualization of photoreceptors in the living human retina [1,2], and these AO ophthalmoscopes have been used to study the properties of cones [3–8] and rods [6,8–10] in both healthy and pathological eyes. AO ophthalmic imaging was first demonstrated with a flood illumination camera [11]. Currently however, the most widely used AO ophthalmic technology is the AO scanning light ophthalmoscope (AOSLO), due to the superior image contrast provided by its axial sectioning, and its potential higher transverse resolution [1,12–16]. In addition to reflectance imaging, the point scanning in AOSLO enables multiple-scattering imaging, which reveals transparent retinal neuronal and vascular structures [17–19], as well as single- and two-photon fluorescence [20–23], and fluorescence lifetime imaging [24]. Finally, AO optical coherence tomography allows for acquisition of 3-dimential images with comparable transverse resolution and an order of magnitude superior axial resolution, only limited by the light source spectrum [25–30]. Confocal AOSLO is able to visualize rods and foveal cones, the smallest photoreceptors in the retina [9]. In the past decade, the benefits presented by non-confocal AOSLO modalities have been explored, with split detector AOSLO [17] providing a number of advantages. Split detector AOSLO has a reduced ability to visualize rods in comparison to confocal AOSLO, but

due to the different source of contrast, split detector AOSLO is often able to reduce ambiguity in identifying cones, especially in diseased eyes [31].

To utilize the captured images for clinical or research purposes, quantitative metrics such as rod and cone density, spacing, size, and luminance are often measured [32]. Generally, before these quantitative metrics can be calculated, each individual rod or cone in a region of interest (ROI) must be localized. The current gold standard method of manually marking these photoreceptors is highly subjective [33] and time consuming, which acts as a bottleneck limiting the clinical utilization of AOSLO systems. To combat this problem, several automated algorithms have been developed to detect cones in ophthalmic AO images taken from healthy [34–50] and pathological [31,48,50] subjects. To date, no well validated automatic method for detecting individual rods in AOSLO images or method for classifying between cones and rods has been published.

In recent years, deep learning has achieved state-of-the-art results for a variety of image processing tasks. Of particular note are convolutional neural networks (CNNs), which apply a sequence of transforming layers to an image with weights learned directly from training data [51]. CNNs have been utilized for a variety of tasks in ophthalmic image processing including classification [52–56], segmentation [57–62], and image enhancement [63]. CNNs have been used to achieve state-of-the-art performances for cone localization in ophthalmic AO images in healthy [46,49] and pathologic [31,50] eyes. Our recent work [31] showed that a CNN using multimodal confocal and split detector AOSLO information could improve the performance of detecting cones in subjects with achromatopsia (ACHM) by utilizing the complimentary information captured in both modalities.

In this work, we present the first validated method for automatically detecting and classifying rod and cone photoreceptors in AOSLO images. We develop a novel CNN semantic segmentation architecture that combines information from both the confocal and split detector AOSLO modalities. We show that our method is able to accurately localize and classify rods and cones in AOSLO images from both healthy subjects and from those with ACHM. We validate the results of our method against the current gold standard of manual marking.

2. Review

2.1. AOSLO photoreceptor imaging

AOSLO imaging is able to visualize rod and cone photoreceptors within the human retina [9]. Although the size and density of cones and rods vary as a function of retinal eccentricity [64], rods are generally smaller and more numerous than cones in healthy subjects. Rods are often more difficult to visualize, which may be due to their small size [9]. Figures 1(a) and 1(b) provide an example of co-registered confocal and split detector AOSLO images taken from a healthy subject. It can be seen that the confocal image is able to visualize both the cones and relatively small rods. The split detector image is not able to visualize rods, but it has high contrast in visualizing cones and can be helpful for reducing ambiguities in identifying cone photoreceptors [31,45].

Figures 1(c) and 1(d) show an example of simultaneously captured confocal and split detector AOSLO images from a subject with ACHM. ACHM is a genetic retinal disorder characterized by a lack of cone function resulting in color blindness, photophobia, nystagmus, and severely reduced visual acuity. In confocal AOSLO images of subjects with ACHM, cones appear as dark spots instead of the bright spots seen when imaging healthy subjects [17,65]. Because rods can still be visualized, it is sometimes possible to infer the positions of cones by finding areas surrounded by rods. Split detector AOSLO imaging is able to visualize remnant cone structures in ACHM, and visualized cones have similar features as compared to cones imaged in healthy subjects. Even though split detector AOSLO can visualize most cones in ACHM, there is still ambiguity in determining some cone locations due to poor contrast seen in pathologic subjects [33], and using multimodal information has been shown to be beneficial in identifying cones
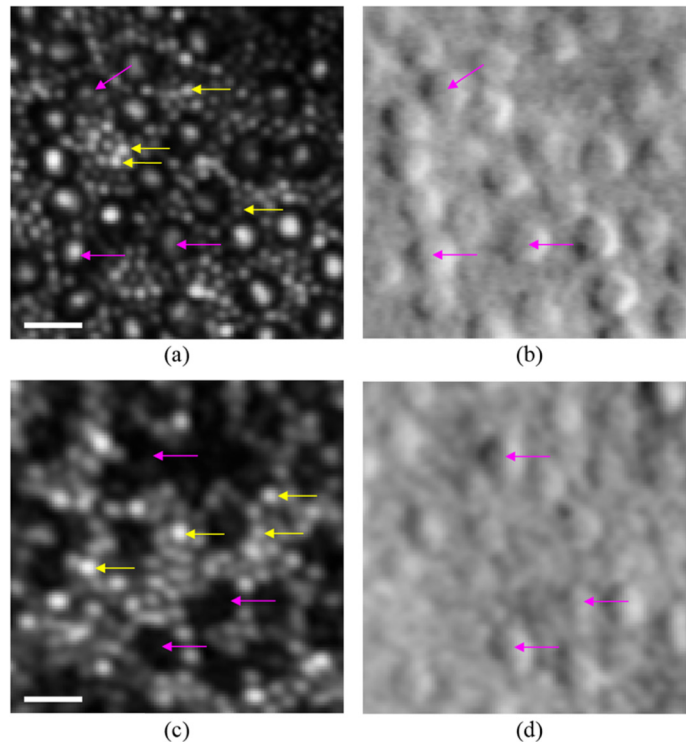
Fig. 1. Rod and cone photoreceptor visualization on AOSLO. **(a)** Confocal AOSLO image at 7° from the fovea in a normal subject. **(b)** Co-registered non-confocal split detector AOSLO image from the same location as (a). **(c)** Confocal AOSLO image at 3° from the fovea in a subject with ACHM. **(d)** Simultaneously captured split detector AOSLO image from the same location as (c). Cone photoreceptor examples are shown with magenta arrows, and rod photoreceptor examples are shown with yellow arrows. *Scale bars: 10 μm.*

[31]. Additionally, it has been seen that rods may sometimes be visualized with split detector AOSLO imaging in ACHM subjects [17].

2.2. Convolutional neural networks

CNNs apply a series of transforming layers to complete a specific task. These networks learn filters to extract relevant features directly from training data. Of note for medical imaging processing is U-net [66], a semantic segmentation network that takes a full-sized image and outputs a classification for each pixel in the original image. U-net and its variants have been used for a variety of semantic segmentation [61,66–68] and detection [68] tasks across multiple imaging modalities.

The performance and function of a CNN depends on the layers that compose it. Specific layer types used in this work include convolutional, batch normalization, rectified linear units (ReLU), max pooling, concatenation (i.e. fusion), unpooling, and soft-max layers. Convolutional layers convolve an input of size $W \times H \times D$ (before padding) with N kernels of size $F \times G \times D$ with a stride of 1 which produces an output of size $W \times H \times N$, where the output can be considered a stack of N feature maps. For each of these N feature maps, a potentially different bias value is added. The weight values for the kernel and the bias values are learned automatically through training. Batch normalization layers [69] normalize their inputs based on mean and variance statistics, which can reduce internal covariate shift to decrease overfitting during training. ReLU

layers [70] transform their inputs by setting all negative values to 0, which speeds up the training process and improves the performance of the network by adding a source of non-linearity [71]. Max pooling layers apply a max operation over the first two dimensions of their input in a $P \times Q$ window with a stride of 2, which effectively down-samples the input by a factor of 2 in the first two dimensions. Concatenation layers combines two inputs of size $A \times B \times Y$ and $A \times B \times Z$ into a single output of size $A \times B \times (Y + Z)$. Unpooling layers are used in a decoder portion of a network to up-sample their input while preserving the spatial information from the max pooling layer in the corresponding encoder portion of the network [72]. Finally, a soft-max [73] layer takes an input of size $I \times J \times C$, where C is the number of classes, and applies the soft-max function across the third dimension to get the probability of each point in the first two dimension belonging to each class. For semantic segmentation, $I \times J$ is the size of the original image.

3. Methods

The steps for training and testing our proposed rod and cone detection and classification algorithm are outlined in Fig. 2. In the training phase, we used manual rod and cone markings on paired confocal and split detector images to create label and weight maps. These maps and the original images were then used to train a semantic segmentation CNN to create probability maps of rod and cone locations, on which the parameters for rod and cone localization were learned. In the testing phase, we then used the trained CNN and learned detection parameters to detect and classify rods and cones in paired confocal and split detector images that had not been previously seen.
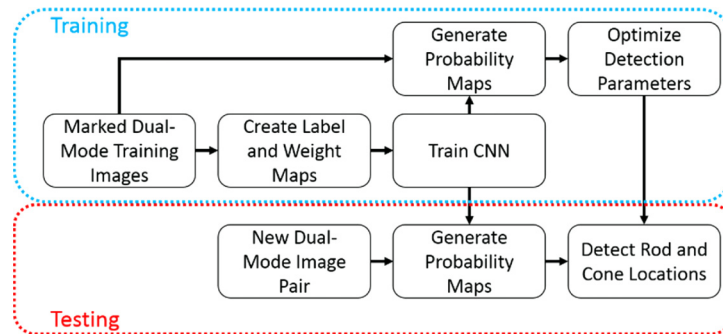

Fig. 2. Outline of the CNN AOSLO rod and cone detection algorithm.

3.1. Data sets

The images for the data sets used in this work were acquired at the Medical College of Wisconsin using a previously described AOSLO system [13,17], which simultaneously captures both confocal and split detector images. The images were obtained from the Advanced Ocular Imaging Program image bank (MCW, Milwaukee, Wisconsin). Images were acquired from healthy subjects and subjects with ACHM. For each subject, a series of image sequences was captured along the temporal meridian over a range of eccentricities using a 1.0° field of view for each sequence. For the healthy subjects, image sequences were repeated at each location over a 0.05 D focus range at 0.005 intervals. Each image sequence was strip-registered and averaged using between 19 and 70 frames, as previously described [17,74]. For healthy subjects, processed images from the multiple focus positions were manually aligned and overlapping regions were then cropped and used to create a new image sequence that was strip-registered and averaged once more, resulting in a further enhanced image. We extracted ROIs from the averaged images

to form our data sets. Lateral scale/sampling for each subject was calculated using axial length measurements from an IOL Master (Carl Zeiss Meditec Inc., Dublin, California, USA).

We used separate data sets for the healthy and ACHM cases. Our healthy data set consisted of 40 confocal and split detector image pairs from 8 subjects. For each subject, there were 5 image pairs that were approximately evenly spread between 3° and 7° from the fovea. The average image size for the healthy data set was 115×115 μm^2. Our ACHM data set consisted of 49 confocal and split detector image pairs from 7 subjects. For each subject, there were 7 image pairs that were approximately evenly spread between 1° and 7° from the fovea. The average image size for the ACHM data set was 116×116 μm^2. All image pairs from both data sets had their cones and rods marked by two manual graders independently. Both modalities were used when creating the markings, resulting in a single set of markings for each confocal and split detector image pair for each grader. The first manual grader marked a total of 7847 cones and 32111 rods across the healthy data set, and a total of 3204 cones and 32641 rods across the ACHM data set. The second manual grader marked a total of 7853 cones and 31664 rods across the healthy data set, and a total of 3071 cones and 28418 rods across the ACHM data set.

3.2. Labeling and weighting

Training our semantic segmentation CNN to classify and detect rod and cone locations requires label and weight maps for the training data. For each confocal and split detector image pair in the training data set, both of size $I{\times}J$, we generated a single label map of size $I{\times}J$. Each pixel in the label map corresponds to the same location in the confocal and split detector images and classifies those pixels as one of three possible classes: rod, cone, or background. Because our manual markings (Figs. 3(c) and 3(d)) are only photoreceptor locations and not segmentations, we developed a method similar to the one presented for cell detection in Falk *et al.* [68] to use the manual markings to create label maps that could be used for photoreceptor detection and classification. We rounded the coordinates of the cone markings from the first grader to the nearest pixel location and labeled all pixels within a 3.5 pixel radius of the coordinates as the cone class. Next, we labeled all pixels in a 2×2 pixel block around the manual rod coordinates from the first grader as the rod class. If there would be any overlap between the cone and rod labeled pixels, we labeled them as rods. We labeled all remaining pixels as background to produce the final label map (Fig. 3(e)). The sizes of the cone and rod masks were chosen empirically to be smaller than the sizes of rods and cones observed in a separate data set.

For each label map, we also generated a weight map of the same size that determined how much weight to place on correctly classifying each pixel when training the CNN. Because there was a large difference in the number of labels for each class, we adjusted the weights to balance for this difference in class representation. The weight was determined by:

$$w(x) = \begin{cases} L_{Background}/L_{Rod} & l(x) : Rod \\ L_{Background}/L_{Cone} & l(x) : Cone \\ 1 & l(x) : Background \end{cases}, \tag{1}$$

where $l(x)$ is the label for pixel x, and $w(x)$ is the associated weight for pixel x. L_{Rod}, L_{Cone}, and $L_{Background}$ are the number of pixels throughout the entire training data set labeled as rod, cone, or background, respectively. An example weight map is shown in Fig. 3(f).

3.3. Convolutional neural network

We built a novel dual-mode semantic segmentation network which we call the rod and cone CNN (RAC-CNN). The architecture of the network is shown in Fig. 4, and was based on the encoder-decoder architectures with skip connections seen in U-net [66] and ReLayNet [61],
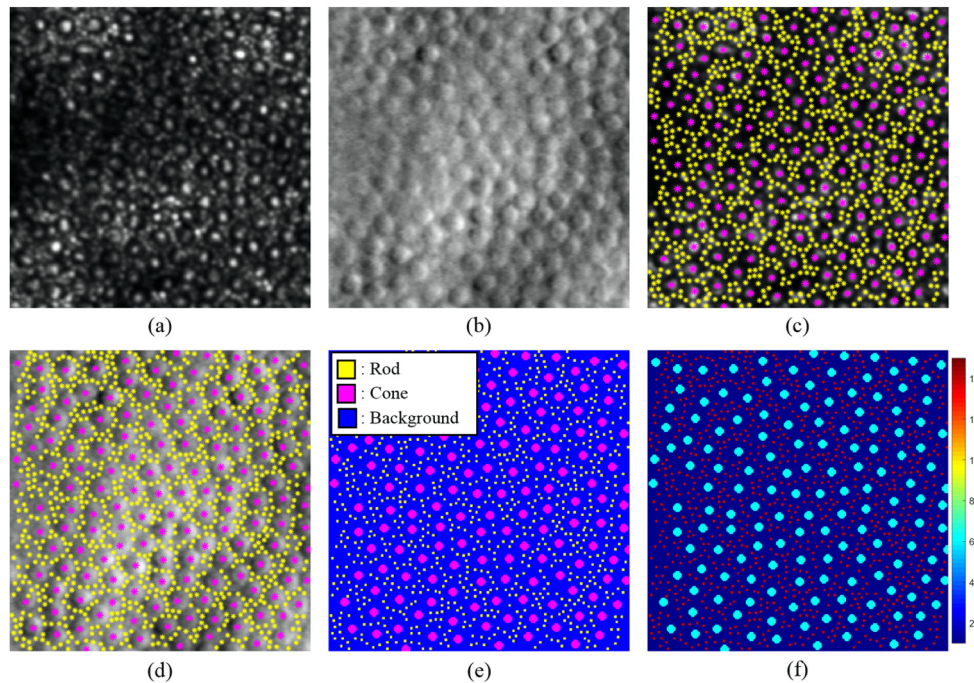
Fig. 3. Creating label and weight maps from AOSLO image pairs. **(a)** Confocal AOSLO image. **(b)** Co-registered non-confocal split detector AOSLO image from the same location. **(c-d)** Manually marked rod positions shown in yellow and cone positions shown in magenta on the confocal image shown in (a) and on the split detector image shown in (b). **(e)** Label map generated from the markings in (c-d). **(f)** Weight map corresponding to the label map in (e).

along with the dual-mode nature and filter sizes used in the DM-LF-CNN [31] used previously for AOSLO images. Our RAC-CNN is composed of two structurally identical paths, with one path taking a confocal image as the input and the other taking the matching split detector image. Each path consists of three contracting encoder blocks, followed by a transition block, and then three expanding decoder blocks. Each encoder block consists of a convolutional layer, a batch normalization layer, a ReLU layer, and a max pooling layer in that order. The transition block has a convolutional layer, a batch normalization layer, and a ReLU layer. Each decoder block consists of an unpooling layer that uses the max pooling indices from the corresponding encoder block, a concatenation layer that combines the input with the output from the ReLU layer from the corresponding encoder block (known as a skip connection [75]), a convolutional layer, a batch normalization layer, and a ReLU layer. All convolutional layers in the encoder, transition, and decoder blocks used 64 kernels of size 5×5×D (where D is the size of the third dimension of the input into the convolutional layer), and all max pooling layers had windows of size 2×2. The output of both paths is concatenated together and put through a final convolutional layer with 3 kernels of size 1×1×128 and a soft-max layer, which outputs the probability of each pixel pair in the original images belonging to each class.

The weight and bias parameters must first be learned before a network can be used for inference. We trained separate networks for the healthy and ACHM cases. First, we initialized the weights of the convolutional layers by drawing from a Gaussian distribution with a standard deviation of $\sqrt{2/K}$, where K is the number of nodes in one kernel, and initialized the biases of the convolutional layers to 0 [66]. The network parameters were then learned using stochastic
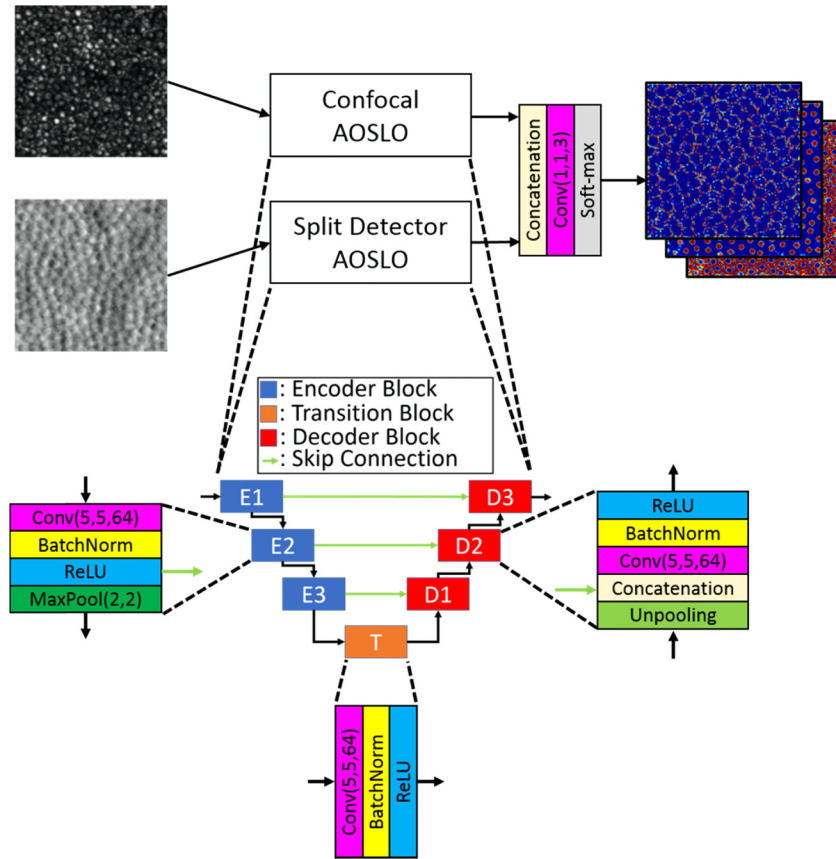
Fig. 4. The rod and cone CNN (RAC-CNN) architecture, which consists of the following layers: convolutional (Conv(*F*,*G*,*N*) where *F* and *G* are the kernel sizes in the first two dimensions and *N* is the number of kernels), batch normalization (BatchNorm), ReLU, max pooling (MaxPool(*P*,*Q*) where *P* and *Q* are the window dimensions), unpooling, concatenation, and soft-max. The same structure is used in the split detector AOSLO and confocal AOSLO paths.

gradient descent to minimize the weighted cross-entropy loss:

$$L = -\sum_{x \in \Omega} w(x) \log(p_{l(x)}(x)), \qquad (2)$$

where x is a pixel in the image domain Ω, $w(x)$ is the associated weight for the pixel, and $p_{l(x)}(x)$ is the probability output from the soft-max layer for the pixel x associated with the true class label $l(x)$. As in [66], we used a mini-batch size of 1 so that each step of the stochastic gradient descent was over a single image. Note that when a mini-batch size of 1 is used, batch normalization layers act as instance normalization layers [76]. During inference, we used the mini-batch mean and variance for the batch normalization layers as opposed to the moving average mean and variance which can be inaccurate when small batch sizes are used. We performed the training over 100 epochs, where in each epoch all the training images are seen once. The learning rate was set initially to 0.1 and was gradually lowered to 0.0001 by the final epoch, with logarithmically equally spaced values between epochs. Weight decay was set to 0.0005, and momentum was set to 0.9, which are the default values in MatConvNet [77]. We also used data augmentation in the form of vertical flipping and translations. Every time an image pair is seen during training, there

was a 50% chance for both images to be flipped vertically. To train for translational invariance, a square 200×200 pixel region was randomly chosen and used for training each time an image pair was seen. The same region was used for the paired confocal and split detector images.

3.4. Photoreceptor localization

Finally, we used the probability maps generated by the trained RAC-CNN to detect cone and rod locations for split detector and confocal image pairs. Each image pair is fed into the trained network which outputs a 3-dimensional matrix of size $I{\times}J{\times}3$, where $I{\times}J$ is the size of either of the input images. This matrix can be thought of as a stack of three probability maps, where each probability map corresponds to one of the classes: cones, rods, or background. From this matrix, we pulled out the probability maps for rods (Fig. 5(c)) and cones (Fig. 5(d)) and processed them separately using the method presented in Cunefare *et al.*[31]. In brief, we smoothed each map by convolving it with a Gaussian filter with standard deviation σ to filter out spurious maxima. Next, we applied the extended-maxima transform using MATLAB's *imextendedmax* function [78], which finds connected maximal regions where the probability difference in the region is less than or equal to a set value H and outputs these regions in a binary map (Figs. 5(e) and 5(f)). We found all connected clusters in each binary map to use as potential candidates for rod or cone positions, depending on the probability map used, and eliminated weak candidates by removing any cluster whose maximum value in the filtered probability map was less than a threshold T. Finally, we found the center of mass for all the remaining clusters which were considered to be the rod or cone positions (Figs. 5(g) and 5(h)). We set the values of σ, H, and T automatically by maximizing the average Dice's coefficient (explained in 3.5) across the same training images and manual markings used to train the RAC-CNN over a set of potential parameter combinations. The values of σ, H, and T differed and were found separately for the cone and rod cases, and for the healthy and ACHM data sets.
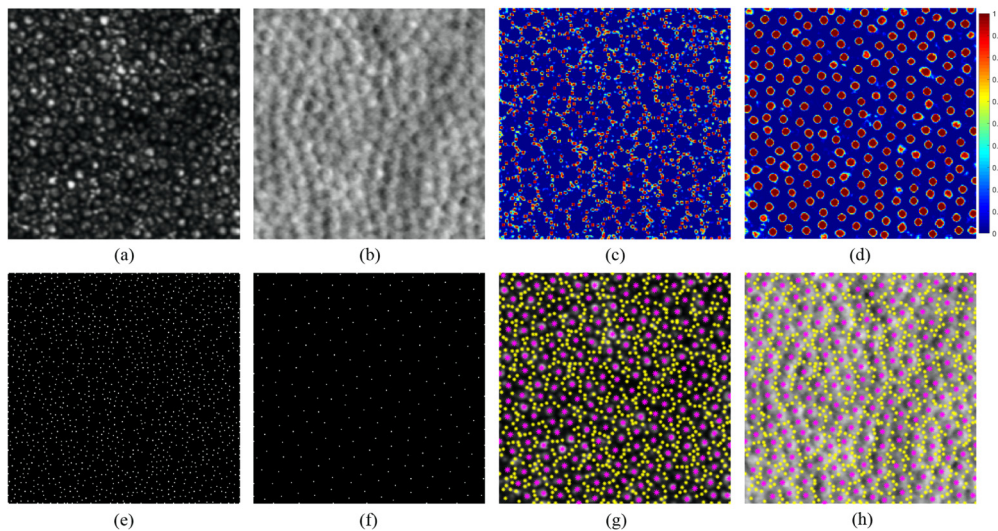

Fig. 5. Detection of rods and cones in confocal and split detector AOSLO image pairs. **(a)** Confocal AOSLO image. **(b)** Co-registered non-confocal split detector AOSLO image from the same location. **(c)** Rod probability map and **(d)** cone probability map generated from (a) and (b) using the trained RAC-CNN. **(e)** Extended maxima of (c). **(f)** Extended maxima of (d). **(g-h)** Detected rods marked in yellow and cones marked in magenta on the confocal image shown in (a) and on the split detector image shown in (b).

3.5. Validation

We validated the RAC-CNN detection method against the current gold standard of manual grading. We used leave-one-subject-out cross validation to evaluate our method, where the images from one subject were held back, and the images from the remaining subjects were used for training the network and photoreceptor localization parameters. The images from the held back subject were then evaluated using the trained network and parameters. This was repeated for all subjects in the data set, so that all subjects could be used for validation without overlap between subjects used for training and testing of the algorithm. We performed the validation separately for the healthy and ACHM data sets. The first set of manual markings was used for validation of our method. In order to compare to state-of-the-art cone detection methods, we also evaluated the performance of the LF-DM-CNN cone detection method [31]. The LF-DM-CNN was retrained and evaluated using leave-one-subject-out cross validation performed separately for the healthy and ACHM data sets. There are no other published methods for rod detection to compare to.

To quantify the performance of our RAC-CNN for detecting rods and cones, we first matched the automatically detected photoreceptors to the photoreceptors marked by the first grader one-to-one for each image pair in a similar fashion as presented in Cunefare *et al.*[31]. We performed the matching separately for rods and cones. An automatically detected photoreceptor was considered a true positive if it was located within a distance d of a manually marked photoreceptor of the same type. The value d was set to the smaller between 0.75 of the median spacing between manually marked photoreceptors of the specific type in the image and an upper limit based on the type of photoreceptor. The upper limit for cones was set to 16 pixels based on the value used in Cunefare *et al.* [31], and the upper limit for rods was set to 5 pixels empirically based on rod size observed on a separate data set. Automatically detected photoreceptors that were not matched to a manually marked photoreceptor of the same type were considered false positives, and manually marked photoreceptors that did not have a matching automatically detected photoreceptor of the same type were considered false negatives. If a manually marked photoreceptor matched to more than one automatically detected photoreceptor of the same type, only the automatically marked photoreceptor with the smallest distance to the manually marked photoreceptor was considered a true positive, and the remaining were considered false positives. Finally, automatically detected and manually marked photoreceptors within 7 pixels of the edges of the image were removed to avoid border artefacts. After matching for one photoreceptor type in an image pair, the number of automatically marked photoreceptors ($N_{Automatic}$) and manually marked photoreceptors (N_{Manual}) of that type can then be expressed as:

$$N_{Automatic} = N_{TP} + N_{FP}, \tag{3}$$

$$N_{Manual} = N_{TP} + N_{FN}, \tag{4}$$

where N_{TP} is the number of true positives, N_{FP} is the number of false positives, and N_{FN} is the number of false negatives. For each image pair, we then calculated the true positive rate, false discovery rate, and Dice's coefficient [79,80] for rods and cones separately as:

$$\text{True positive rate} = N_{TP}/N_{Manual}, \tag{5}$$

$$\text{False discovery rate} = N_{FP}/N_{Automatic}, \tag{6}$$

$$\text{Dice's coefficient} = 2N_{TP}/(N_{Manual} + N_{Automatic}). \tag{7}$$

To assess inter-observer variability, the second set of manual markings was compared to the first set of manual markings following the same procedure. Additionally, the LF-DM-CNN results were computed in the same way for only cones.

4.　Results

We implemented and ran all methods in MATLAB 2017b (The MathWorks, Natick, MA) and used MatConvNet [77] 1.0-beta23 for training and running the RAC-CNN. We ran all experiments on a desktop PC with an i7-5930 K CPU at 3.5 GHz, 64 GB of RAM, and a GeForce GTX TITAN X GPU. Using the trained RAC-CNNs, the average run time for detecting and classifying rods and cones in new images was 0.2 seconds across the healthy data set (average image size of 236×237 pixels) and 0.2 seconds across the ACHM data set (average image size of 231×231 pixels). Using the trained LF-DM-CN, the average run time for detecting only cones was 30.2 seconds across the healthy data set and 29.0 seconds across the ACHM data set. The mean values and standard deviations of the automatically chosen photoreceptor detection parameters σ, H, and T across the validation groups are given in Table 1. The average time to train our RAC-CNN and to learn the detection parameters was under 2 hours for both the healthy and ACHM data sets.

Table 1. Average detection parameters across the healthy and ACHM validation groups (standard deviations shown in parenthesis).

	σ	H	T
Healthy-Rod	0.7 (0.2)	0.01 (0.02)	0.66 (0.12)
Healthy-Cone	2.5 (0.4)	0 (0)	0.48 (0.09)
ACHM-Rod	0.9 (0.2)	0.03 (0.03)	0.63 (0.08)
ACHM-Cone	2.2 (0.7)	0.01 (0.02)	0.66 (0.15)

The performance of the automatic methods with respect to the first set of manual markings over the 40 images in the healthy data set is summarized in Table 2, along with a comparison between the two manual graders. Table 3 shows the same performance comparisons across the 49 images in the ACHM data set. The median Dice's coefficient values across the healthy data set was 0.92 for rods and 0.99 for cones for the RAC-CNN, and 0.90 for rods and 0.98 for cones for the comparison between graders. The median Dice's coefficient values across the ACHM data set was 0.91 for rods and 0.90 for cones for the RAC-CNN, and 0.84 for rods and 0.89 for cones for the comparison between graders.

Table 2. Average performance of the automatic methods and second grader with respect to the first set of manual markings across the healthy data set (standard deviations shown in parenthesis).

	True positive rate	False discovery rate	Dice's coefficient
RAC-CNN (Rods)	**0.92** (0.06)	**0.10** (0.06)	**0.91** (0.04)
Grader #2 (Rods)	0.88 (0.06)	**0.10** (0.06)	0.89 (0.05)
RAC-CNN (Cones)	0.97 (0.04)	0.03 (0.03)	**0.97** (0.03)
LF-DM-CNN [31] (Cones)	**0.98** (0.04)	0.03 (0.04)	**0.97** (0.03)
Grader #2 (Cones)	0.97 (0.03)	**0.02** (0.03)	**0.97** (0.03)

Table 3. Average performance of the automatic methods and second grader with respect to the first set of manual markings across the ACHM data set (standard deviations shown in parenthesis).

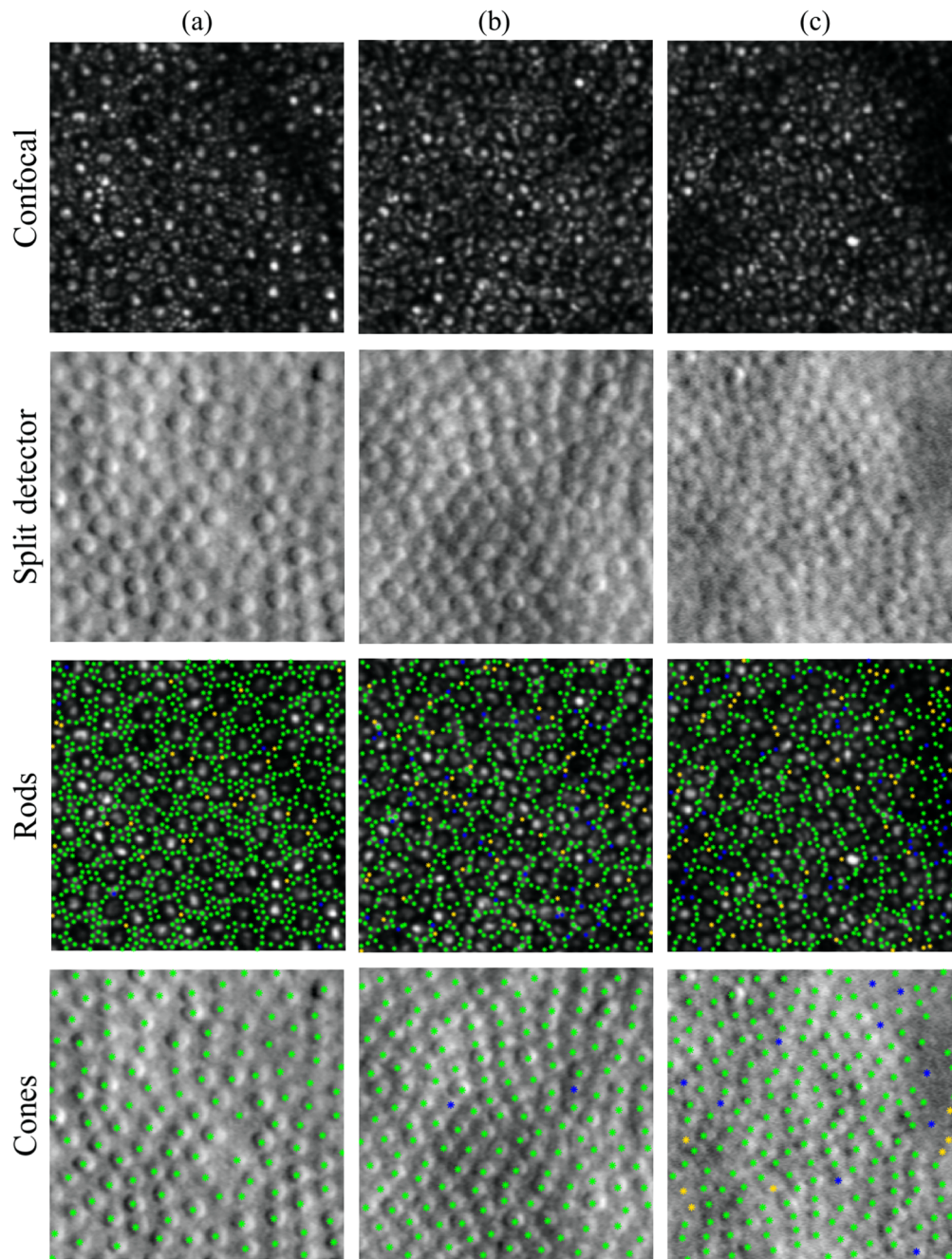|	True positive rate	False discovery rate	Dice's coefficient
RAC-CNN (Rods)	**0.91** (0.07)	0.11 (0.10)	**0.89** (0.05)
Grader #2 (Rods)	0.78 (0.07)	**0.08** (0.08)	0.84 (0.05)
RAC-CNN (Cones)	0.84 (0.14)	**0.08** (0.09)	**0.87** (0.10)
LF-DM-CNN [31] (Cones)	0.86 (0.12)	0.10 (0.10)	**0.87** (0.09)
Grader #2 (Cones)	**0.88** (0.11)	0.12 (0.09)	**0.87** (0.07)

Fig. 6. Performance of the RAC-CNN method on healthy images. Confocal AOSLO images from different subjects are shown on the top row, and the co-registered split detector AOSLO images are shown in the row second from the top. Rod detection results for the RAC-CNN method with respect to the first set of manual markings are shown on the second row from the bottom, and cone detection results are shown on the bottom row. Green points denote true positives, blue denotes false negatives, and gold denotes false positives. Dice's coefficients for the rods and cones are 0.98 and 1 in **(a)**, 0.94 and 0.99 in **(b)**, and 0.91 and 0.95 in **(c)**, respectively.
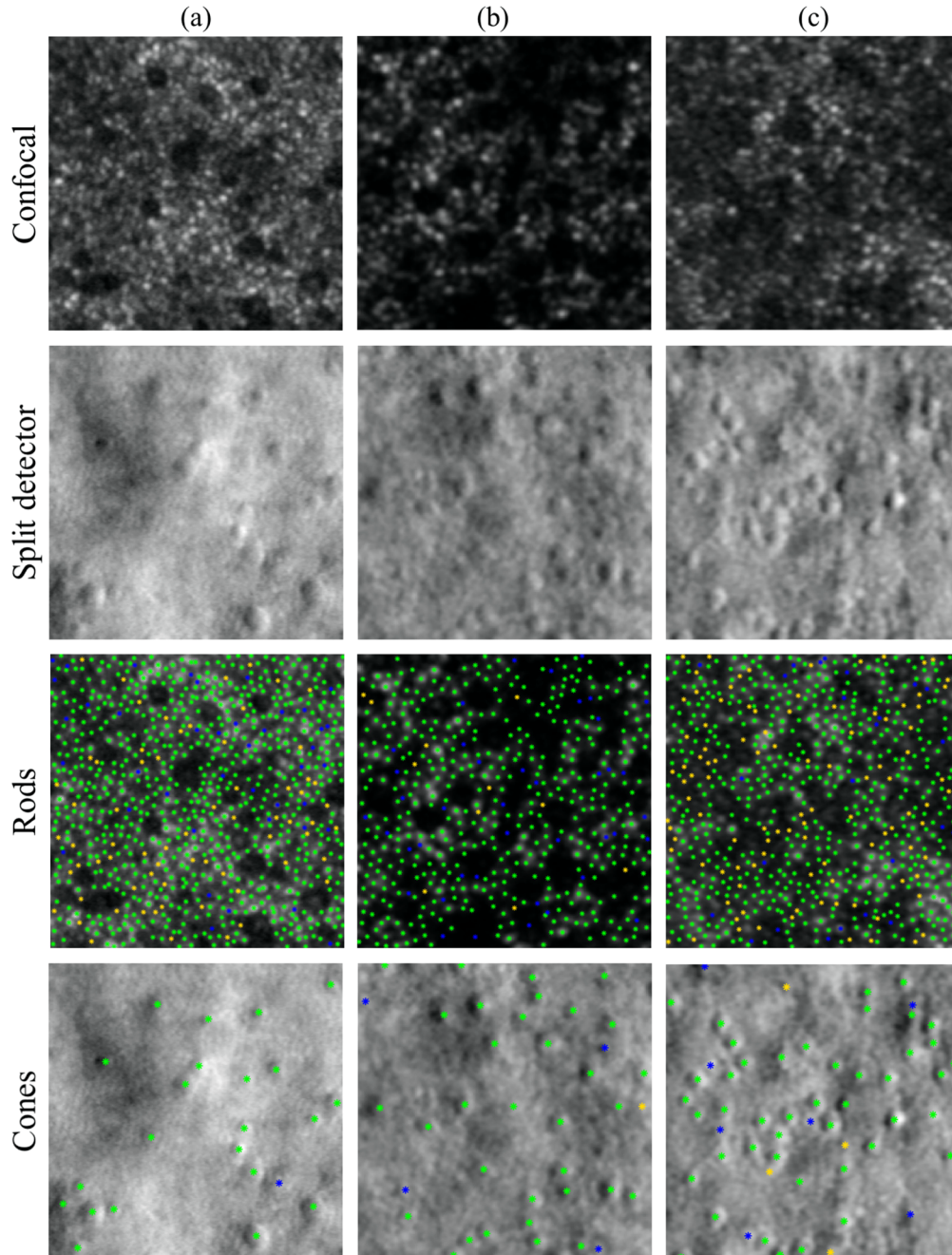
Fig. 7. Performance of the RAC-CNN method on ACHM images. Confocal AOSLO images from different subjects are shown on the top row, and the simultaneously captured split detector AOSLO images are shown in the row second from the top. Rod detection results for the RAC-CNN method with respect to the first set of manual markings are shown on the second row from the bottom, and cone detection results are shown on the bottom row. Green points denote true positives, blue denotes false negatives, and gold denotes false positives. Dice's coefficients for the rods and cones are 0.93 and 0.98 in **(a)**, 0.94 and 0.93 in **(b)**, and 0.89 and 0.88 in **(c)**, respectively.

Figure 6 shows examples of the performance of our RAC-CNN method with respect to markings by the first manual grader on images from the healthy data set, and Fig. 7 provides examples from the ACHM data set. In the marked images, automatically detected photoreceptors that were matched to a manually marked photoreceptor of the same type (true positives) are shown in green, photoreceptors missed by the RAC-CNN method (false negatives) are shown in blue, and automatically detected photoreceptors with no corresponding manually marked photoreceptor (false positive) are shown in gold. For the sake of visual clarity, the rods and cones are shown on separate images. Figure 8 provides comparisons of the performances of the automatic methods for cone detection.
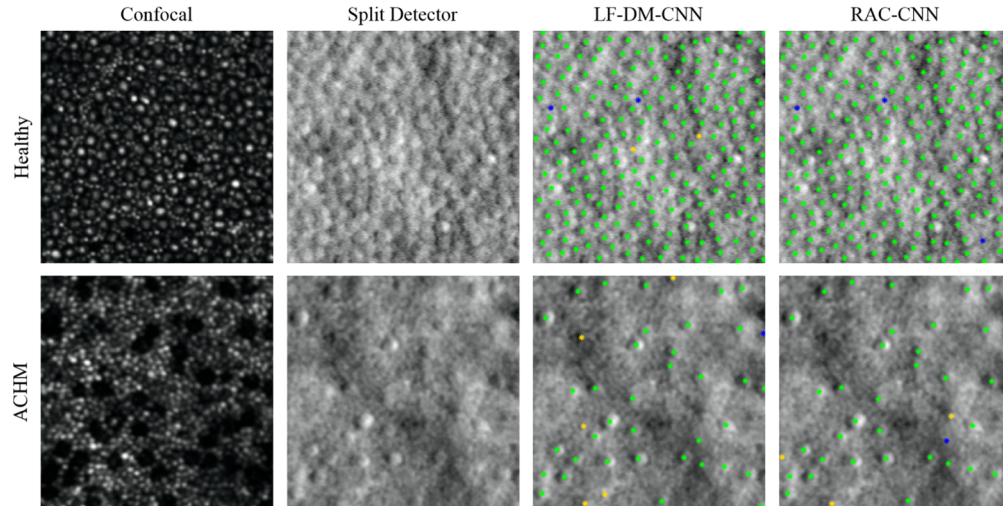

Fig. 8. Performance of the automated algorithms for cone detection in a healthy (top) and ACHM (bottom) image pair. Simultaneously captured confocal and split detector images are shown in the two left columns. Performance with respect to manual cone markings for the RAC-CNN and our previous LF-DM-CNN [31] methods are shown in the right two columns and displayed on the split detector images. Only cones are included in this figure as LF-DM-CNN cannot detect rods. Green points denote true positives, blue denotes false negatives, and gold denotes false positives. Dice's coefficients are 0.99 for both methods for the healthy image pair, and 0.92 for both methods for the ACHM image pair.

5.   Discussion

We developed the RAC-CNN, an automatic deep learning based method for detecting and classifying rod and cone photoreceptors in multimodal AOSLO images. Our semantic segmentation based RAC-CNN worked very fast, able to detect and classify rods and cones in 0.2 seconds on image sizes that would take our previous patch based CNNs [31,46] over 10 seconds to only detect cones. We validated our method on images taken over a wide range of retinal eccentricities from healthy subjects and from subjects with ACHM. We showed that our method had good agreement with the current gold standard of manual grading. Finally, we showed that the performance of our method was similar to that between two different manual graders.

In Table 2, we show the performance of the RAC-CNN, LF-DM-CNN, and manual methods over the healthy data set. There was a lower performance for detecting rods than cones. This is likely due to the increased difficulty in visualizing rods which are smaller than cones in non-foveal regions of the retina. Table 3 shows the performance of the RAC-CNN, LF-DM-CNN, and manual methods across the ACHM data set, which as expected was worse than those reported

for the healthy data set. There is often more uncertainty in identifying photoreceptors, and the images are generally nosier and blurrier in comparison to images taken from healthy subjects. Additionally, the healthy data set had extra averaging done to further enhance the quality of the confocal images which was not done for the ACHM set. Cone detection had a larger drop in performance in comparison to rod detection, likely because ACHM is a cone, rather than rod, dysfunction. On both healthy and ACHM images, performance of our RAC-CNN method was comparable to those reported by the state-of-the-art LF-DM-CNN cone-only detection method [31]. This is encouraging because we had previously shown that LF-DM-CNN cone detection performance is on a par with human grading. Yet, RAC-CNN was roughly 150 times faster and for the first time adds rod segmentation capability with human level accuracy.

As AOSLO is currently commonly utilized for assessment of known ophthalmic diseases, we utilized an automatic method to learn optimal parameters for the healthy and ACHM groups. For the sake of completeness, we tested an alternative scenario in which RAC-CNN and detection parameters are trained with leave-one-subject-out cross validation across both the healthy and ACHM data sets combined (with no indication given whether an image is from a healthy or pathological subject). The performance of the algorithm was identical at the reported precision of the Dice's coefficient value for cone detection in healthy and ACHM groups, and for rod detection in the ACHM group. Only, the Dice's coefficient value for rod detection in the healthy set was reduced from 0.91 to 0.90.

From Table 3 it can be seen that there is noticeable inter-observer variability for detecting photoreceptors in the ACHM data set, especially for rods. This is consistent with previous studies showing the difficulty in marking pathological images [33]. From the table it can be seen that the first manual grader had a tendency to mark more rods than the second grader in the ACHM data set. A CNN trained using the second set of manual markings would be expected to detect less rods than one trained on the first. Yet, if the goal is to measure the longitudinal change in photoreceptor density, it is expected that such biases to largely cancel out when comparing photoreceptor counts by a consistent automatic algorithm in the images of the same subject at different timepoints. Assessment and validation of the automated photoreceptor counting consistency in longitudinal studies is part of our future work.

To assess how each path contributed to the classification of the AOSLO images, we inspected the weights in final convolutional layer, which combines the features from the confocal and split detector paths. For the rod class in healthy subject networks, on average 67% and 33% of the weight magnitude was associated with the confocal and split detector information, respectively; while in ACHM subject networks on average 60% and 40% of the weight magnitude was associated with the confocal and split detector information, respectively. For both the healthy and diseased networks, more weight was given to the confocal modality for detecting rods, as rods are generally less visible in split detector AOSLO. However, the split detector information is still useful for resolving ambiguities sometimes seen in differentiating cones from rods in confocal AOSLO [31,45]. For the cone class in healthy subject networks, on average 59% and 41% of the weight magnitude was associated with the confocal and split detector information, respectively; while in ACHM subject networks on average 41% and 59% of the weight magnitude was associated with the confocal and split detector information, respectively. For detecting cones, more weight was given to the confocal information in the networks trained on healthy data, but more weight was given to the split detector information for the ACHM networks. This is consistent with the loss of ability to directly visualize cones in confocal AOSLO images of ACHM subjects. Finally, for the sake of completeness to further show the importance of using the information of both channels for detecting rods, we retrained the networks to use only the confocal information path for rod detection. We found that rod detection based only on the confocal channel reduced the accuracy as reported by Dice's coefficient from 0.91 to 0.90 in the healthy data set, and from 0.89 to 0.86 in the ACHM data set.

There are a few possible avenues to improve our RAC-CNN method. First, accurate manual segmentations of the rod and cone boundaries could be useful for training the RAC-CNN. As our manual markings only contained the rod and cone positions, the label maps for the data did not reflect the variations seen between individual photoreceptors. This could be especially important for cones, whose sizes can vary significantly at different retinal eccentricities. Accurate segmentations could allow the network to be better trained or allow it to be trained for segmentation as well as classification and localization. Additionally, our method could be improved by increasing the amount of training data. Studies on rod photoreceptors have been limited in part due to the subjective and time-consuming nature of manually marking rods in AOSLO. We hope that by providing an automated method for detecting rod and cone photoreceptors we will enable researchers to perform more studies on the rod mosaic, and in turn generate more data which may then be used to further refine this and other automated methods. Finally, rather than simply averaging the registered AOSLO frames together, a modified network that uses temporal information [81] is expected to further enhance rod and cone localization.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. A. Roorda and J. L. Duncan, "Adaptive optics ophthalmoscopy," Annu. Rev. Vis. Sci. **1**(1), 19–50 (2015).
2. S. A. Burns, A. E. Elsner, K. A. Sapoznik, R. L. Warner, and T. J. Gast, "Adaptive optics imaging of the human retina," Prog. Retinal Eye Res. **68**, 1–30 (2019).
3. A. Roorda and D. R. Williams, "The arrangement of the three cone classes in the living human eye," Nature **397**(6719), 520–522 (1999).
4. O. P. Kocaoglu, S. Lee, R. S. Jonnal, Q. Wang, A. E. Herde, J. C. Derby, W. Gao, and D. T. Miller, "Imaging cone photoreceptors in three dimensions and in time using ultrahigh resolution optical coherence tomography with adaptive optics," Biomed. Opt. Express **2**(4), 748–763 (2011).
5. M. Lombardo, S. Serrao, and G. Lombardo, "Technical factors influencing cone packing density estimates in adaptive optics flood illuminated retinal images," PLoS One **9**(9), e107402 (2014).
6. D. Merino, J. L. Duncan, P. Tiruveedhula, and A. Roorda, "Observation of cone and rod photoreceptors in normal subjects and patients using a new generation adaptive optics scanning laser ophthalmoscope," Biomed. Opt. Express **2**(8), 2189–2201 (2011).
7. K. E. Stepien, W. M. Martinez, A. M. Dubis, R. F. Cooper, A. Dubra, and J. Carroll, "Subclinical photoreceptor disruption in response to severe head trauma," Arch. Ophthalmol. **130**(3), 400–402 (2012).
8. H. Song, E. A. Rossi, L. Latchney, A. Bessette, E. Stone, J. J. Hunter, D. R. Williams, and M. Chung, "Cone and rod loss in stargardt disease revealed by adaptive optics scanning light ophthalmoscopydistribution of cone and rod loss in stargardt diseasedistribution of cone and rod loss in stargardt disease," JAMA Ophthalmol. **133**(10), 1198–1203 (2015).
9. A. Dubra, Y. Sulai, J. L. Norris, R. F. Cooper, A. M. Dubis, D. R. Williams, and J. Carroll, "Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope," Biomed. Opt. Express **2**(7), 1864–1876 (2011).
10. R. F. Cooper, A. M. Dubis, A. Pavaskar, J. Rha, A. Dubra, and J. Carroll, "Spatial and temporal variation of rod photoreceptor reflectance in the human retina," Biomed. Opt. Express **2**(9), 2577–2589 (2011).

11. J. Liang, D. R. Williams, and D. T. Miller, "Supernormal vision and high-resolution retinal imaging through adaptive optics," J. Opt. Soc. Am. A **14**(11), 2884–2892 (1997).

12. A. Roorda, F. Romero-Borja, I. I. I. W. Donnelly, H. Queener, T. Hebert, and M. Campbell, "Adaptive optics scanning laser ophthalmoscopy," Opt. Express **10**(9), 405–412 (2002).

13. A. Dubra and Y. Sulai, "Reflective afocal broadband adaptive optics scanning ophthalmoscope," Biomed. Opt. Express **2**(6), 1757–1768 (2011).

14. T. DuBose, D. Nankivil, F. LaRocca, G. Waterman, K. Hagan, J. Polans, B. Keller, D. Tran-Viet, L. Vajzovic, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Handheld adaptive optics scanning laser ophthalmoscope," Optica **5**(9), 1027–1036 (2018).

15. N. Sredar, O. E. Fagbemi, and A. Dubra, "Sub-airy confocal adaptive optics scanning ophthalmoscopy," Transl. Vis. Sci. Techn. **7**(2), 17 (2018).

16. T. B. DuBose, F. LaRocca, S. Farsiu, and J. A. Izatt, "Super-resolution retinal imaging using optically reassigned scanning laser ophthalmoscopy," Nat. Photonics **13**(4), 257–262 (2019).

17. D. Scoles, Y. N. Sulai, C. S. Langlo, G. A. Fishman, C. A. Curcio, J. Carroll, and A. Dubra, "In vivo imaging of human cone photoreceptor inner segments," Invest. Ophthalmol. Visual Sci. **55**(7), 4244–4251 (2014).

18. E. A. Rossi, C. E. Granger, R. Sharma, Q. Yang, K. Saito, C. Schwarz, S. Walters, K. Nozato, J. Zhang, T. Kawakami, W. Fischer, L. R. Latchney, J. J. Hunter, M. M. Chung, and D. R. Williams, "Imaging individual neurons in the retinal ganglion cell layer of the living eye," Proc. Natl. Acad. Sci. U. S. A. **114**(3), 586–591 (2017).

19. K. A. Sapoznik, T. Luo, A. de Castro, L. Sawides, R. L. Warner, and S. A. Burns, "Enhanced retinal vasculature imaging with a rapidly configurable aperture," Biomed. Opt. Express **9**(3), 1323–1333 (2018).

20. D. C. Gray, W. Merigan, J. I. Wolfing, B. P. Gee, J. Porter, A. Dubra, T. H. Twietmeyer, K. Ahmad, R. Tumbar, F. Reinholz, and D. R. Williams, "In vivo fluorescence imaging of primate retinal ganglion cells and retinal pigment epithelial cells," Opt. Express **14**(16), 7144–7158 (2006).

21. J. J. Hunter, B. Masella, A. Dubra, R. Sharma, L. Yin, W. H. Merigan, G. Palczewska, K. Palczewski, and D. R. Williams, "Images of photoreceptors in living primate eyes using adaptive optics two-photon ophthalmoscopy," Biomed. Opt. Express **2**(1), 139–148 (2011).

22. A. Pinhas, M. Dubow, N. Shah, T. Y. Chui, D. Scoles, Y. N. Sulai, R. Weitz, J. B. Walsh, J. Carroll, A. Dubra, and R. B. Rosen, "In vivo imaging of human retinal microvasculature using adaptive optics scanning light ophthalmoscope fluorescein angiography," Biomed. Opt. Express **4**(8), 1305–1317 (2013).

23. R. Sharma, D. R. Williams, G. Palczewska, K. Palczewski, and J. J. Hunter, "Two-photon autofluorescence imaging reveals cellular structures throughout the retina of the living primate eyetwo-photon autofluorescence imaging," Invest. Ophthalmol. Visual Sci. **57**(2), 632–646 (2016).

24. J. A. Feeks and J. J. Hunter, "Adaptive optics two-photon excited fluorescence lifetime imaging ophthalmoscopy of exogenous fluorophores in mice," Biomed. Opt. Express **8**(5), 2483–2495 (2017).

25. B. Hermann, E. J. Fernández, A. Unterhuber, H. Sattmann, A. F. Fercher, W. Drexler, P. M. Prieto, and P. Artal, "Adaptive-optics ultrahigh-resolution optical coherence tomography," Opt. Lett. **29**(18), 2142–2144 (2004).

26. R. Zawadzki, S. Jones, S. Olivier, M. Zhao, B. Bower, J. Izatt, S. Choi, S. Laut, and J. Werner, "Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal in vivo imaging," Opt. Express **13**(21), 8532–8546 (2005).

27. C. Torti, B. Považay, B. Hofer, A. Unterhuber, J. Carroll, P. K. Ahnelt, and W. Drexler, "Adaptive optics optical coherence tomography at 120,000 depth scans/s for non-invasive cellular phenotyping of the living human retina," Opt. Express **17**(22), 19382–19400 (2009).

28. Y. Jian, J. Xu, M. A. Gradowski, S. Bonora, R. J. Zawadzki, and M. V. Sarunic, "Wavefront sensorless adaptive optics optical coherence tomography for in vivo retinal imaging in mice," Biomed. Opt. Express **5**(2), 547–559 (2014).

29. J. Polans, D. Cunefare, E. Cole, B. Keller, P. S. Mettu, S. W. Cousins, M. J. Allingham, J. A. Izatt, and S. Farsiu, "Enhanced visualization of peripheral retinal vasculature with wavefront sensorless adaptive optics optical coherence tomography angiography in diabetic patients," Opt. Lett. **42**(1), 17–20 (2017).

30. Z. Liu, K. Kurokawa, F. Zhang, J. J. Lee, and D. T. Miller, "Imaging and quantifying ganglion cells and other transparent neurons in the living human retina," Proc. Natl. Acad. Sci. U. S. A. **114**(48), 12803–12808 (2017).

31. D. Cunefare, C. S. Langlo, E. J. Patterson, S. Blau, A. Dubra, J. Carroll, and S. Farsiu, "Deep learning based detection of cone photoreceptors with multimodal adaptive optics scanning light ophthalmoscope images of achromatopsia," Biomed. Opt. Express **9**(8), 3740–3756 (2018).

32. P. Godara, M. Wagner-Schuman, J. Rha, T. B. Connor, K. E. Stepien, and J. Carroll, "Imaging the photoreceptor mosaic with adaptive optics: Beyond counting cones," in *Retinal Degenerative Diseases* (Springer US, 2012), 451–458.

33. M. A. Abozaid, C. S. Langlo, A. M. Dubis, M. Michaelides, S. Tarima, and J. Carroll, "Reliability and repeatability of cone density measurements in patients with congenital achromatopsia," in *Advances in Experimental Medicine and Biology*, C. Bowes Rickman, M. M. LaVail, R. E. Anderson, C. Grimm, J. Hollyfield, and J. Ash, eds. (Springer International Publishing, Cham, 2016), pp. 277–283.

34. K. Y. Li and A. Roorda, "Automated identification of cone photoreceptors in adaptive optics retinal images," J. Opt. Soc. Am. A **24**(5), 1358–1363 (2007).

35. B. Xue, S. S. Choi, N. Doble, and J. S. Werner, "Photoreceptor counting and montaging of en-face retinal images from an adaptive optics fundus camera," J. Opt. Soc. Am. A **24**(5), 1364–1372 (2007).

36. D. H. Wojtas, B. Wu, P. K. Ahnelt, P. J. Bones, and R. P. Millane, "Automated analysis of differential interference contrast microscopy images of the foveal cone mosaic," J. Opt. Soc. Am. A **25**(5), 1181–1189 (2008).

37. A. Turpin, P. Morrow, B. Scotney, R. Anderson, and C. Wolsley, "Automated identification of photoreceptor cones using multi-scale modelling and normalized cross-correlation," in *Image analysis and processing – iciap* 2011, G. Maino and G. Foresti, eds. (Springer Berlin Heidelberg, 2011), pp. 494–503.

38. R. Garrioch, C. Langlo, A. M. Dubis, R. F. Cooper, A. Dubra, and J. Carroll, "Repeatability of in vivo parafoveal cone density and spacing measurements," Optom. Vis. Sci. **89**(5), 632–643 (2012).

39. F. Mohammad, R. Ansari, J. Wanek, and M. Shahidi, "Frequency-based local content adaptive filtering algorithm for automated photoreceptor cell density quantification," in *Proceedings of IEEE International Conference on Image Processing*, (IEEE, 2012), 2325–2328.

40. S. J. Chiu, C. A. Toth, C. Bowes Rickman, J. A. Izatt, and S. Farsiu, "Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming," Biomed. Opt. Express **3**(5), 1127–1140 (2012).

41. S. J. Chiu, Y. Lokhnygina, A. M. Dubis, A. Dubra, J. Carroll, J. A. Izatt, and S. Farsiu, "Automatic cone photoreceptor segmentation using graph theory and dynamic programming," Biomed. Opt. Express **4**(6), 924–937 (2013).

42. R. F. Cooper, C. S. Langlo, A. Dubra, and J. Carroll, "Automatic detection of modal spacing (Yellott's ring) in adaptive optics scanning light ophthalmoscope images," Ophthalmic Physiol. Opt. **33**(4), 540–549 (2013).

43. L. Mariotti and N. Devaney, "Performance analysis of cone detection algorithms," J. Opt. Soc. Am. A **32**(4), 497–506 (2015).

44. D. M. Bukowska, A. L. Chew, E. Huynh, I. Kashani, S. L. Wan, P. M. Wan, and F. K. Chen, "Semi-automated identification of cones in the human retina using circle hough transform," Biomed. Opt. Express **6**(12), 4676–4693 (2015).

45. D. Cunefare, R. F. Cooper, B. Higgins, D. F. Katz, A. Dubra, J. Carroll, and S. Farsiu, "Automatic detection of cone photoreceptors in split detector adaptive optics scanning light ophthalmoscope images," Biomed. Opt. Express **7**(5), 2036–2050 (2016).

46. D. Cunefare, L. Fang, R. F. Cooper, A. Dubra, J. Carroll, and S. Farsiu, "Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks," Sci. Rep. **7**(1), 6620 (2017).

47. J. Liu, H. Jung, A. Dubra, and J. Tam, "Automated photoreceptor cell identification on nonconfocal adaptive optics images using multiscale circular voting," Invest. Ophthalmol. Visual Sci. **58**(11), 4477–4489 (2017).

48. C. Bergeles, A. M. Dubis, B. Davidson, M. Kasilian, A. Kalitzeos, J. Carroll, A. Dubra, M. Michaelides, and S. Ourselin, "Unsupervised identification of cone photoreceptors in non-confocal adaptive optics scanning light ophthalmoscope images," Biomed. Opt. Express **8**(6), 3081–3094 (2017).

49. M. Heisler, M. J. Ju, M. Bhalla, N. Schuck, A. Athwal, E. V. Navajas, M. F. Beg, and M. V. Sarunic, "Automated identification of cone photoreceptors in adaptive optics optical coherence tomography images using transfer learning," Biomed. Opt. Express **9**(11), 5353–5367 (2018).

50. B. Davidson, A. Kalitzeos, J. Carroll, A. Dubra, S. Ourselin, M. Michaelides, and C. Bergeles, "Automatic cone photoreceptor localisation in healthy and stargardt afflicted retinas using deep learning," Sci. Rep. **8**(1), 7911 (2018).

51. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, (Springer, 2012), 1097–1105.

52. V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, and J. Cuadros, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA **316**(22), 2402–2410 (2016).

53. M. J. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," IEEE Trans. Med. Imag. **35**(5), 1273–1284 (2016).

54. M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," Invest. Ophthalmol. Visual Sci. **57**(13), 5200–5206 (2016).

55. S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," Biomed. Opt. Express **8**(2), 579–592 (2017).

56. A. D. Desai, C. Peng, L. Fang, D. Mukherjee, A. Yeung, S. J. Jaffe, J. B. Griffin, and S. Farsiu, "Open-source, machine and deep learning-based automated algorithm for gestational age estimation through smartphone lens imaging," Biomed. Opt. Express **9**(12), 6038–6052 (2018).

57. P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," IEEE Trans. Med. Imag. **35**(11), 2369–2380 (2016).

58. Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," IEEE Trans. Med. Imag. **35**(1), 109–118 (2016).

59. H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer International Publishing, 2016), 132–139.

60. 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 **8**(5), 2732–2744 (2017).

61. 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 **8**(8), 3627–3642 (2017).

62. 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 **9**(6), 2681–2698 (2018).

63. X. Fei, J. Zhao, H. Zhao, D. Yun, and Y. Zhang, "Deblurring adaptive optics retinal images using deep convolutional neural networks," Biomed. Opt. Express **8**(12), 5675–5687 (2017).

64. C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," J. Comp. Neurol. **292**(4), 497–523 (1990).

65. C. S. Langlo, E. J. Patterson, B. P. Higgins, P. Summerfelt, M. M. Razeen, L. R. Erker, M. Parker, F. T. Collison, G. A. Fishman, C. N. Kay, J. Zhang, R. G. Weleber, P. Yang, D. J. Wilson, M. E. Pennesi, B. L. Lam, J. Chiang, J. D. Chulay, A. Dubra, W. W. Hauswirth, and J. Carroll, "Residual foveal cone structure in cngb3-associated achromatopsia," Invest. Ophthalmol. Visual Sci. **57**(10), 3984–3995 (2016).

66. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Springer International Publishing, 2015), 234–241.

67. H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *Medical Image Understanding and Analysis* (Springer International Publishing, 2017), 506–517.

68. T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger, "U-net: Deep learning for cell counting, detection, and morphometry," Nat. Methods **16**(1), 67–70 (2019).

69. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International Conference on Machine Learning*, (2015), 448–456.

70. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, (IEEE, 2010), 807–814.

71. K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *IEEE 12th International Conference on Computer Vision*, (IEEE, 2009), 2146–2153.

72. H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, (2015), 1520–1528.

73. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).

74. A. Dubra and Z. Harvey, "Registration of 2D images from fast scanning ophthalmic instruments," in *Biomedical Image Registration* (Springer Berlin Heidelberg, 2010), 60–71.

75. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 3431–3440.

76. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," arXiv preprint arXiv:1607.08022 (2016).

77. A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*, (ACM, Brisbane, Australia, 2015), pp. 689–692.

78. P. Soille, *Morphological Image Analysis: Principles and Applications* (Springer Science & Business Media, 2013).

79. L. R. Dice, "Measures of the amount of ecologic association between species," Ecology **26**(3), 297–302 (1945).

80. T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons," Biol. Skr. **5**(1), 1–34 (1948).

81. S. Soltanian-Zadeh, K. Sahingur, S. Blau, Y. Gong, and S. Farsiu, "Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning," Proc. Natl. Acad. Sci. U. S. A. **116**(17), 8554–8563 (2019).