



OPEN

# Neural network assisted annotation and analysis tool to study *in-vivo* foveolar cone photoreceptor topography

Aleksandr Gutnikov<sup>1,5</sup>, Patrick Hähn-Schumacher<sup>2,5</sup>, Julius Ameln<sup>1</sup>, Shekoufeh Gorgi Zadeh<sup>2,4</sup>, Thomas Schultz<sup>2,3</sup> & Wolf Harmening<sup>1</sup>✉

The foveola, the central region of the human retina, plays a crucial role in sharp color vision and is challenging to study due to its unique anatomy and technical limitations in imaging. We present ConeMapper, an open-source MATLAB software that integrates a fully convolutional neural network (FCN) for the automatic detection and analysis of cone photoreceptors in confocal adaptive optics scanning light ophthalmoscopy (AOSLO) images of the foveal center. The FCN was trained on a dataset of 49 healthy retinas and showed improved performance over previously published neural networks, particularly in the central fovea, achieving an  $F_1$  score of 0.9769 across the validation set, critically reducing analysis time. In addition to automatic cone detection, ConeMapper provides efficient manual annotation tools, visualizations and topographical analysis, offering users detailed metrics for further analysis. ConeMapper is freely available, with ongoing development aimed at enhancing functionality and adaptability to different retinal imaging modalities.

**Keywords** AOSLO, Neural networks, FCN, Confocal imaging

The foveola, the central 1-degree diameter of the human retina, is a sought-after study target for ophthalmology, vision science, and the neurosciences<sup>1–4</sup>. This is in part due to its key role for human vision, being the retinal area for sharp color vision<sup>5–7</sup>, and at the same time, the historical difficulty to study its delicate anatomy with both *ex-vivo*<sup>8</sup> and *in-vivo* approaches<sup>9,10</sup>. With the recent refinement of high-resolution retinal imaging techniques equipped with adaptive optics, researchers can directly observe the mosaic of the light-sensitive cone photoreceptors and their outer segments in the foveola of the living eye<sup>11–14</sup>, and study its functional role for human vision in health and disease<sup>15–18</sup>.

*In-vivo* imaging and topographical analysis of the foveal center remain challenging, even under optimal imaging conditions—such as minimized ocular aberrations, clear optical media and minimal eye movements. This is because the outer segments of foveolar cones are long and thin, thinner than anywhere else in the retina<sup>19,20</sup>. Their minimal diameter, typically less than 2  $\mu\text{m}$ , approaches the lateral resolution limit governed by the laws of diffraction of light focused in a human eye<sup>21</sup>. Combined with their orderly arrangement, with an almost crystalline packing geometry and a steep density gradient towards a central peak<sup>22,23</sup>, imaging with insufficient lateral resolution produces spurious patterns. Such patterns, similar in appearance to cellular structure, may mask the true underlying mosaic and hinders image interpretation<sup>24</sup>. Moreover, the outer segment's wave-guiding properties cause a natural variation in reflectivity in confocal images which can make individual cells disappear from the mosaic<sup>25–28</sup>.

In the past years, image analysis and automatic cell annotation techniques have been put forward to quantify rod and cone photoreceptor cells in confocal and non-confocal retinal images, as they are produced by modalities such as adaptive optics scanning light ophthalmoscopy (AOSLO). Such solutions span a variety of approaches, including image histogram analysis<sup>29</sup>, and multiscale modelling and normalized cross-correlation<sup>30,31</sup>. More recently, neural networks (NNs) have been developed to generalize cell identification from a manually labelled training data set<sup>32–36</sup>. While NNs have been shown to work reliably in non-foveal areas where cell density is

<sup>1</sup>Department of Ophthalmology, University Hospital Bonn, Bonn 53127, Germany. <sup>2</sup>b-it and Computer Science Department, University of Bonn, Bonn 53115, Germany. <sup>3</sup>Lamarr Institute for Machine Learning and Artificial Intelligence, <https://lamarr-institute.org/>. <sup>4</sup>Present address: Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg. <sup>5</sup>A. Gutnikov and P. Hähn-Schumacher: these authors contributed equally to this work. ✉email: wolf.harmening@ukbonn.de

relatively constant, they have not been trained nor applied for foveolar analysis. The gold standard for foveolar image annotation thus remains manual labeling by an expert image reader<sup>3,17,37</sup>. Manual image annotation is, however, extremely time-consuming, prone to error, and highly dependent of internal annotation criteria employed by the reader in the interpretation of images that are affected by the challenges described above<sup>38,39</sup>. While individual research labs have developed their own analysis strategies that can lead to difficulties when data is shared between sites, there is currently no open-source solution for researchers to efficiently annotate and analyze foveolar AOSLO imagery.

Here, we present an open-source software package that integrates an improved state-of-the-art neural network with an approachable user interface for fast, reliable automatic detection, efficient manual relabeling and basic topographical analysis of the human foveolar cone mosaic.

## Methods

A MATLAB application (“ConeMapper”) was developed that integrates a fully convolutional neural network (FCN) with additional annotation, visualization and analysis functionality. ConeMapper was developed in MATLAB 2022b App Designer. The FCN and post-processing steps were developed in Python with PyTorch library and are called as external Python scripts during runtime of the Matlab app. To gauge our FCN results, two published neural networks were trained and evaluated with identical data sets. For the Cunefare CNN, the MatConvNet 1.0-beta25 library was used.

### Data set

The dataset of confocal detector AOSLO images has been acquired for another study focusing on foveal cone topography<sup>40</sup>. It consisted of intensity image montages of 49 healthy retinas (ages, average: 25 years, range: 10–44 years), each covering an approximate 2-degree diameter of the retina centered on the foveola (see Figure 1, top left). The AOSLO system was described in more detail previously<sup>41,42</sup>. The digital resolution was 600 pixels per degree of visual angle. Imaging wavelengths were 711, 788, or 840 nm. Each montage was composed of 9 summed and normalized images which were computed from 10-second AOSLO video recordings. Individual video frames were stabilized offline using a refined strip-based registration technique based on an earlier published algorithm<sup>43</sup>, and then pixel-wise summed and normalized. This process eliminated uncorrelated detector noise. Individual summated images were automatically aligned using software previously described<sup>44</sup> and manually blended in Corel PhotoPaint (v21.0.0.593, Corel Corporation, Ottawa, Canada) to create a continuous image montage. In each montage, cone center locations were annotated using then available annotation tools and by manual correction.

For network training, the dataset was randomly split into a training set of 39 montages and a test set of 10 montages (80:20 split). Each montage was cropped into randomly chosen, overlapping 256 patches of size 256 × 256 pixels and processed with their corresponding labels. Before cropping, montage borders were eroded to counteract image artifacts such as noise and striping. We used a binary structuring element representing a von Neumann neighbourhood, which took only directly adjacent pixels into account. Due to the limited data availability, only 64 erosion iterations were performed on each montage, which eliminated most artifacts while keeping as much image content as possible (Fig. 1, top right). Erosion was also employed during recognition as a pre-processing step, though the number of iterations was limited to two. This is because each iteration removes part of the border information. During the training phase, it is crucial to retain as much clear data as possible, whereas in the recognition phase, we aim to utilize as much image data as we can. Two iterations of erosion were sufficient to eliminate the most significant artifacts at the borders.

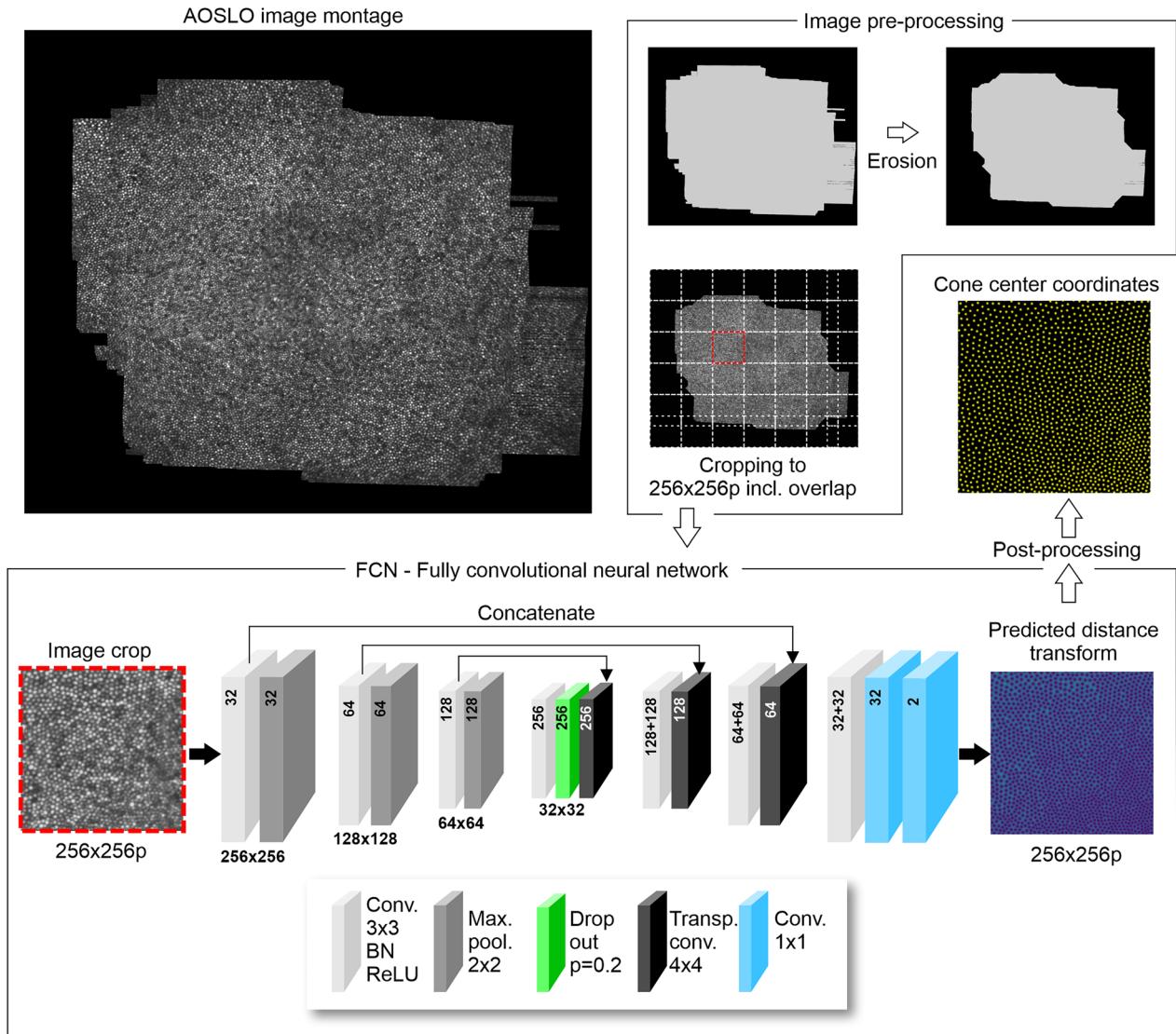
### Fully convolutional network

The network architecture represented a modified version of U-Net<sup>45</sup> and closely resembled a fully convolutional network proposed earlier<sup>34</sup>. It consisted of an encoder-decoder like structure with no fully connected layers after the decoder. A number of modifications to the earlier implementation were employed, such as different label representations, input sizes and number of filters, mainly to satisfy an increased network capacity (Figure 1). Our solution employed a single forward pass to process an image.

The encoder consisted of multiple blocks at a given resolution, which are connected by max pooling layers decreasing the resolution. Each block consisted of  $k$  sub-blocks, each containing a  $3 \times 3$  convolutional layer with zero-padding, a batch normalization layer and a ReLU layer. The number of filters is increased by a factor of 2 within each block. Best performances were obtained using  $k = 2$  sub-blocks at each resolution. As in the original implementation, we chose 3 decreases in resolution by pooling which was able to capture all cone sizes present in the images.

In contrast to the original U-Net, Hamwood et al.<sup>34</sup> inserted a dropout layer at the lowest resolution between the encoder part and decoder part of the neural network due to the need for regularization when training with a limited amount of data. However, their chosen dropout probability of  $p = 0.5$  led to worse results than a more lenient value of  $p = 0.2$  with our available data.

The decoder consisted of multiple blocks at a given resolution as well. The output of a previous resolution block is upscaled to doubled resolution using transposed convolution with a kernel size of  $4 \times 4$ , a stride of  $2 \times 2$  and zero-padding of  $1 \times 1$ , following the reference implementation of Hamwood et al.<sup>34</sup>. In addition to the upscaled feature maps, the output of the encoder block of the same resolution is concatenated using skip connections before being processed by the respective block. Skip connections have proved to be beneficial in deep neural networks when reconstructing after decreasing the resolution to not lose structural information<sup>46</sup>. Like the encoder blocks, the decoder blocks consist of  $k = 2$  sub-blocks, each with a  $3 \times 3$  convolutional layer with zero-padding, a batch normalization layer and a ReLU layer. However, the number of filters is decreased by the decoder block by a factor of 2.



**Figure 1.** Image data workflow during automatic cone detection. AOSLO image montages were eroded and cropped (top) to serve as image data for a fully convolutional network (FCN, bottom). From the resulting predicted distance transform, cone center locations were derived for each crop and stitched back into the full montage.

While U-Net has been originally designed for segmentation<sup>45</sup>, we modified it to perform regression due to our chosen label representation of a non-negative distance transform instead of class probabilities. We replace the last softmax layer of the neural network with a  $1 \times 1$  convolutional layer. Hence, we receive a single channel  $256 \times 256$  network output of a predicted distance transform.

### Post-processing

Finally, to localize cones from the FCN result, we implemented the same approach suggested by Cunefare et al.<sup>32</sup>, by thresholding and determining local maxima in the NN probability map. To further minimize possible manual work after automatic cone detection, a regularity-aware particle system (RAPS) was implemented. This attempts to take advantage of the observed regularity in the GT cone locations to shift cones into place. In the particle system, each cone holds internal and external energy that dictates its location relative to the surround. The whole energy of the system is:  $E = (1.0 - \alpha) \times E_{ext} + \frac{\alpha}{2} \times E_{int}$ . Here,  $\alpha$  is a control parameter. The external term  $E_{ext}$  imposes the environmental information given by the predicted distance transform  $D$  of the retinal image onto the particles:  $E_{ext} = \sum_{p \in P} D(p)$ .  $E_{int}$  are internal forces that impose the regularity of the cone mosaic onto the particles by modeling the influence  $I$  of neighboring particles. Each particle  $p \in P$  contributes to the internal term by its internal energy, which is the sum of all influences from neighboring particles  $n \in N(P)$ :

$E_{int} = \sum_{p \in P} \sum_{n \in N(P)} I(p, n)$ . Optimal cone locations were then found by minimizing the total energy, E, of the system.

### Validation

For validation, an additional data set of 12 image montages was employed that was not used during NN training/testing. The validation data set was obtained in the same way as described earlier. To quantify and compare the performance of our FCN, the automatically found cone locations were matched one-to-one to the ground truth (GT) coordinates which were found by manual labeling of an expert image reader. An automatically found cone was considered a true positive (TP) if it was located within some distance  $d$  of the GT coordinates. The value  $d$  was found empirically and set to 0.35 of the minimal distance between two cones across all the images. Automatically detected cones that were not matched to a GT cone were considered false positives (FP), and GT locations that did not have a matching automatically detected cone were considered false negatives (FN). If a GT cone was matched to more than one automatically marked cone, only the one with the smallest distance was considered a TP, and the remaining were considered FP. Finally, we were comparing the cones only in image boundaries formed by GT sets. Automatically detected cones outside of that boundary were excluded from analysis. From TP and FP, the true positive rate (TPR), false detection rate (FDR) and  $F_1$  score ( $F_1$ ) were computed by,

$$\begin{aligned} TPR &= TP / (TP + FN), \\ FDR &= FP / (TP + FP), \\ F_1 &= 2 * TP / (2 * TP + FP + FN). \end{aligned}$$

Because the seemingly subtle differences in such metrics may be sometimes hard to translate into real-world impact, we added an approximate estimation of correction time needed for a manual re-labelling step, i.e. how much time is required to fix the errors in the NN annotation by a human image reader. This approximation is based on the assumption that an expert reader requires 3 seconds to fix one cone position manually, either by adding a missing FN, or by removing a FP.

To evaluate the RAPS, an additional metric was used. The Chamfer Distance is defined as

$$ChamferDistance = \frac{1}{|C|} \sum_{c \in C} d_p(c)$$

where  $C$  denotes the set of all ground-truth cone locations,  $P$  is the set of predicted cone locations and  $d_p(c)$  represents the minimum distance between a ground-truth cone location and any predicted cone location.

### Cone density metrics

A commonly used quantitative metric for analyzing photoreceptor mosaics is cone density, defined as the number of cones per unit area. Three different ways of density calculation were implemented in ConeMapper:

1. Voronoi: A map of two-dimensional cone density was computed based on the combined area of the Voronoi cells of the nearest 150 cones to any given point in the map. Vicinity is based on Euclidean distance between points. This method was suggested by Reiniger et al.<sup>17</sup>.
2. ICD: Density is computed at each cone location from the individual average inter cone distance (ICD) and assuming perfect hexagonal packing, by  $density = 1/HexArea$ ;  $HexArea = ICD^2 \times \sqrt{3}/2$ . ICD for a given cone was computed as the average distance to all its neighbours. Neighbouring cones are all those that share a common edge with the given cone in the Voronoi diagram. For every point in the map which is not a cone location, density is interpolated by using MATLAB function *scatteredInterpolant* with nearest neighbour method.
3. Yellot's ring: Density is computed based on the spatial frequency content of the image, found by analysis of symmetrical patterns in the Fourier transformed image. This approach does not require annotation. We implemented an algorithm developed and described by Cooper et al.<sup>31</sup>.

One of the features of ConeMapper is a density z-score map, visualizing the difference of a current density map to the group average. For this, group average density maps based on the 49 full montage images were computed for left and right eye separately. The z-score value,  $z$ , at each map point is then calculated as,

$$z = (x - \mu) / \sigma$$

where  $x$  is the density value at given point,  $\mu$  is the average density across all eyes at given point, and  $\sigma$  is the density standard deviation across all eyes at given point.

Further, derived from such two-dimensional density maps, density profiles were computed along either horizontal or vertical meridians or as radial averages, in each case centered on the cone density centroid (CDC). The CDC was found in each density map as the weighted centroid of all density values within the contour of the 20% highest density values<sup>17</sup>. Consistently, retinal eccentricity was defined as distance from the CDC throughout all analysis.

## Data and software availability

Image datasets used for training, testing and validation are available at <https://doi.org/10.17632/r87cvk4mp8.1>. The complete software package of ConeMapper, including example data sets, documentation and roadmap, is available at <https://github.com/ukb-aoslo/ConeMapper>.

## Results

### Automatic foveolar cone detection

To evaluate automatic annotation performance of our software, key performance metrics were computed for a validation data set and compared to the performance of two published neural networks (Hamwood FCN<sup>34</sup>, Cunefare CNN<sup>32</sup>), and to a conventional fast peak find algorithm (FPF, [Fast 2d Peak Finder MATLAB File Exchange](#)) on the same set.

Across all montages from the validation data set, the two NNs and our solution show similar performance, with our solution performing best, on average (Table 1). As expected, the FPF algorithm does not work well for the challenging task of cone identification in the foveal center. The average  $F_1$  Scores across the validation set were 0.8723 for FPF, 0.9517 for Cunefare CNN, 0.9449 for Hamwood FCN, and 0.9769 for our solution. Individual metrics computed for each montage are presented in the Supplemental Information. In some cases, all NNs perform similarly, in others, differences can be quite drastic (compare supplemental Tables S1, S2, S3).

Because of the steep gradient in cone density towards the foveal center and the associated rapid change in structural appearance (Figure 2), we additionally compared the performance metrics for the bounding box within 0.3 deg radius from ground truth CDC location (Tables S4, S5, S6). Results show that our solution excels especially in this region, with mean  $F_1$  scores of 0.6799 for FPF, 0.9779 for Cunefare CNN, 0.9799 for Hamwood FCN, and 0.9915 for our solution. Inclusion of the regularity-aware particle system (RAPS) did improve cone localization relative to the FCN result slightly. The Chamfer distance was 0.602 without and 0.557 with RAPS. However, the RAPS had no impact on  $F_1$  scores, because it did not add or remove cones.

On our test-system running Windows 11 and Matlab R2022b with a Intel Core i5 11600K CPU and NVIDIA RTX3070 graphics card with 8Gb of VRAM, average recognition time of our FCN was 141 seconds per montage without the RAPS and 177 seconds with RAPS.

When the performance metrics were translated to the real-world problem of manual re-labelling after automatic detection, it became clear that our solution outperformed the current best solution by more than 2.5-fold. The median estimated correction time of our FCN was 45 minutes, presenting a significant advantage over median correction times of 141 minutes for Hamwood FCN, 117 minutes for Cunefare CNN, and 350 minutes for FPF (Table S7).

### Cone density maps and profiles

The uncorrected two-dimensional density maps and density radial averages were also compared to GT data. The comparison in Figures 3, 5 confirms that our solution outperforms the current solutions, especially in the foveal center. We also compared density maps resulting from the spatial frequency image analysis<sup>31</sup> to GT, because this method is independent of annotation quality. Density difference maps (Figure 4) better highlight the typical problematic regions for each approach. Overall, the typical problematic regions were the foveal center (< 0.2 deg radii from CDC), because of light interference artifacts, and above 1 deg radii from CDC, due to border artifacts and the increased occurrence of rod photoreceptors that are similar in appearance as the smallest cones.

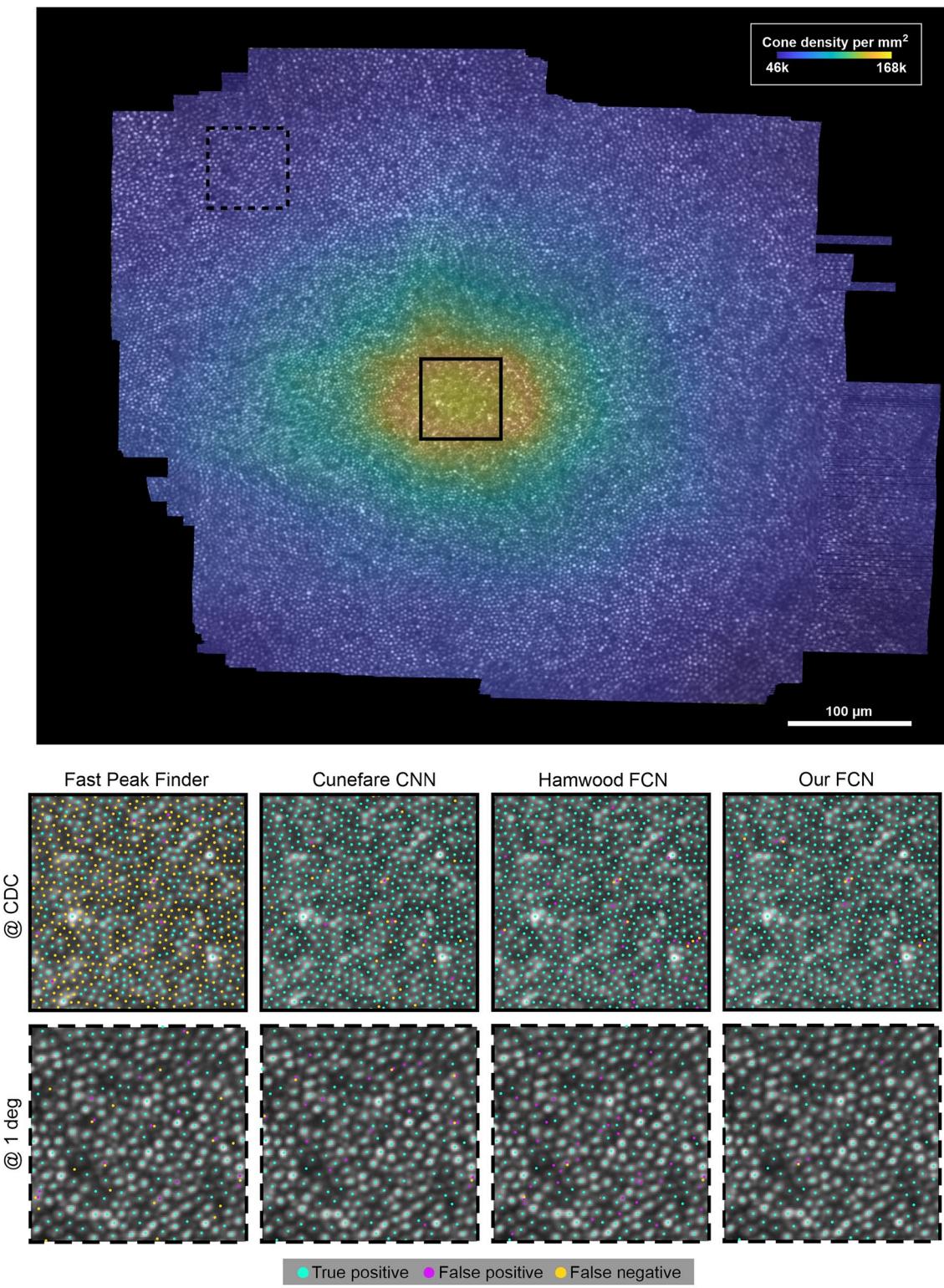
The density radial profiles in Figure 5 show the deviation from GT of each method. It is clearly visible that most of the methods struggled close to the CDC and at radii>1 degree away from CDC. Our FCN performs better at the CDC in most eyes (but compare data for BAK1093R). Interestingly, the Yello's Ring approach systematically underestimated density in all cases.

### ConeMapper annotation, visualization and analysis features

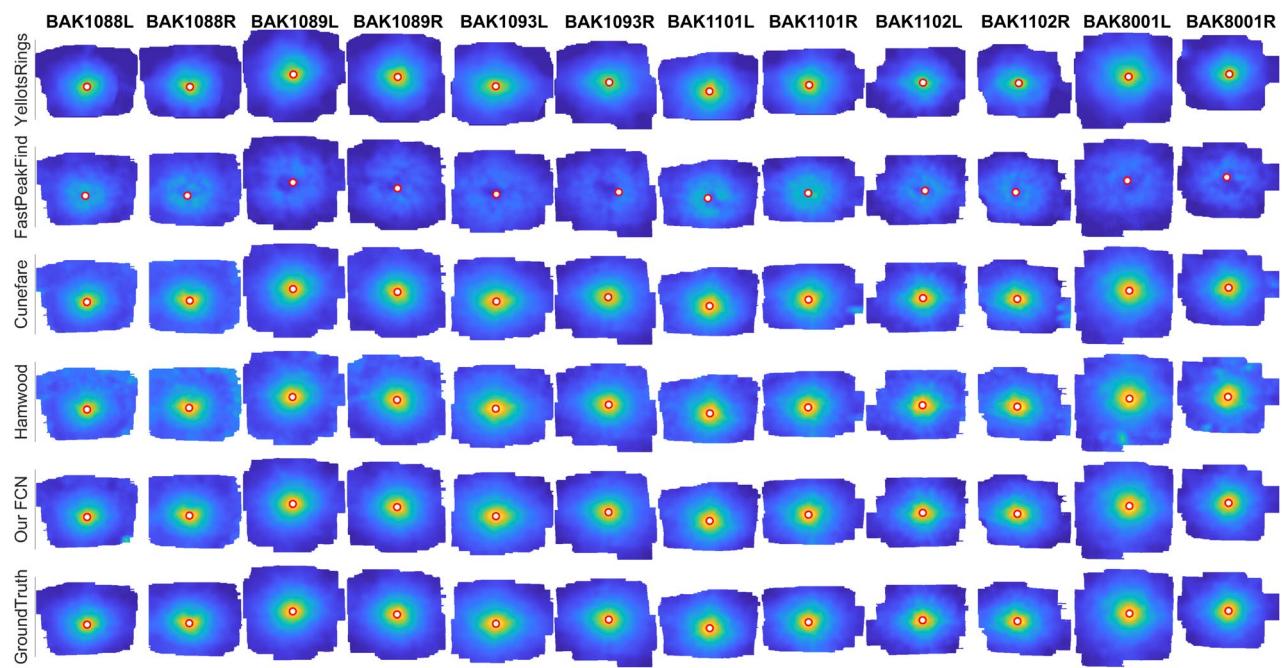
As we have seen, automatic cone detection was improved, but still leaves us with undetected or falsely labeled cones. On average, this amounted to about 900 errors in a 2-degree foveal montage, and to about 40 errors in a central zone (0.3 deg radii around CDC). In a newly acquired image, ground truth data is not available, of course. We have thus equipped the ConeMapper software with tools that make the manual correction process as easy and efficient as possible. For this, an intuitive annotation interface, several visualization options and basic

	FastPeakFinder	Cunefare	Hamwood	Our FCN
Avg TPR	0.9531	0.9484	0.9197	0.9778
Avg FDR	0.1946	0.0447	0.0278	0.0238
Avg $F_1$	0.8723	0.9517	0.9449	0.9769
Avg TPR (0.3 deg)	0.9597	0.9848	0.9681	0.9931
Avg FDR (0.3 deg)	0.4681	0.0289	0.0080	0.0101
Avg $F_1$ (0.3 deg)	0.6799	0.9779	0.9799	0.9915
Median Correction Time (min)	350	117	141	45

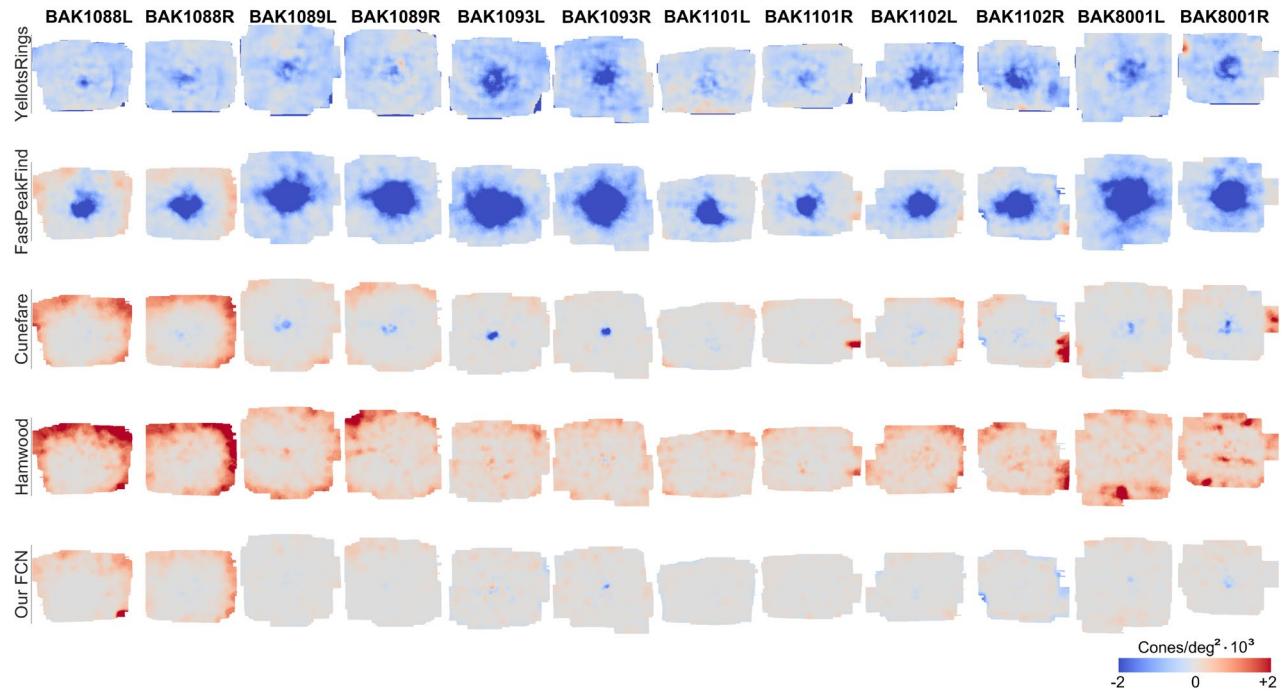
**Table 1.** Comparative performance metrics for automatic foveolar cone detection. TPR: true positive rate, FDR: false detection rate,  $F_1$ :  $F_1$  score. Except for FDR and correction time, higher values indicate better performance.



**Figure 2.** Automatic cone detection quality. Top: GT density map of eye BAK1011R. Density is indicated by color. The squares show the regions presented at higher magnification in the bottom panel, at the CDC (solid line), and at 1-deg eccentricity (dashed line). Box size is 64  $\mu\text{m}$ . Bottom: Detection results from the Fast peak find algorithm, Cunefare CNN, Hamwood FCN, and our FCN. TP, FP and FN are indicated by colors. All neural networks were trained on the same data set.



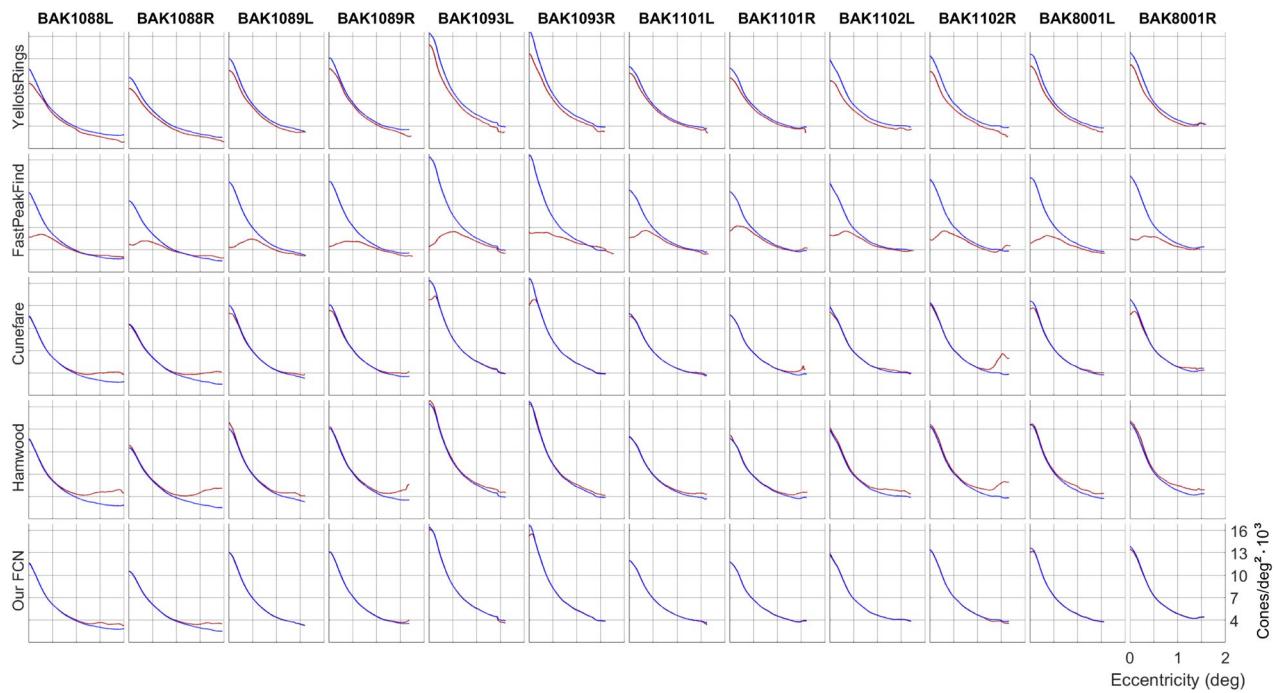
**Figure 3.** Density maps of the validation data set. The red-white marker in the center represents the CDC. Each column represents one montage, rows are the different recognition algorithms results. The bottom row is GT. Colors are cone density, normalized for GT density in each eye.



**Figure 4.** Density difference from GT of the validation set. Red color indicates overestimation, blue is underestimation. No difference is gray.

analysis features were developed. We briefly describe the main features here. For complete documentation and more examples, the reader is referred to the help section of our app.

The ConeMapper interface is made up of individual windows for maximum flexibility in window placement to accommodate different working and display environments. The main interface window from which most



**Figure 5.** Radially averaged density profiles of the validation set. Red lines are the profile based on the given recognition algorithm. The blue lines are GT profiles.

functions are controlled is narrow so it can be placed left or right of the image window. The image window can be freely resized, zoomed and panned with intuitive mouse controls. Digital and image scaling is displayed for repeatable settings. In general, the most important display settings are stored and carried over to future sessions (Figure 6).

To activate cone annotation, the ALT-key has to be held while inputs (deletions and additions) are given with mouse clicks directly in the image window (Figure 7). Several visualizations can be employed to aid this process. What we have found to be most useful are: a grid overlay to keep track of the annotation progress, an unfilled Voronoi representation of the current cone locations which updates in real-time. For intermittent checks, a filled Voronoi cell colormap based on number of neighbors, and a z-score difference map are helpful to detect problematic areas that need re-labelling. Voronoi maps are re-generated with every change in the annotation, while the z-score map is recomputed only after computation of the associated density map. Density and density z-score maps are computed as described in the Methods section which requires a few seconds, depending on image size.

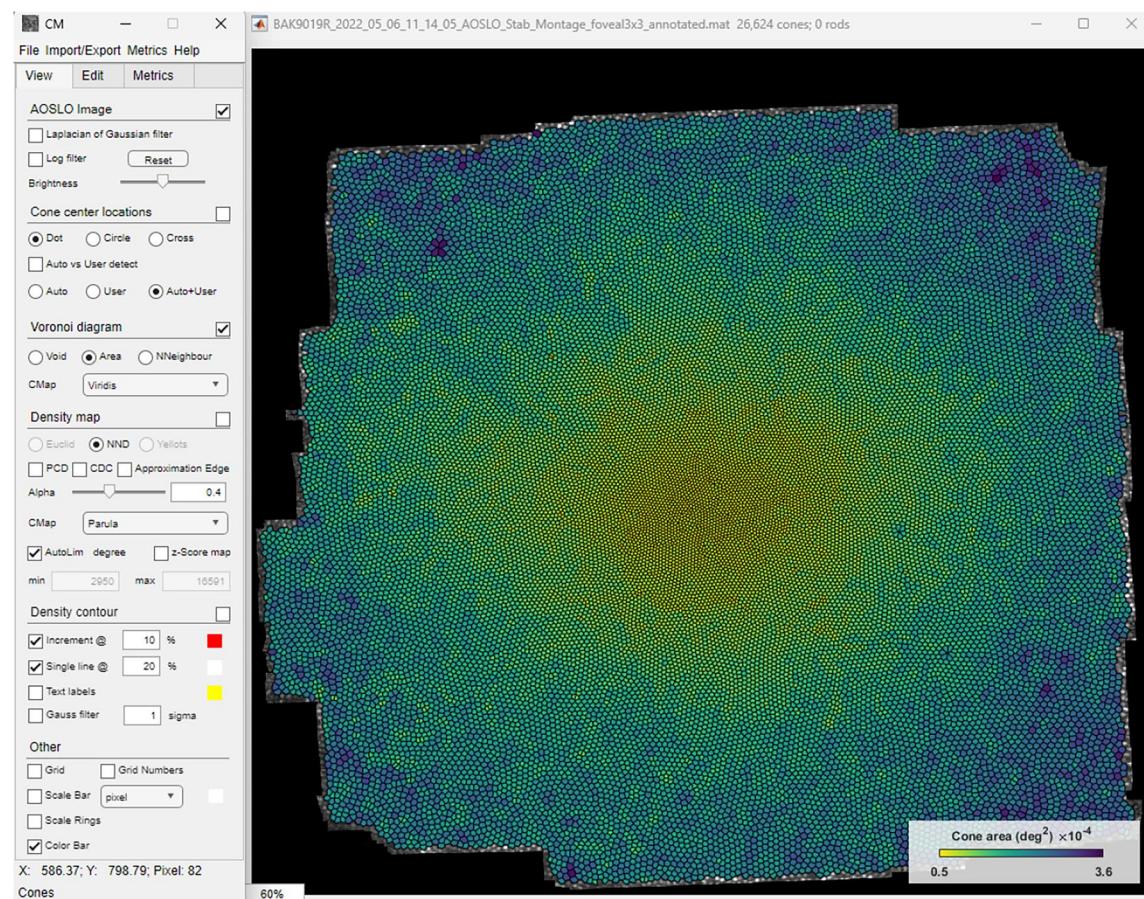
Most of the visualizations were implemented as layers, so they can be shown on top of each other at the same time, or in isolation (Figure 8). Other useful visualizations are markers for peak cone density (PCD) and the cone density centroid (CDC), the latter defining 0,0 eccentricity. Density contours at selectable steps can be shown to better gauge overall topography. Voronoi diagrams can be chosen to be unfilled, or filled by colormaps representing number of neighbours or area of the cell. Scale rings allows to precisely determine the distance from the CDC.

Finally, an analysis window can be summoned to display several quantities such as key density values (CDC, PCD) and their locations, and a radial average density plot which displays the current profile in comparison to the group average. (Figure 9). Group average data can be exchanged to arbitrary datasets in the code. The user can switch between a selection of image scales and units (e.g. mm or minutes of arc) for all displayable content. For such conversions, users have to input the image magnification factor (pixel per degree of visual angle) and retinal magnification factor (micron per degree of visual angle). Such values are stored together with annotation and density maps into a single .mat file upon saving (or ending) the session.

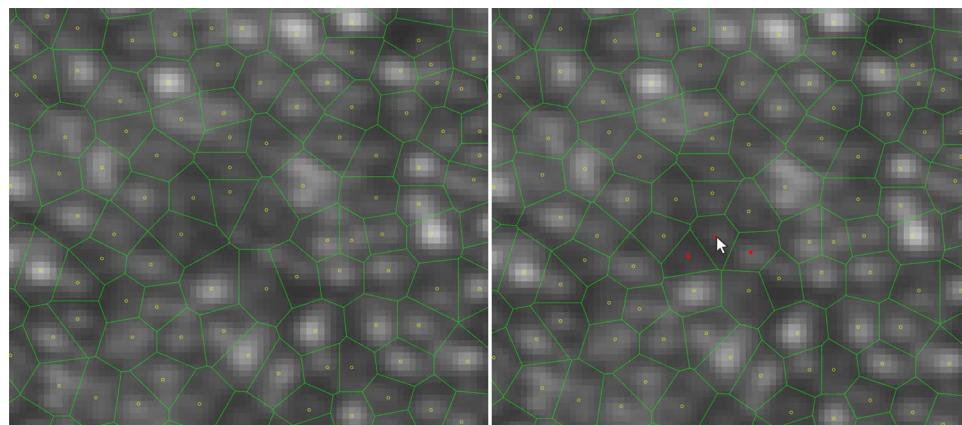
## Discussion

We developed a software package that delivers good performance for automatic detection and manual annotation of foveolar cones and topographical analysis of en-face foveolar images. It is free to use, open source, user friendly and adjustable. Our cone recognition FCN showed improved performance when compared to published solutions, especially in the most central part of the foveola. ConeMapper allows easy manual error correction and gives instruments to immediately assess the result. Moreover, it allows extraction of basic topographical information useful for future analysis such as density profiles, two-dimensional density maps, and topographical landmarks.

Foveolar *in-vivo* image interpretation to quantify cell topography is difficult because of the blur ceiling governed by the imaging wavelength and the aberrations and numerical aperture of the eye<sup>21</sup>. Second, cell

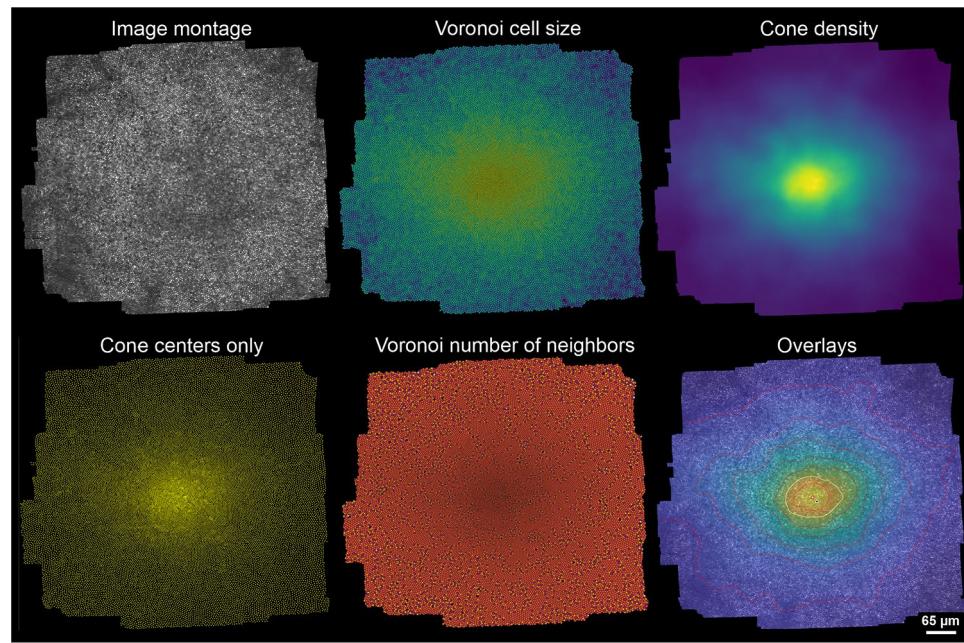


**Figure 6.** ConeMapper main interface with an example fully annotated AOSLO montage and Voronoi diagram indicating cell area.

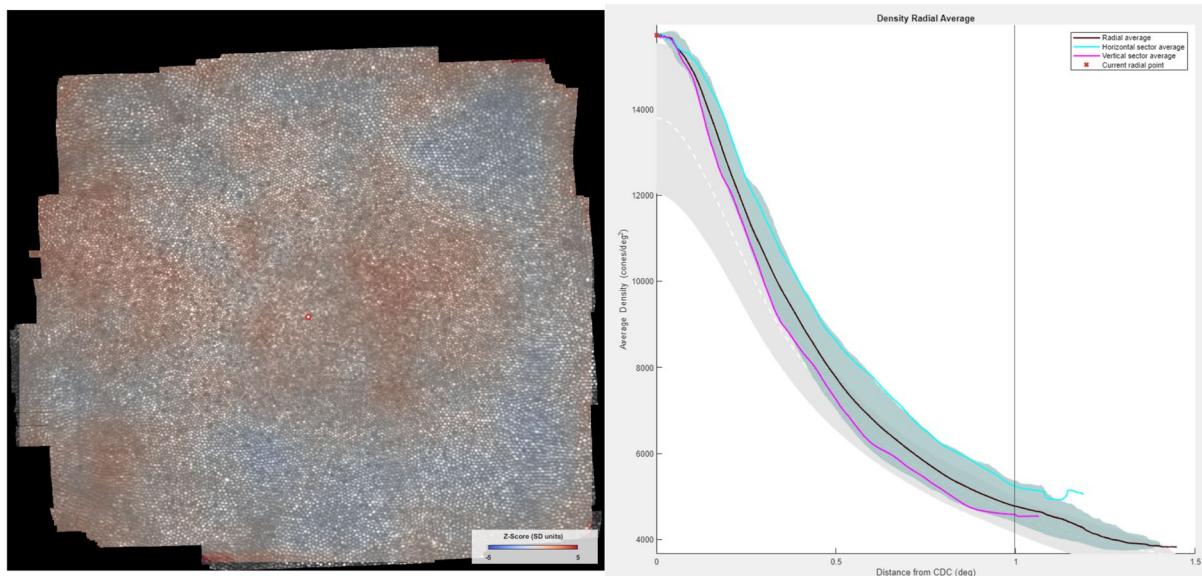


**Figure 7.** Manual annotation. The red dots in the center were manually added. In Cone Mapper by Alt+ Left Mouse Button cone is placed under the mouse position, by Alt + Right Mouse Button cone under the mouse is removed.

appearance in confocal AOSLO images is not straightforward, a consequence of the optical properties of the cells outer segments and their interaction with light<sup>25–28</sup>. Third, the sheer number of cells make manual annotation infeasible. Neural networks seem ideal to tackle cell detection in this instance, but require to be trained to the specific image features present in foveolar images. For example, an attempt to detect cones in our dataset with a NN that has been trained with image material from a 1-deg eccentricity region (Cunefare original data set<sup>32</sup>), was not able to generalize its excellent detection performance to the foveal center (Average  $F_1$  score derived



**Figure 8.** Example visualization options of ConeMapper. Visualization types are given in panel titles, colors indicate the entities given. Some or all of the individual options can be combined in layered overlays.



**Figure 9.** Z-Score map and Density Radial Average plot for image at figure 6. Gray zone on the background represents the mean  $\pm 1$  SD of radial density distribution.

from our data set: 0.8452), and was en par with the inferior Fast Peak Find algorithm. Fully-automated pipelines that can handle large image eccentricities exist but lack high-confidence cone detection in the foveal center<sup>47</sup>.

Our improved FCN reduced manual correction times more than 2.5-fold compared to current solutions. Adding the regularity-aware particle system resulted in minor improvements to cone center localization which were outweighed by additional computation time. We thus decided to deactivate the particle system by default (it can be activated in the source code). In our experience, the remaining manual annotation error correction still takes time, but is no longer a major bottleneck in the image analysis pipeline. Other manual steps like image quality assessment, image selection and montage blending require a similar amount or more time than cone annotation.

It is important to recapitulate that successful *in-vivo* foveolar cell analysis can only be as good as the image material used. One of the key requirements for good image quality is sufficient lateral resolution. Our solution

is thus tailored to work with confocal AOSLO foveal images. It has not been tested with split-detector images because this modality does not provide the resolution to resolve the small cones of the foveola<sup>48</sup>. Similarly, flood-illumination adaptive optics imaging does not provide the necessary lateral resolution<sup>36</sup>. Adaptive optics ocular coherence tomography (AO-OCT) has been shown to resolve foveolar cones and rods in select eyes<sup>49</sup>, but a larger image dataset is not available yet.

For future improvements to ConeMapper, we publish a [roadmap](#) of planned additions and changes alongside our software on github. These include, for example, batch processing, better figure export and ROI analysis tools. We decided to make our software and the development process open access so that there is potential for other groups to freely use and add to it. For example, ConeMapper allows choice of different NNs for automatic cell detection, which could be useful for images from outside the foveal center. For other imaging modalities such as split detection, offset pinhole, or AO-OCT imaging, specialized NNs can be added or the existing ones re-trained. In conclusion, we hope that our software package will allow researchers to speed up foveolar image analysis by making the annotation process more reliable, comfortable and comparable.

Received: 22 November 2024; Accepted: 18 June 2025

Published online: 04 July 2025

## References

- Poletti, M., Rucci, M. & Carrasco, M. Selective attention within the foveola. *Nat. Neurosci.* **20**, 1413–1417. <https://doi.org/10.1038/nn.4622> (2017).
- Sinha, R. et al. Cellular and circuit mechanisms shaping the perceptual properties of the primate fovea. *Cell* **168**, 413–426.e12. <https://doi.org/10.1016/j.cell.2017.01.005> (2017).
- Wang, Y. et al. Human foveal cone photoreceptor topography and its dependence on eye length. *Elife* **8**, e47148. <https://doi.org/10.7554/elife.47148> (2019).
- Bucci, A. et al. Synchronization of visual perception within the human fovea. *Nat. Neurosci.* **43**, 127–143 (2025).
- Yuodelis, C. & Hendrickson, A. A qualitative and quantitative analysis of the human fovea during development. *Vision. Res.* **26**, 847–855. [https://doi.org/10.1016/0042-6989\(86\)90143-4](https://doi.org/10.1016/0042-6989(86)90143-4) (1986).
- Cohen, M. A., Botch, T. L. & Robertson, C. E. The limits of color awareness during active, real-world vision. *Proc. Natl. Acad. Sci.* **117**, 13821–13827. <https://doi.org/10.1073/pnas.1922294117> (2020).
- Tuten, W. S. & Harmening, W. M. Foveal vision. *Curr. Biol.* **31**, R701–R703. <https://doi.org/10.1016/j.cub.2021.03.097> (2021).
- Curcio, C. A., Packer, O. & Kalina, R. E. A whole mount method for sequential analysis of photoreceptor and ganglion cell topography in a single retina. *Vision. Res.* **27**, 9–15. [https://doi.org/10.1016/0042-6989\(87\)90137-4](https://doi.org/10.1016/0042-6989(87)90137-4) (1987).
- Roorda, A. & Duncan, J. L. Adaptive optics ophthalmoscopy. *Annual Review of Vision Science* **1**, 19–50. <https://doi.org/10.1146/annurev-vision-082114-035357> (2015).
- Burns, S. A., Elsner, A. E., Sapoznik, K. A., Warner, R. L. & Gast, T. J. Adaptive optics imaging of the human retina. *Prog. Retin. Eye Res.* **68**, 1–30. <https://doi.org/10.1016/j.preteyeres.2018.08.002> (2019).
- Dubra, A. & Sulai, Y. Reflective afocal broadband adaptive optics scanning ophthalmoscope. *Biomed. Opt. Express* **2**, 1757–1768. <https://doi.org/10.1364/BOE.2.001757> (2011).
- Ameln, J., Reiniger, J. L., Hess, K., Holz, F. G. & Harmening, W. M. Supernormal foveal photoreceptor density in alport syndrome: A case report. *Eur. J. Ophthalmol.* **33**, NP51–NP54. <https://doi.org/10.1177/11206721221093197> (2022).
- Kreis, J. & Carroll, J. Applications of adaptive optics imaging for studying conditions affecting the fovea. *Annual Rev. Vis. Sci.* **10**, 239–262. <https://doi.org/10.1146/annurev-vision-102122-100022> (2024).
- Moon, B. et al. Alignment, calibration, and validation of an adaptive optics scanning laser ophthalmoscope for high-resolution human foveal imaging. *Appl. Opt.* **63**, 730–742. <https://doi.org/10.1364/AO.504283> (2024).
- Roorda, A. Adaptive optics for studying visual function: A comprehensive review. *J. Vis.* **11**, 6. <https://doi.org/10.1167/11.5.6> (2011).
- Tuten, W. S. et al. Spatial summation in the human fovea: Do normal optical aberrations and fixational eye movements have an effect?. *J. Vis.* **18**, 6. <https://doi.org/10.1167/18.8.6> (2018).
- Reiniger, J. L., Domdei, N., Holz, F. G. & Harmening, W. M. Human gaze is systematically offset from the center of cone topography. *Curr. Biol.* **31**, 4188–4193.e3. <https://doi.org/10.1016/j.cub.2021.07.005> (2021).
- Witten, J. L., Lukyanova, V. & Harmening, W. M. Sub-cone visual resolution by active, adaptive sampling in the human foveola. *Elife* **13**, RP98648. <https://doi.org/10.7554/elife.98648> (2024).
- Yamada, E. Some Structural Features of the Fovea Centralis in the Human Retina. *Arch. Ophthalmol.* **82**, 151–159. <https://doi.org/10.1001/archophth.1969.00990020153002> (1969).
- Domdei, N. et al. Cone density is correlated to outer segment length and retinal thickness in the human foveola. *Invest. Ophthalmol. Vis. Sci.* **64**, 11. <https://doi.org/10.1167/iosv.64.15.11> (2023).
- Campbell, F. W. & Green, D. G. Optical and retinal factors affecting visual resolution. *J. Physiol.* **181**, 576–593. <https://doi.org/10.1113/jphysiol.1965.sp007784> (1965).
- Curcio, C. A. & Sloan, K. R. Packing geometry of human cone photoreceptors: Variation with eccentricity and evidence for local anisotropy. *Vis. Neurosci.* **9**, 169–180. <https://doi.org/10.1017/S0952523800009639> (1992).
- Pum, D., Ahnelt, P. K. & Grasl, M. Iso-orientation areas in the foveal cone mosaic. *Vis. Neurosci.* **5**, 511–523. <https://doi.org/10.1017/S095252380000687> (1990).
- Putnam, N. M., Hammer, D. X., Zhang, Y., Merino, D. & Roorda, A. Modeling the foveal cone mosaic imaged with adaptive optics scanning laser ophthalmoscopy. *Opt. Express* **18**, 24902–24916. <https://doi.org/10.1364/OE.18.024902> (2010).
- Pallikaris, A., Williams, D. R. & Hofer, H. The reflectance of single cones in the living human eye. *Invest. Ophthalmol. Vis. Sci.* **44**, 4580–4592. <https://doi.org/10.1167/iosv.03-0094> (2003).
- Bruce, K. S. et al. Normal perceptual sensitivity arising from weakly reflective cone photoreceptors. *Invest. Ophthalmol. Vis. Sci.* **56**, 4431–4438. <https://doi.org/10.1167/iosv.15-16547> (2015).
- Meadway, A. & Sincich, L. C. Light reflectivity and interference in cone photoreceptors. *Biomed. Opt. Express* **10**, 6531–6554. <https://doi.org/10.1364/BOE.10.006531> (2019).
- Bensinger, E., Wang, Y. & Roorda, A. Patches of dysflective cones in eyes with no known disease. *Invest. Ophthalmol. Vis. Sci.* **63**, 29–29. <https://doi.org/10.1167/iosv.63.1.29> (2022).
- 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. <https://doi.org/10.1364/JOSAA.24.001364> (2007).
- Turpin, A., Morrow, P., Scotney, B., Anderson, R. & Wolsey, C. Automated identification of photoreceptor cones using multi-scale modelling and normalized cross-correlation. In *Image Analysis and Processing - ICIAP 2011*, 494–503 (Springer (eds Maino, G. & Foresti, G. L.) (Berlin Heidelberg, Berlin, Heidelberg, 2011).

31. Cooper, R. F., Aguirre, G. K. & Morgan, J. I. W. Fully automated estimation of spacing and density for retinal mosaics. *Translational Vision Science & Technology* **8**, 26. <https://doi.org/10.1167/tvst.8.5.26> (2019).
32. Cunefare, D. et al. Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. *Sci. Rep.* **7**, 6620. <https://doi.org/10.1038/s41598-017-07103-0> (2017).
33. Davidson, B. et al. Automatic cone photoreceptor localisation in healthy and stargardt afflicted retinas using deep learning. *Sci. Rep.* **8**, 7911. <https://doi.org/10.1038/s41598-018-26350-3> (2018).
34. Hamwood, J., Alonso-Caneiro, D., Sampson, D. M., Collins, M. J. & Chen, F. K. Automatic detection of cone photoreceptors with fully convolutional networks. *Translational Vision Science & Technology* **8**, 10. <https://doi.org/10.1167/tvst.8.6.10> (2019).
35. Li, K., Yin, Q., Ren, J., Song, H. & Zhang, J. Automatic quantification of cone photoreceptors in adaptive optics scanning light ophthalmoscope images using multi-task learning. *Biomed. Opt. Express* **13**, 5187–5201. <https://doi.org/10.1364/BOE.471426> (2022).
36. Wooning, S. et al. Automated cone photoreceptor detection in adaptive optics flood illumination ophthalmoscopy. *Ophthalmology Science* **5**, 100675. <https://doi.org/10.1016/j.xops.2024.100675> (2025).
37. Cava, J. A. et al. Assessing interocular symmetry of the foveal cone mosaic. *Investigative Ophthalmology & Visual Science* **61**, 23. <https://doi.org/10.1167/iovs.61.14.23> (2020).
38. Morgan, J. I. W., Chen, M., Huang, A. M., Jiang, Y. Y. & Cooper, R. F. Cone identification in choroideremia: Repeatability, reliability, and automation through use of a convolutional neural network. *Translational Vision Science & Technology* **9**, 40. <https://doi.org/10.1167/tvst.9.2.40> (2020).
39. Wynne, N. et al. Intergrader agreement of foveal cone topography measured using adaptive optics scanning light ophthalmoscopy. *Biomed. Opt. Express* **13**, 4445–4454. <https://doi.org/10.1364/BOE.460821> (2022).
40. Ameln, J. et al. In-vivo cone photoreceptor topography of the normal human foveola. *Investigative Ophthalmology & Visual Science* **64**, 1040 (2023).
41. Domdei, N. et al. Ultra-high contrast retinal display system for single photoreceptor psychophysics. *Biomed. Opt. Express* **9**, 157–172. <https://doi.org/10.1364/BOE.9.000157> (2018).
42. Domdei, N., Linden, M., Reiniger, J. L., Holz, F. G. & Harmening, W. M. Eye tracking-based estimation and compensation of chromatic offsets for multi-wavelength retinal microstimulation with foveal cone precision. *Biomed. Opt. Express* **10**, 4126–4141. <https://doi.org/10.1364/BOE.10.004126> (2019).
43. Stevenson, S. B., Roorda, A. & Kumar, G. Eye tracking with the adaptive optics scanning laser ophthalmoscope. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, 195–198, <https://doi.org/10.1145/1743666.1743714> (Association for Computing Machinery, New York, NY, USA, 2010).
44. Chen, M. et al. Multi-modal automatic montaging of adaptive optics retinal images. *Biomed. Opt. Express* **7**, 4899–4918. <https://doi.org/10.1364/BOE.7.004899> (2016).
45. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer International Publishing, Cham, 2015).
46. K. He, X. Zhang, S. Ren & J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, <https://doi.org/10.1109/CVPR.2016.90> (2016).
47. Cooper, R. F., Kalaparambath, S., Aguirre, G. K. & Morgan, J. I. W. Morphology of the normative human cone photoreceptor mosaic and a publicly available adaptive optics montage repository. *Sci. Rep.* **14**, 23166. <https://doi.org/10.1038/s41598-024-7427-4-y> (2024).
48. Scoles, D. et al. In vivo imaging of human cone photoreceptor inner segments. *Investigative Ophthalmology & Visual Science* **55**, 4244–4251. <https://doi.org/10.1167/iovs.14-14542> (2014).
49. Felberer, F. et al. Adaptive optics slo/ocf for 3d imaging of human photoreceptors in vivo. *Biomed. Opt. Express* **5**, 439–456. <https://doi.org/10.1364/BOE.5.000439> (2014).

## Author contributions

Aleksandr Gutnikov developed the ConeMapper software package, performed network training and analysis and wrote the main manuscript draft. Patrick Hähn-Schumacher developed and trained the Fully-Convolutional-Network as part of his Master thesis which was included in the software package. Julius Ameln recorded and annotated all retinal images. Shekoufeh Gorgi Zadeh and Thomas Schultz acquired funding, provided guidance and expertise in image processing and convolutional neural networks. Wolf Harmening acquired funding, provided guidance and expertise in AOSLO imaging and image interpretation, tested the software package, edited the manuscript, and oversaw the project. All authors reviewed the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL. Open Access funding enabled and organized by Projekt DEAL. Funded by the German Research Foundation (funding code 399370883), and by the Federal Ministry of Education and Research within the project “BNTrAinee” (funding code 16DHBK1022). This work was supported by the Open Access Publication Fund of the University of Bonn.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-08028-9>.

**Correspondence** and requests for materials should be addressed to W.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

# Neural network assisted annotation and analysis tool to study *in-vivo* foveolar cone photoreceptor topography

Aleksandr Gutnikov<sup>1,+</sup>, Patrick Hähn-Schumacher<sup>2,+</sup>, Julius Ameln<sup>1</sup>, Shekoufeh Gorgi Zadeh<sup>2,†</sup>, Thomas Schultz<sup>2,3</sup>, and Wolf Harmening<sup>1,\*</sup>

<sup>1</sup>Department of Ophthalmology, University Hospital Bonn, 53127, Germany

<sup>2</sup>b-it and Computer Science Department, University of Bonn, 53115, Germany

<sup>3</sup>Lamarr Institute for Machine Learning and Artificial Intelligence, <https://lamarr-institute.org/>

\*Corresponding author: [wolf.harmening@ukbonn.de](mailto:wolf.harmening@ukbonn.de)

<sup>+</sup>these authors contributed equally to this work

<sup>†</sup>current affiliation: Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg

## Supplementary information

**Table S1.** Individual true positive rates (TPR) for all montages in the validation set. Higher values indicate better performance.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	0.9085	0.8679	0.8294	0.9226
BAK1088R	0.9215	0.8897	0.8487	0.9370
BAK1089L	0.9582	0.9575	0.9250	0.9877
BAK1089R	0.9643	0.9585	0.9283	0.9886
BAK1093L	0.9676	0.9767	0.9559	0.9833
BAK1093R	0.9513	0.9769	0.9568	0.9783
BAK1101L	0.9667	0.9711	0.9552	0.9922
BAK1101R	0.9734	0.9711	0.9625	0.9919
BAK1102L	0.9632	0.9601	0.9262	0.9903
BAK1102R	0.9562	0.9374	0.9168	0.9840
BAK8001L	0.9546	0.9657	0.9237	0.9889
BAK8001R	0.9515	0.9480	0.9082	0.9884
Average	0.9531	0.9484	0.9197	0.9778
STD	0.0192	0.0348	0.0418	0.0229

**Table S2.** Individual false detection rate (FDR) for all montages in the validation set. Lower values indicate better performance.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	0.1590	0.0674	0.0522	0.0447
BAK1088R	0.1784	0.0588	0.0442	0.0372
BAK1089L	0.2283	0.0458	0.0261	0.0201
BAK1089R	0.2168	0.0400	0.0222	0.0160
BAK1093L	0.2397	0.0318	0.0196	0.0234
BAK1093R	0.2438	0.0335	0.0233	0.0297
BAK1101L	0.1448	0.0354	0.0215	0.0185
BAK1101R	0.1109	0.0308	0.0193	0.0162
BAK1102L	0.1554	0.0417	0.0249	0.0205
BAK1102R	0.1797	0.0478	0.0297	0.0220
BAK8001L	0.2414	0.0449	0.0243	0.0186
BAK8001R	0.2367	0.0583	0.0265	0.0189
Average	0.1946	0.0447	0.0278	0.0238
STD	0.0456	0.0117	0.0101	0.0089

**Table S3.** Individual  $F_1$  score for all montages in the validation set. Higher values indicate better performance.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	0.8734	0.8991	0.8847	0.9387
BAK1088R	0.8687	0.9147	0.8991	0.9497
BAK1089L	0.8549	0.9559	0.9488	0.9838
BAK1089R	0.8644	0.9592	0.9524	0.9863
BAK1093L	0.8515	0.9724	0.9680	0.9799
BAK1093R	0.8426	0.9717	0.9666	0.9743
BAK1101L	0.9075	0.9679	0.9667	0.9868
BAK1101R	0.9293	0.9702	0.9715	0.9878
BAK1102L	0.9001	0.9592	0.9500	0.9849
BAK1102R	0.8830	0.9448	0.9428	0.9810
BAK8001L	0.8454	0.9604	0.9490	0.9852
BAK8001R	0.8471	0.9448	0.9397	0.9848
Average	0.8723	0.9517	0.9449	0.9769
STD	0.0277	0.0231	0.0271	0.0159

**Table S4.** Individual true positive rates (TPR) for all montages in the validation set. Radius of the area is 0.3 deg with the center in ground truth CDC point. Higher values indicate better performance.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	0.9820	0.9925	0.9856	0.9991
BAK1088R	0.9750	0.9879	0.9726	0.9951
BAK1089L	0.9483	0.9817	0.9666	0.9956
BAK1089R	0.9572	0.9816	0.9708	0.9971
BAK1093L	0.9467	0.9798	0.9683	0.9791
BAK1093R	0.9352	0.9720	0.9675	0.9783
BAK1101L	0.9638	0.9926	0.9846	0.9975
BAK1101R	0.9878	0.9950	0.9845	0.9982
BAK1102L	0.9660	0.9893	0.9600	0.9962
BAK1102R	0.9743	0.9870	0.9536	0.9958
BAK8001L	0.9495	0.9898	0.9700	0.9977
BAK8001R	0.9313	0.9688	0.9326	0.9880
Average	0.9597	0.9848	0.9681	0.9931
STD	0.0181	0.0083	0.0148	0.0073

**Table S5.** Individual false detection rate (FDR) for all montages in the validation set. Radius of the area is 0.3 deg with the center in ground truth CDC point. Lower values indicate better performance.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	0.3390	0.0075	0.0016	0.0016
BAK1088R	0.4302	0.0224	0.0081	0.0078
BAK1089L	0.5466	0.0380	0.0079	0.0084
BAK1089R	0.5595	0.0365	0.0072	0.0059
BAK1093L	0.5858	0.0440	0.0128	0.0184
BAK1093R	0.5873	0.0524	0.0180	0.0297
BAK1101L	0.3987	0.0110	0.0040	0.0040
BAK1101R	0.2661	0.0029	0.0015	0.0018
BAK1102L	0.4115	0.0202	0.0057	0.0082
BAK1102R	0.4280	0.0188	0.0045	0.0064
BAK8001L	0.5182	0.0290	0.0069	0.0059
BAK8001R	0.5467	0.0639	0.0176	0.0229
Average	0.4681	0.0289	0.0080	0.0101
STD	0.1044	0.0186	0.0055	0.0088

**Table S6.** Individual  $F_1$  scores for all montages in the validation set. Radius of the area is 0.3 deg with the center in ground truth CDC point. Higher values indicate better performance.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	0.7901	0.9925	0.9920	0.9988
BAK1088R	0.7192	0.9827	0.9822	0.9937
BAK1089L	0.6135	0.9717	0.9792	0.9936
BAK1089R	0.6033	0.9724	0.9817	0.9956
BAK1093L	0.5762	0.9677	0.9777	0.9804
BAK1093R	0.5727	0.9597	0.9747	0.9743
BAK1101L	0.7406	0.9908	0.9903	0.9967
BAK1101R	0.8421	0.9961	0.9915	0.9982
BAK1102L	0.7314	0.9845	0.9768	0.9940
BAK1102R	0.7208	0.9841	0.9741	0.9947
BAK8001L	0.6392	0.9803	0.9814	0.9959
BAK8001R	0.6098	0.9521	0.9569	0.9825
Average	0.6799	0.9779	0.9799	0.9915
STD	0.0890	0.0134	0.0096	0.0079

**Table S7.** The estimate time needed for an experienced user to correct the errors of the algorithm in minutes. This estimate assumes that it takes 3 seconds to correct one error (FP+FN). It is thus very approximate. Actual correction times highly depend on eccentricity, image quality and the presence of rods in the region of interest.

Subject	FastPeakFinder	Cunefare	Hamwood	Our FCN
BAK1088L	351	301	355	179
BAK1088R	326	231	282	134
BAK1089L	386	130	155	48
BAK1089R	383	127	152	43
BAK1093L	428	88	105	64
BAK1093R	455	89	106	81
BAK1101L	234	86	90	35
BAK1101R	172	76	73	31
BAK1102L	247	107	135	40
BAK1102R	272	139	147	47
BAK8001L	351	100	133	37
BAK8001R	369	147	167	41
Median	350	117	141	45