7-26-2017

Open Source Software for Automatic Detection of Cone Photoreceptors in Adaptive Optics Ophthalmoscopy Using Convolutional Neural Networks

David Cunefare
Duke University

Leyuan Fang
Duke University

Robert F. Cooper
Marquette University

Alfredo Dubra
Marquette University

Joseph Carroll
Marquette University

See next page for additional authors

Authors
David Cunefare, Leyuan Fang, Robert F. Cooper, Alfredo Dubra, Joseph Carroll, and Sina Farsiu

SCIENTIFIC REPORTS

Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks

David Cunefare[1], Leyuan Fang[1], Robert F. Cooper[2,3], Alfredo Dubra[4], Joseph Carroll[5,6] & Sina Farsiu[1,7]

Imaging with an adaptive optics scanning light ophthalmoscope (AOSLO) enables direct visualization of the cone photoreceptor mosaic in the living human retina. Quantitative analysis of AOSLO images typically requires manual grading, which is time consuming, and subjective; thus, automated algorithms are highly desirable. Previously developed automated methods are often reliant on ad hoc rules that may not be transferable between different imaging modalities or retinal locations. In this work, we present a convolutional neural network (CNN) based method for cone detection that learns features of interest directly from training data. This cone-identifying algorithm was trained and validated on separate data sets of confocal and split detector AOSLO images with results showing performance that closely mimics the gold standard manual process. Further, without any need for algorithmic modifications for a specific AOSLO imaging system, our fully-automated multi-modality CNN-based cone detection method resulted in comparable results to previous automatic cone segmentation methods which utilized ad hoc rules for different applications. We have made free open-source software for the proposed method and the corresponding training and testing datasets available online.

The structure of the retinal photoreceptor mosaic has long been of scientific and clinical interest[1]. A key group of technologies used for visualizing photoreceptors *in vivo* is adaptive optics (AO) ophthalmoscopy[2–9]. AO retinal images have been used to analyze various facets of normal[2, 10–16] and pathologic photoreceptor mosaics[17–24]. The most widely used of these technologies is the confocal AO scanning light ophthalmoscope (AOSLO)[3], which is capable of visualizing the smallest photoreceptors, rods and foveal cones[13] (indeed, AO enhanced optical coherence tomography systems may also visualize rods and foveal cones[25]). More recently, a new generation of AOSLO modalities has been developed that takes advantage of non-confocal information that is lost in confocal systems[26–29]. Non-confocal split detector AOSLO has been shown to provide complementary information to that obtained from confocal AOSLO. It has been suggested that combining multiple modalities could be beneficial for accurately analyzing photoreceptor mosaic properties[30], or for improving the performance of image processing applications such as registration[31].

To quantify geometrical properties of the photoreceptor mosaic, often the locations of each individual cone in an image must be found. Since manual grading of these images is subjective and time consuming, several automated methods for detecting cones in AO images have been developed[32]. These automated algorithms use techniques including local intensity maxima detection[33–36], model based correlation[37], circular Hough transform[38],

[1]Department of Biomedical Engineering, Duke University, Durham, NC, 27708, USA. [2]Department of Ophthalmology, Scheie Eye Institute, University of Pennsylvania, Philadelphia, PA, 19104, USA. [3]Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104, USA. [4]Department of Ophthalmology, Stanford University, Palo Alto, CA, 94303, USA. [5]Department of Biomedical Engineering, Marquette University, Milwaukee, WI, 53233, USA. [6]Department of Ophthalmology & Visual Sciences, Medical College of Wisconsin, Milwaukee, WI, 53226, USA. [7]Department of Ophthalmology, Duke University Medical Center, Durham, NC, 27710, USA. Correspondence and requests for materials should be addressed to D.C. (email: david.cunefare@duke.edu)
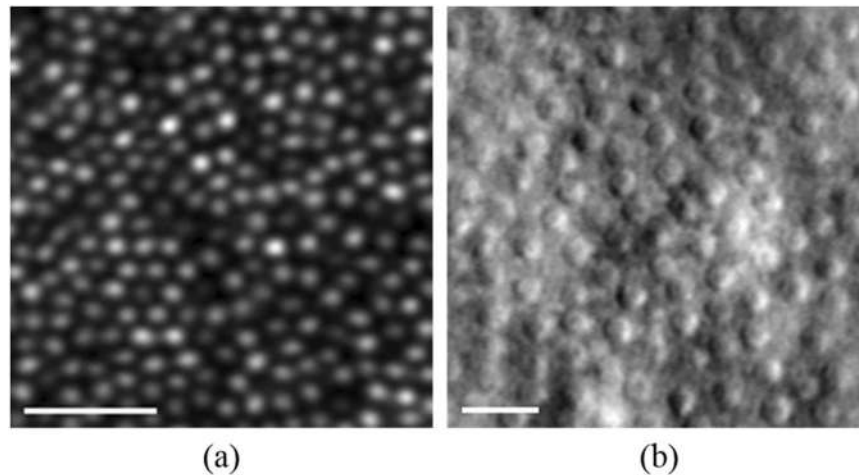
Figure 1. AOSLO cone imaging. (**a**) Confocal AOSLO image at 0.65° from the fovea. (**b**) Split detector AOSLO image at approximately 8° from the fovea. *Scale bars*: 20 μm.
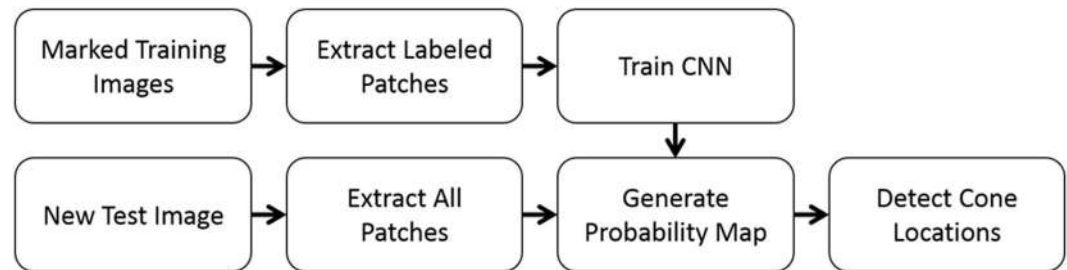
Figure 2. CNN based cone detection algorithm schematic.

graph-theory and dynamic programming (GTDP)[39], and estimation of cone spatial frequency[40–42]. These fixed mathematical model based methods have shown good performance for the problems they were designed for (e.g. specific imaging modality, resolution, cone density, etc.). However, reliance on ad hoc rules and specific algorithmic parameters does not allow them to become generalizable to alternative imaging conditions. As an example, in our previous publications, despite developing a highly accurate segmentation method for confocal AOSLO images, we had to devise a completely new segmentation method in order to detect cones in split detector AOSLO images due to the disparity in cone appearances between the two modalities, as can be seen in Fig. 1.

Novel machine learning techniques provide an alternative approach where features of interest are learned directly from data, allowing for a higher degree of adaptability[43]. Unlike classic model-based techniques, one can conceptually use the same core machine learning algorithm for a variety of image segmentation problems by using different training images. Of note, deep convolution neural networks (CNNs)[44–47] have shown high performance in a multitude of image analysis tasks including over a range of fields, including ophthalmic imaging. CNNs have been used for the segmentation of retinal blood vessels[48–50], detection of diabetic retinopathy[51] and hemorrhage[52] in retinal fundus images, and classification of pathology[53] or segmentation of retinal layers[54] in optical coherence tomography images.

In this work, we present a fully automatic method for training a CNN from manually segmented AOSLO images, and then for using the trained CNN in order to detect cones in previously unseen images. We validated the performance of our method on two separate data sets from confocal and split detector AOSLO systems. To promote future research, we provide an open-source MATLAB implementation of our algorithm.

Methods

Our CNN based cone detection method is outlined in Fig. 2. In brief, we first take a data set of images from an imaging modality (e.g. confocal AOSLO) with manually marked cones and extract patches centered around cone and non-cone locations using the markings. We then use these patches to train a CNN classifier. To detect cones in previously unseen images from this imaging modality, we then classify overlapping patches centered at each pixel to create a probability map. This probability map is then processed to locate the individual cone locations. To detect cones in an alternative AOSLO modality (e.g. split detector AOSLO), all that needs to be done is to exchange the previous training dataset with one from the new imaging modality.

Data Sets. We used two separate data sets for the confocal and split detector AOSLO cases. These images were previously used for development and validation of cone segmentation algorithms[39, 42].
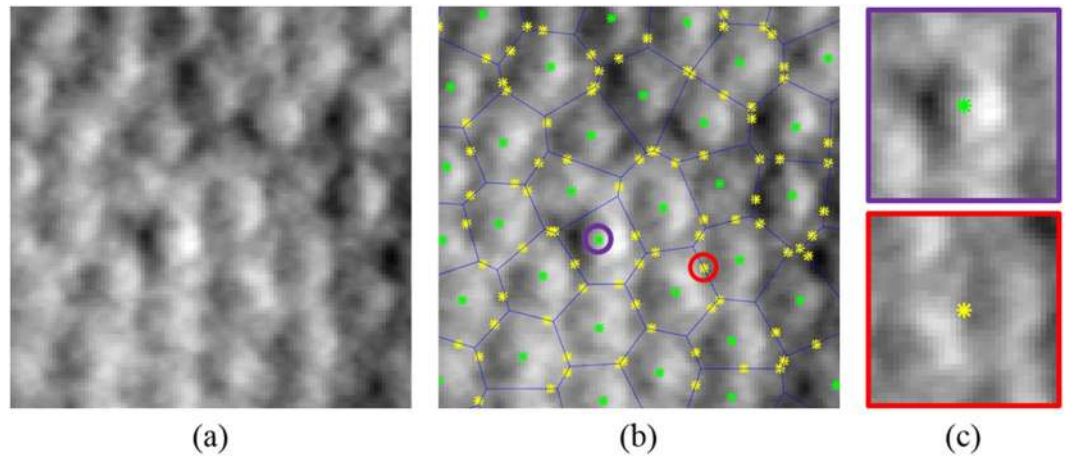
Figure 3. Extraction of labeled patches from cone images. (**a**) Original cropped split detector AOSLO image. (**b**) Image (**a**) with Voronoi diagram overlain in blue. Manually marked cones are shown in green and randomly generated locations along Voronoi edges in yellow. (**c**) Example cone (top-purple) and non-cone (bottom-red) patches (size 33×33 pixels) from positions circled in (**b**) with center marked.

The confocal AOSLO experiments were based on the Garrioch *et al.*[36] data set, which was later used by Chiu *et al.*[39] to train and validate the GTDP cone segmentation method. In brief, this data set consists of 840 images (150×150 pixels) from 21 subjects (20 normal and one subject with deuteranopia) acquired with a previously described confocal AOSLO system[8, 13] with a 0.96° field-of-view. For each subject, four positions 0.65° from the center of fixation (bottom left, bottom right, top left, and top right) were imaged, and 10 averaged images were acquired per location. Axial length measurements were acquired with an IOL Master (Carl Zeiss Meditec, Dublin, CA) and used to determine the lateral resolution of the images. A region-of-interest (ROI) of size 150×150 pixels was extracted from the center of each image and used for analysis. The ROIs' size ranged from $58 \times 58\,\mu m^2$ to $71 \times 71\,\mu m^2$. An expert grader semi-automatically marked cone photoreceptors in all images[39].

The split detector AOSLO experiments used an extension of the data set used by Cunefare *et al.* for training and validating the AFLD cone detection method[42]. The original set contained 112 images (average image size of 208×208 pixels) from 14 subjects with normal vision obtained from the Advanced Ocular Imaging Program image bank (MCW, Milwaukee, Wisconsin). These images were acquired using a previously described split detector AOSLO system[8, 28] with a 1.0° field-of-view. For each subject, eight images were acquired from a single randomly selected meridian (superior, temporal, inferior, and nasal) over a range of eccentricities (500 to 2800 μm). Lateral scale/sampling for each subject was determined using an IOL Master (Carl Zeiss Meditec Inc., Dublin, California, USA). An ROI was extracted from each image and used for analysis. The size of each ROI was chosen as a function of retinal eccentricity so that the ROI would contain approximately 100 cones[55]. In addition to the published data set in Cunefare *et al.*[42], we acquired a new set of 152 split detector images (ROIs with an average size of 216×216 pixels) from four subjects with normal vision using the same split detector system. The ROIs' size across the entire split detector set ranged from $93 \times 93\,\mu m^2$ to $111 \times 111\,\mu m^2$. An expert grader manually marked cone photoreceptors in all images.

We learned separate CNN weights and algorithmic parameters for the cases of confocal images, split detector images, and a combination of both modalities. For the case of confocal AOSLO imaging, we used 200 images from 5 subjects for training the CNN, and 640 images from the remaining 16 subjects for validating the proposed method. The images from the subject with deuteranopia were used in the validation data set. For the case of split detector AOSLO, we used 184 images from 8 subjects for training. We evaluated the performance of the CNN by using the same 80 images from 10 subjects that were used for validation of our previous study[42]. For the combined case, we added both the confocal and split detector training data sets to form a single mixed training data set. In this case, the network was blinded to the type of AOSLO image. There was no overlap in subjects between the training and validation sets for either the confocal or split detector cases.

Image Preprocessing and Patch extraction. We normalized all images so that their intensity values stretched between 0 and 255. Each neural network was trained by first extracting patches centered on cone and non-cone pixel locations. For the set corresponding to cone locations, we extracted patches of size 33×33 pixels centered on every manually marked position in the training images. The patch size was chosen empirically to be large enough to contain cones and their surrounding features. Our training data sets did not include direct manual markings for the non-cone areas. Thus, we devised a simple technique using Voronoi diagrams[33] to select the non-cone locations. First, the manually marked cone positions are used to generate a Voronoi diagram, as shown in Fig. 3(b), where each cell consists of all positions that have a smaller Euclidian distance to the contained cone position than any other cone position. The boundaries between these cells are called Voronoi edges, which we assumed to correspond with the space between individual cones. Subsequently, in order to generate the non-cone set, we randomly selected a single point from each Voronoi edge for all training images, rounded to the nearest pixel value, and then extracted a 33×33 pixel patch around this location (Fig. 3(c)). Patches that would extended outside of the image were excluded. Over 34000 cone and 88000 non-cone patches were extracted from

| Layer number	Type	Input size	Filter size	Stride	Number of kernels/nodes
1	Convolution	$33 \times 33 \times 1$	$5 \times 5 \times 1$	1	32
2	Batch Normalization	$33 \times 33 \times 32$	—	—	—
3	Max Pooling	$33 \times 33 \times 32$	3×3	2	—
4	ReLU	$16 \times 16 \times 32$	—	—	—
5	Convolution	$16 \times 16 \times 32$	$5 \times 5 \times 32$	1	32
6	Batch Normalization	$16 \times 16 \times 32$	—	—	—
7	ReLU	$16 \times 16 \times 32$	—	—	—
8	Average Pooling	$16 \times 16 \times 32$	3×3	2	—
9	Convolution	$8 \times 8 \times 32$	$5 \times 5 \times 32$	1	64
10	Batch Normalization	$8 \times 8 \times 64$	—	—	—
11	ReLU	$8 \times 8 \times 64$	—	—	—
12	Average Pooling	$8 \times 8 \times 64$	3×3	2	—
13	Fully Connected	$4 \times 4 \times 64$	$4 \times 4 \times 64$	—	64
14	Batch Normalization	$1 \times 1 \times 64$	—	—	—
15	ReLU	$1 \times 1 \times 64$	—	—	—
16	Fully Connected	$1 \times 1 \times 64$	$1 \times 1 \times 64$	—	2
17	Soft-max	$1 \times 1 \times 2$	—	—	—

Table 1. Architecture of the CNN.

the confocal training data set, and over 20000 cone and 49000 non-cone patches were extracted from the split detector training set.

Convolutional Neural Network. We slightly modified a Cifar[44] CNN architecture implemented in the MatConvNet[56] CNN toolbox. In general, a simple CNN accepts an input image or patch and applies a sequence of transforming layers in order to extract features and finally classify the input. Table 1 shows the sequence of layers used in our network. Convolutional layers convolve an input of size $W \times H \times D$ (before padding) with N kernels of size $w \times h \times D$ to get an output of size $W \times H \times N$. The output can be thought of as a stack of N feature maps, each generated by the corresponding kernel. For each of these feature maps, the CNN adds a potentially different bias value. The pooling layers apply either a max or average operation over the first two dimensions in a 3×3 window while down sampling the input by a factor of 2. This lowers the computational burden and improves robustness to small image distortions by increasing spatial invariance[57]. Rectified linear units (ReLU) layers[58] transform their inputs by setting all negative values to 0. ReLU layers are used to speed up the training process[44] and improve the performance of the network by adding non-linearity[57]. Batch normalization layers[59] normalize their inputs based on mean and variance statistics. These layers are used to reduce internal covariate shift which speeds up the training process and helps prevent overfitting. Fully connected layers consolidate all information from their preceding layer. Each output node in a fully connected layer is obtained by a weighted combination of all values from the previous layer with an additional bias term added for each node. The final fully connected layer provides a score for each class (cone and non-cone), which are input into a soft-max layer[60] that outputs the probability of the original patch belonging to each class. Figure 4(a) provides an illustrative example of all layer types used in our CNN.

In order for the CNNs to be effective, the convolutional and fully connected weights and biases must be learned. We trained separate networks for the confocal and split detector cases, which we label the confocal CNN (C-CNN) and split detector CNN (SD-CNN). Each network was trained using the extracted training patches and their class labels as inputs. Additionally, we trained a combined network, which we labeled the mixed CNN (M-CNN), using the patches from both the confocal and split detector training images without providing the network any indication which modality the patches originated from. The initial weight parameters are randomly initialized, and the initial biases are set to zero. Then, the weights and biases are refined through stochastic gradient descent and back-propagation[46]. The training data is split into mini-batches with 100 patches per mini-batch, and each iteration of the gradient descent occurs over one mini-batch. We repeated this process for all batches spanning the training data set (called an epoch[48]), and trained over 45 epochs for both networks. The weight learning rate was initially set to 0. 001 for all layers except the last fully connected layer (layer 12) where it was set to 0.0001. Bias learning rates were similarly set to 0.1 for all layers except the last fully connected layer where it was set to 0.01. All learning rates were decreased by a factor of 10 twice over training, at epochs 31 and 41. Weight decay was left at the default value of 0.0001 in the MatConvNet[56] CNN toolbox. Figure 4(b–d) displays the feature maps output after each sequence of convolutional, batch normalization, pooling, and ReLU layers in the trained SD-CNN for a single cone input patch.

Cone Detection. After training the CNNs, we used them to find cone locations within an image. To do this, we first extracted a 33×33 pixel patch around every pixel in the image after normalizing the image's intensity values. Symmetric padding by mirroring the pixel values at the edge by half of the patch length (16 pixels) was used for patches that extended beyond the edges of the image. We then used the corresponding CNN to determine the probability that each patch is centered on a cone. The associated probability for each patch is then used to
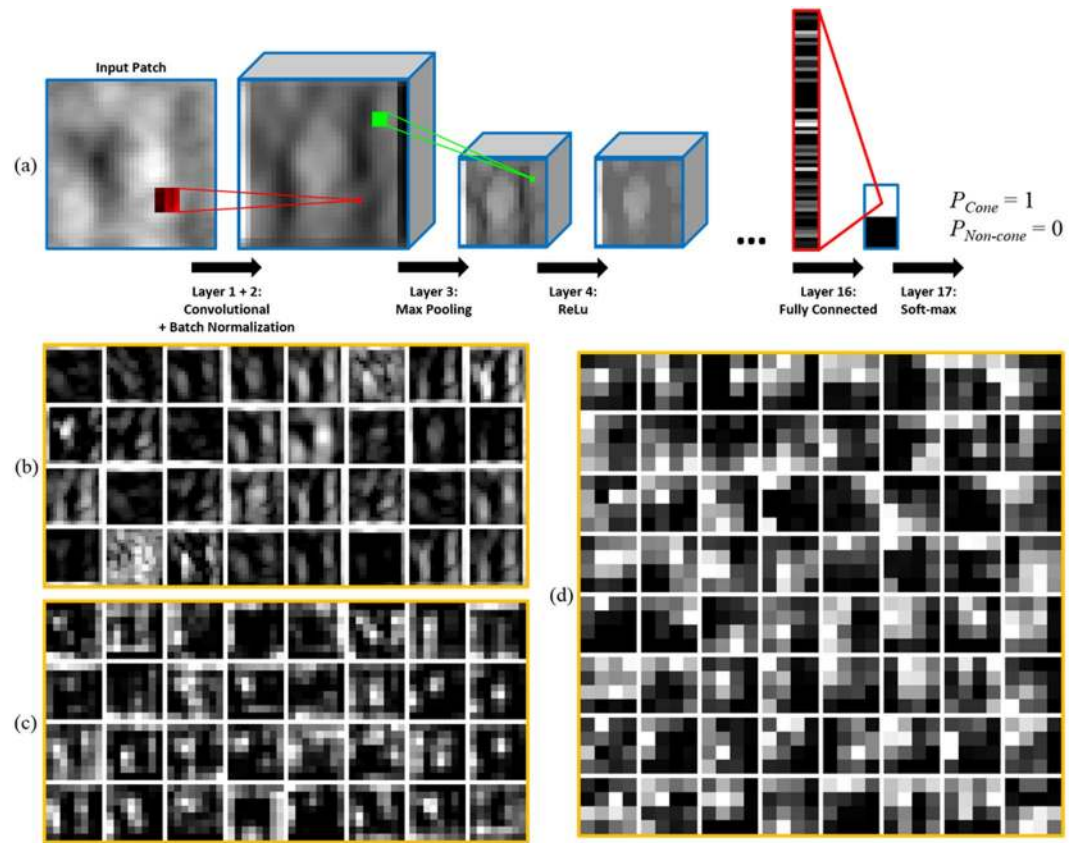
Figure 4. Convolutional network properties. (**a**) Visualization of convolutional, pooling, ReLu, fully connected, and soft-max layers. (**b**–**d**) Output feature maps from (**b**) Layer 4 (**c**) Layer 8, and (**d**) Layer 12 when the input patch from (**a**) is processed by the trained SD-CNN.

construct a probability map with the same dimensions as the original image, as shown in Fig. 5(b). We smoothed the map with a Gaussian filter with standard deviation σ. We then applied the extended-maxima transform using MATLAB's *imextendedmax* function[61], which finds maximal regions where the height difference in the region is less than or equal to H. This results in a binary image as shown in Fig. 5(c). We then found all connected clusters in the binary image to use as candidates for cone locations, and eliminated weak candidates by removing any cluster whose maximum value in the filtered probability map is less than a threshold T. Finally, the locations of cones in the image are determined by finding the centers of the remaining clusters (Fig. 5(d)). The values σ, H, and T were automatically chosen to be 1.3, 0, and 0.3 for the C-CNN case; 2, 0.1, and 0.5 for the SD-CNN case; and 0.4, 0.25, and 0.9 for the M-CNN case. These automatically chosen parameters were learned by maximizing the average Dice's coefficient (as explained in the following section) across the training images over a set of potential parameter combinations.

Validation. We validated the results of our method by comparing to the current gold standard of manual grading across the validation data sets. In the case of M-CNN, we provided no explicit information to the network about the type of the AOSLO image to be tested (confocal or split-detector). Additionally, we compared the performance of our CNN based method to the performance of the GTDP cone detection method[39] for the confocal set, and to the AFLD method[42] for the split detector set. To quantify the performance of the automatic methods, we matched automatically marked cones to manually marked cones one-to-one. In brief, an automatic cone was considered a true positive if it was located within some distance d of a manually marked cone. The value d was set to 0.75 of the median spacing between manually marked cones in the image. Automatically detected cones that were not matched to a manually marked cone were considered as false positives, and manually marked cones that did not have a matching automatically detected cone were considered as false negatives. If a manually marked cone was matched to more than one automatically marked cone, only the one with the smallest distance was considered a true positive, and the remaining were considered false positives. Finally, we removed manually and automatically marked cones that were within 7 pixels of the edges of the image to remove border artifacts. For each image, the number of automatically marked cones ($N_{Automatic}$) and manually marked cones (N_{Manual}) can then be broken down into the number of true positives (N_{TP}), false positives (N_{FP}), and false negatives (N_{FN}) as:

$$N_{Automatic} = N_{TP} + N_{FP}, \qquad (1)$$
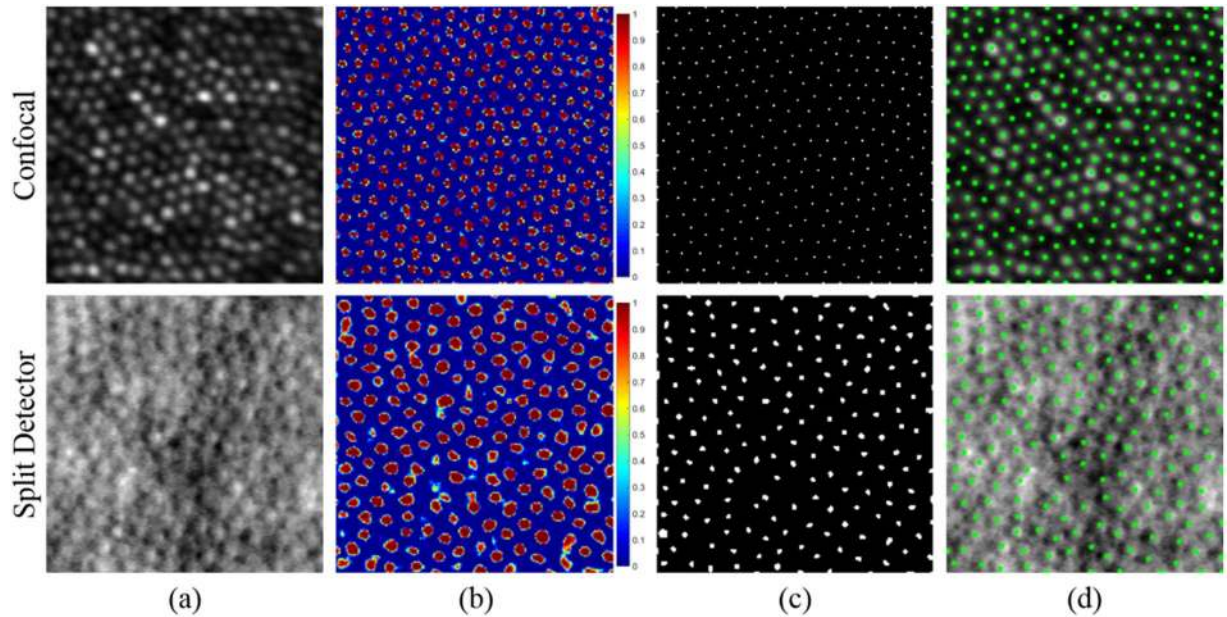
Figure 5. Detection of cones in confocal (top row) and split detector (bottom row) AOSLO images. (**a**) Original images. (**b**) Probability maps generated from (**a**) using the C-CNN (top) and SD-CNN (bottom). (**c**) Extended maxima of (**b**). (**d**) Detected cones marked in green on the image shown in (**a**).

$$N_{Manual} = N_{TP} + N_{FN}. \tag{2}$$

In order to evaluate the performance of the automatic methods, we then calculate the true positive rate, false discovery rate, and Dice's coefficient[62, 63] for each image as:

$$\text{True positive rate} = N_{TP}/N_{Manual}, \tag{3}$$

$$\text{False discovery rate} = N_{FP}/N_{Automatic}, \tag{4}$$

$$\text{Dice's coefficient} = 2N_{TP}/(N_{Manual} + N_{Automatic}). \tag{5}$$

Additionally, we also evaluated cone density measurements from each method, which is a commonly used quantitative metric for analyzing photoreceptor mosaics[64]. Cone density is defined as the ratio of the number of cones in an image to the area of that image. We compared cone densities calculated from each automatic method to the cone density from manual grading using Bland-Altman analysis[65].

Data Availability. The datasets generated during and analyzed during the current study are available at https://github.com/DavidCunefare/CNN-Cone-Detection. These include both training and validation datasets for confocal and split detector AOSLO and the corresponding open source software.
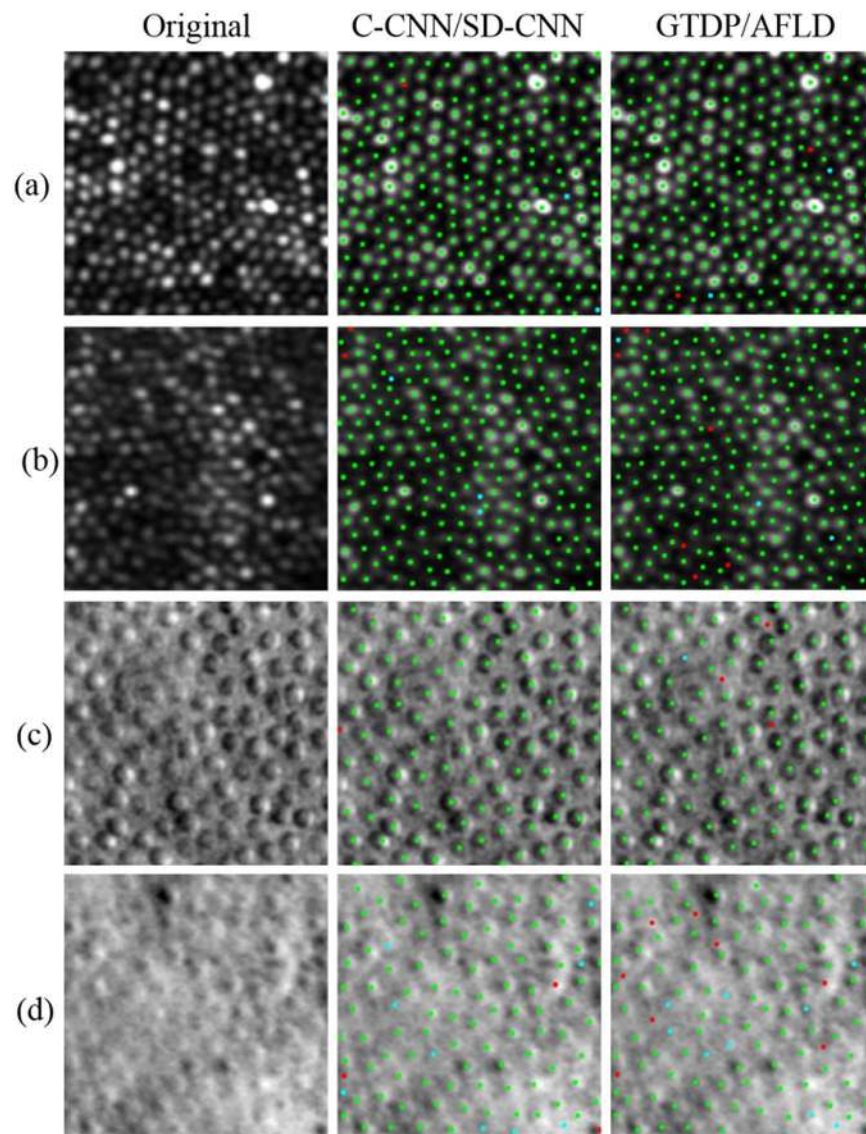
Results

We implemented the CNN based detection method in MATLAB 2016b (The MathWorks, Natick, MA) with MatConvNet[56] 1.0-beta23. We ran the algorithm on a desktop PC with an i7-5930K CPU at 3.5 GHz, 64 GB of RAM, and a GeForce GTX TITAN X GPU. The average run time for our CNN based detection method on a new image after training was 6 seconds per image for the confocal AOSLO data set (image size of 150 by 150 pixels) and 12 seconds per image for the split detector AOSLO data set (average image size of 206.5 by 206.5 pixels). These times were the same regardless of whether the C-CNN, SD-CNN, or M-CNN was used. The total training time was under 3 hours for the C-CNN and SD-CNN, and under 6 hours for the M-CNN. This included the time for extracting training patches, training the CNN, saving the probability maps of all training images, and choosing the detection parameters. Note that this offline training only needs to be completed once.

Figure 6 displays the results of the automated algorithms in comparison to manual grading in four examples. In the marked images, a green point indicates an automatically detected cone that was matched to a manually marked cone (true positive), a cyan point indicates a cone missed by the automatic algorithm (false negative), and a red point indicates an automatic marking with no corresponding manually marked cone (false positive).

The performances of the automated algorithms in comparison to manual grading are summarized in Tables 2 and 3. Table 2 reports the performance of the C-CNN, M-CNN and GTDP methods across the confocal AOSLO validation data set, and Table 3 reports the performance of the SD-CNN, M-CNN and AFLD methods across the split detector AOSLO validation data set. The median Dice's coefficients across the confocal validation set were

Figure 6. Performance of the automated algorithms on confocal (**a**,**b**) and split detector (**c**,**d**) images. Original Images are shown in the left column, C-CNN for confocal and SD-CNN for split detector cone detection results in the middle column, and AFLD for split detector and GTDP for confocal results in the right column. Green points denote true positives, cyan denotes false negatives, and red denotes false positives. Dice's coefficients for our CNN based method are 0.994 in (**a**), 0.990 in (**b**), 0.995 in (**c**), and 0.941 in (**d**). Dice's coefficient for the GTDP method are 0.992 in (**a**) and 0.978 in (**b**), and for the AFLD method are 0.979 in (**c**) and 0.911 in (**d**).

	True positive rate	False discovery rate	Dice's coefficient
C-CNN	0.989 (0.012)	0.008 (0.014)	0.990 (0.010)
M-CNN	0.975 (0.026)	0.003 (0.007)	0.986 (0.015)
GTDP	0.990 (0.011)	0.015 (0.018)	0.988 (0.012)

Table 2. Average performance of the C-CNN, M-CNN, and GTDP methods with respect to manual marking across the confocal validation data set (standard deviations shown in parenthesis).

0.993 for the C-CNN, 0.990 for the M-CNN, and 0.991 for the GTDP methods. The median Dice's coefficients across the split detector validation set were 0.970 for the SD-CNN, 0.968 for the M-CNN, and 0.958 for the AFLD methods. As noted previously, one subject used in the confocal validation data set had deuteranopia. The average true positive rate, false discovery rate, and Dice's coefficient for this subject using the C-CNN were found to be 0.993, 0.007, and 0.993, respectively. It can be seen that the performance of the general purpose CNN method is very comparable to the custom designed GTDP and AFLD techniques.
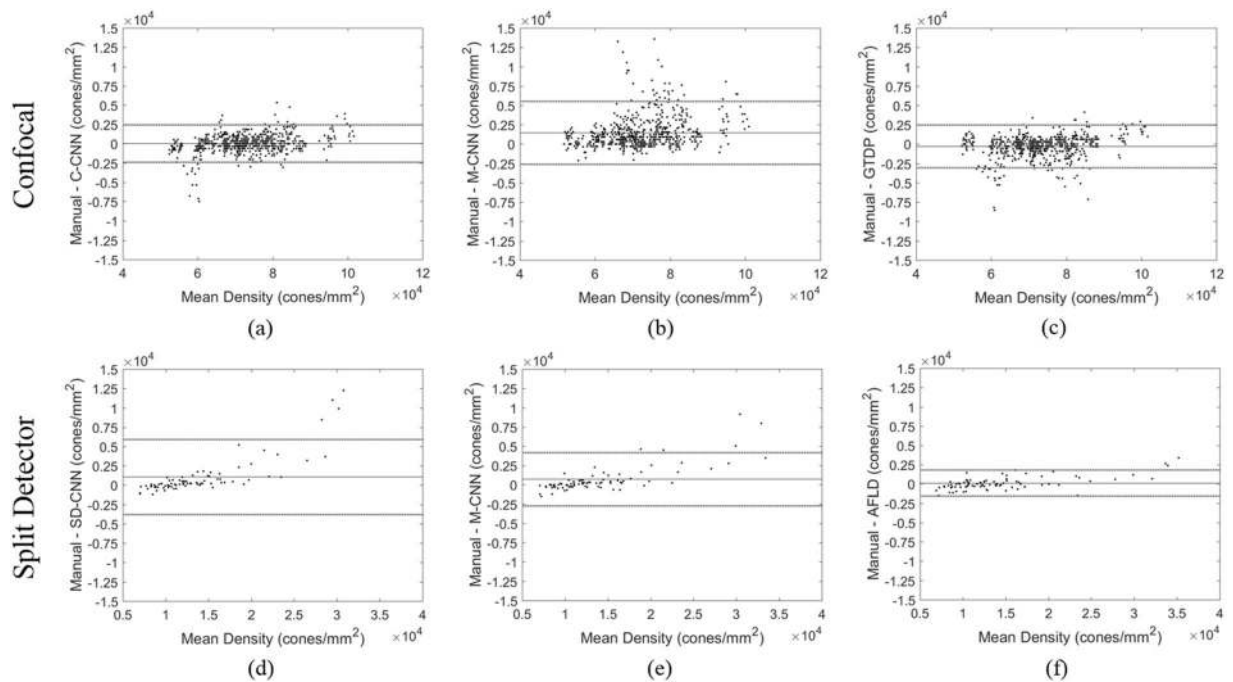
Figure 7. Bland-Altman plots comparing cone density for (**a**) manual – C-CNN on confocal, (**b**) manual – M-CNN on confocal, (**c**) manual – GTDP on confocal, (**d**) manual – SD-CNN on split detector, (**e**) manual – M-CNN on split detector, and (**f**) manual – AFLD on split detector. The solid black line shows the expected value of the difference, and the dotted lines show the 95% confidence limits of agreement.

	True positive rate	False discovery rate	Dice's coefficient
SD-CNN	0.943 (0.075)	0.027 (0.034)	0.955 (0.045)
M-CNN	0.949 (0.057)	0.034 (0.036)	0.956 (0.034)
AFLD	0.961 (0.031)	0.054 (0.045)	0.952 (0.028)

Table 3. Average performance of the SD-CNN, M-CNN, and AFLD methods with respect to manual marking across the split detector validation data set (standard deviations shown in parenthesis).

Bland-Altman plots comparing cone density measurements between the automatic methods and manual grading are shown in Fig. 7. The central solid line shows the average difference between methods, and the surrounding dotted lines show the 95% confidence limits.

Discussion

In this work, we developed an automatic CNN based method for detecting cone photoreceptors in AOSLO images. Using manually marked images from a particular AOSLO imaging modality, our method can train a CNN to extract features of interest and classify cones in unseen images from the same imaging modality. We tested our method on images from confocal and split detector AOSLO systems, and showed that or method had good agreement with the current gold standard of manual marking for both modalities. We have made the open-source code for our CNN based method freely available online to allow other researchers to test and modify the algorithm for their specific applications.

As can be seen in Tables 2 and 3, our CNN based algorithm performed comparably to the state-of-the-art GTDP and AFLD cone detection methods with respect to manual grading. This is highly encouraging, because the noted previous techniques heavily utilized modality specific ad hoc rules, which limited their application to other imaging scenarios and require further algorithm modification and development for new imaging scenarios. Our CNN approach, on the other hand, required no ad hoc rules, and was applied without any algorithmic modification to confocal and split detector AOSLO images. All that was needed to adapt the algorithm to confocal and split detector AOSLO was the corresponding training datasets. Additionally, in the case of the M-CNN, our algorithm was able to learn network filter weights to accurately detect cones in split detector and confocal images without any knowledge of which modality the image was acquired with.

For both the confocal and split detector AOSLO data sets, our CNN-based method had a slightly worse true positive rate but a slightly better false discovery rate than the other corresponding algorithm. This is also reflected in Fig. 7, where the CNN based method is biased towards underestimating the number of cones, especially in images with higher cone densities. This could be improved by using training data sets with a larger representation of images with high cone densities. Moreover, poor inter-observer agreement in grading AOSLO images[66] can negatively affect the performance of learning based methods such as CNN. Utilization of datasets graded by multiple observers is expected to improve performance.

The quantitative performance metrics presented for the GTDP and AFLD algorithms have negligible differences in comparison to those reported in Chiu *et al*.[39] and Cunefare *et al*.[42], respectively, for two reasons. First, these differences arise from changes made to the methods for quantifying performance (e.g. the number of cropped pixels in the boundaries) so that a uniform method could be used for both data sets. Second, we used a subset of the validation set used in Chiu *et al*.[39] for training the CNN, which was naturally excluded from the validation data set in this work.

It should be noted that there is space for improvement our proposed algorithm. We choose the network architecture and hyper-parameters empirically to provide good performance for both confocal and split detector AOSLO images. It is possible these parameters could be optimized to provide better performance, especially for a single imaging modality. Additionally, we expect to further improve the performance of the method for specific applications by applying custom pre-processing steps such as adaptive denoising[67].

Since our CNN based method learned to detect cones directly from training data, it can be easily extended to other imaging modalities or imaging conditions with different features (e.g. images from different retinal eccentricities that may contain rods or images of diseased eyes). Extending the algorithm to images from subjects with retinal pathologies is of particular importance. Several retinal diseases lead to alterations of the cone mosaic (e.g. age-related macular degeneration, achromatopsia, retinitis pigmentosa, and Usher syndrome), and quantitative analysis of these mosaics captured with AOSLO is potentially useful for the characterization, early diagnosis, and prognosis of these diseases[30, 68]. By sharing the open source software of this paper freely available online, we encourage our colleagues to test this algorithm for new imaging conditions. Of course, due to the variability of cone photoreceptor manifestation in AOSLO images of different diseases or eccentricities, we expect that utilizing appropriate new training data sets that match the test data set (e.g. training and testing on the same eccentricity or disease condition) will improve the performance of the algorithm. Detailed evaluation of our algorithm's performance for different diseases, imaging eccentricities, and images containing rod photoreceptors is part of our ongoing work.

References

1. Sawides, L., de Castro, A. & Burns, S. A. The organization of the cone photoreceptor mosaic measured in the living human retina. *Vision Res.* **132**, 34–44 (2017).
2. Roorda, A. & Williams, D. R. The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–522 (1999).
3. Roorda, A. *et al*. Adaptive optics scanning laser ophthalmoscopy. *Opt. Express* **10**, 405–412 (2002).
4. Zawadzki, R. *et al*. Adaptive-optics optical coherence tomography for high-resolution and high-speed 3d retinal *in vivo* imaging. *Opt. Express* **13**, 8532–8546 (2005).
5. Merino, D., Dainty, C., Bradu, A. & Podoleanu, A. G. Adaptive optics enhanced simultaneous en-face optical coherence tomography and scanning laser ophthalmoscopy. *Opt. Express* **14**, 3345–3353 (2006).
6. Torti, C. *et al*. Adaptive optics optical coherence tomography at 120,000 depth scans/s for non-invasive cellular phenotyping of the living human retina. *Opt. Express* **17**, 19382–19400 (2009).
7. Ferguson, R. D. *et al*. Adaptive optics scanning laser ophthalmoscope with integrated wide-field retinal imaging and tracking. *J. Opt. Soc. Am. A* **27**, A265–A277 (2010).
8. Dubra, A. & Sulai, Y. Reflective afocal broadband adaptive optics scanning ophthalmoscope. *Biomed. Opt. Express* **2**, 1757–1768 (2011).
9. Jonnal, R. S., Kocaoglu, O. P., Wang, Q., Lee, S. & Miller, D. T. Phase-sensitive imaging of the outer retina using optical coherence tomography and adaptive optics. *Biomed. Opt. Express* **3**, 104–124 (2012).
10. Roorda, A. & Williams, D. R. Optical fiber properties of individual human cones. *J. Vis.* **2**, 404–412 (2002).
11. Kitaguchi, Y. *et al*. *In vivo* measurements of cone photoreceptor spacing in myopic eyes from images obtained by an adaptive optics fundus camera. *Jpn. J. Ophthalmol.* **51**, 456–461 (2007).
12. Chui, T. Y., Song, H. & Burns, S. A. Adaptive-optics imaging of human cone photoreceptor distribution. *J. Opt. Soc. Am. A* **25**, 3021–3029 (2008).
13. Dubra, A. *et al*. Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope. *Biomed. Opt. Express* **2**, 1864–1876 (2011).
14. Pircher, M. *et al*. Temporal changes of human cone photoreceptors observed *in vivo* with SLO/OCT. *Biomed. Opt. Express* **2**, 100–112 (2011).
15. Kocaoglu, O. P. *et al*. Imaging cone photoreceptors in three dimensions and in time using ultrahigh resolution optical coherence tomography with adaptive optics. *Biomed. Opt. Express* **2**, 748–763 (2011).
16. Lombardo, M., Serrao, S. & Lombardo, G. Technical factors influencing cone packing density estimates in adaptive optics flood illuminated retinal images. *PLoS ONE* **9**, e107402 (2014).
17. Choi, S. S. *et al*. *In vivo* imaging of the photoreceptor mosaic in retinal dystrophies and correlations with visual function. *Invest. Ophthalmol. Vis. Sci.* **47**, 2080–2092 (2006).
18. Duncan, J. L. *et al*. High-resolution imaging with adaptive optics in patients with inherited retinal degeneration. *Invest. Ophthalmol. Vis. Sci.* **48**, 3283–3291 (2007).
19. Choi, S. S., Zawadzki, R. J., Greiner, M. A., Werner, J. S. & Keltner, J. L. Fourier-domain optical coherence tomography and adaptive optics reveal nerve fiber layer loss and photoreceptor changes in a patient with optic nerve drusen. *J. Neuroophthalmol.* **28**, 120–125 (2008).
20. Ooto, S. *et al*. High-resolution imaging of resolved central serous chorioretinopathy using adaptive optics scanning laser ophthalmoscopy. *Ophthalmology* **117**, 1800–1809 (2010).
21. Merino, D., Duncan, J. L., Tiruveedhula, P. & Roorda, A. Observation of cone and rod photoreceptors in normal subjects and patients using a new generation adaptive optics scanning laser ophthalmoscope. *Biomed. Opt. Express* **2**, 2189–2201 (2011).

22. Kitaguchi, Y., Kusaka, S., Yamaguchi, T., Mihashi, T. & Fujikado, T. Detection of photoreceptor disruption by adaptive optics fundus imaging and fourier-domain optical coherence tomography in eyes with occult macular dystrophy. *Clin. Ophthalmol.* **5**, 345–351 (2011).

23. Stepien, K. *et al.* Subclinical photoreceptor disruption in response to severe head trauma. *Arch. Ophthalmol.* **130**, 400–402 (2012).

24. Lammer, J. *et al.* Cone photoreceptor irregularity on adaptive optics scanning laser ophthalmoscopy correlates with severity of diabetic retinopathy and macular edemacone mosaic irregularity in diabetic eyes on AOSLO. *Invest. Ophthalmol. Vis. Sci.* **57**, 6624–6632 (2016).

25. Felberer, F. *et al.* Adaptive optics SLO/OCT for 3d imaging of human photoreceptors *in vivo*. *Biomed. Opt. Express* **5**, 439–456 (2014).

26. Chui, T. Y. P., VanNasdale, D. A. & Burns, S. A. The use of forward scatter to improve retinal vascular imaging with an adaptive optics scanning laser ophthalmoscope. *Biomed. Opt. Express* **3**, 2537–2549 (2012).

27. Scoles, D., Sulai, Y. N. & Dubra, A. *In vivo* dark-field imaging of the retinal pigment epithelium cell mosaic. *Biomed. Opt. Express* **4**, 1710–1723 (2013).

28. Scoles, D. *et al.* *In vivo* imaging of human cone photoreceptor inner segments. *Invest. Ophthalmol. Vis. Sci.* **55**, 4244–4251 (2014).

29. Rossi, E. A. *et al.* Imaging individual neurons in the retinal ganglion cell layer of the living eye. *Proc. Natl. Acad. Sci. USA* **114**, 586–591 (2017).

30. Roorda, A. & Duncan, J. L. Adaptive optics ophthalmoscopy. *Annu. Rev. Vis. Sci.* **1**, 19–50 (2015).

31. Chen, M. *et al.* Multi-modal automatic montaging of adaptive optics retinal images. *Biomed. Opt. Express* **7**, 4899–4918 (2016).

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

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

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

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

36. Garrioch, R. *et al.* Repeatability of *in vivo* parafoveal cone density and spacing measurements. *Optom. Vis. Sci.* **89**, 632–643 (2012).

37. Turpin, A., Morrow, P., Scotney, B., Anderson, R. & Wolsley, C. Automated identification of photoreceptor cones using multi-scale modelling and normalized cross-correlation. In *Image analysis and processing – iciap 2011* Vol. 6978 (eds Giuseppe Maino & GianLuca Foresti) 494–503 (Springer Berlin Heidelberg, 2011).

38. Bukowska, D. M. *et al.* Semi-automated identification of cones in the human retina using circle hough transform. *Biomed. Opt. Express* **6**, 4676–4693 (2015).

39. Chiu, S. J. *et al.* Automatic cone photoreceptor segmentation using graph theory and dynamic programming. *Biomed. Opt. Express* **4**, 924–937 (2013).

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

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

42. Cunefare, D. *et al.* Automatic detection of cone photoreceptors in split detector adaptive optics scanning light ophthalmoscope images. *Biomed. Opt. Express* **7**, 2036–2050 (2016).

43. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).

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

45. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

46. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural. Netw.* **61**, 85–117 (2015).

47. Greenspan, H. & Ginneken, B. v. & Summers, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Trans. Med. Imag.* **35**, 1153–1159 (2016).

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

49. Li, Q. *et al.* A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans. Med. Imag.* **35**, 109–118 (2016).

50. Fu, H. *et al.* Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International Conference on Medical Image Computing and Computer-Assisted Intervention.* (eds Sebastien Ourselin *et al.*) 132–139 (Springer International Publishing, 2016).

51. Gulshan, V., Peng, L. & Coram, M. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).

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

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

54. Fang, L. *et al.* 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**, 2732–2744 (2017).

55. Cooper, R. F., Wilk, M. A., Tarima, S. & Carroll, J. Evaluating descriptive metrics of the human cone mosaicdescriptive metrics of the human cone mosaic. *Invest. Ophthalmol. Vis. Sci.* **57**, 2992–3001 (2016).

56. Vedaldi, A. & Lenc, K. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia.* 689–692 (ACM, 2015).

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

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

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

60. Bishop, C. M. Pattern recognition and machine learning. (Springer, 2006).

61. Soille, P. *Morphological image analysis: Principles and applications.* (Springer Science & Business Media, 2013).

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

63. Sørensen, T. 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**, 1–34 (1948).

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

65. Martin Bland, J. & Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**, 307–310 (1986).

66. Abozaid, M. A. *et al.* Reliability and repeatability of cone density measurements in patients with congenital achromatopsia. In *Advances in experimental medicine and biology* Vol. 854 (eds Catherine Bowes Rickman *et al.*) 277–283 (Springer International Publishing, 2016).

67. Fang, L., Li, S., Cunefare, D. & Farsiu, S. Segmentation based sparse reconstruction of optical coherence tomography. *IEEE Trans. Med. Imag.* **36**, 407–421 (2017).

68. Sun, L. W. *et al.* Assessing photoreceptor structure in retinitis pigmentosa and Usher syndrome. *Invest. Ophthalmol. Vis. Sci.* **57**, 2428–2442 (2016).

Acknowledgements

Author Contributions

D.C. developed the method, wrote the first draft of the manuscript, and prepared all figures and tables. L.F. provided expertise in convolutional neural networks. R.C., A.D., and J.C. acquired data used and provided expertise in AOSLO imaging and image interpretation. S.F. initiated and supervised the project and provided guidance and expertise in image processing. All authors reviewed and edited the manuscript.

Additional Information

Competing Interests: D.C., S.F., J.C., and R.C. have filed an invention disclosure for the work presented in this paper.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.