Deriving Visual Cues from Deep Learning to Achieve Subpixel Cell Segmentation in Adaptive Optics Retinal Images

Jianfei Liu, Christine Shen, Tao Liu, Nancy Aguilera, and Johnny Tam(✉)

National Eye Institute, National Institutes of Health, Bethesda, MD, USA
johnny@nih.gov

Abstract. Direct visualization of photoreceptor cells, specialized neurons in the eye that sense light, can be achieved using adaptive optics (AO) retinal imaging. Evaluating photoreceptor cell morphology in retinal diseases is important for monitoring the onset and progression of blindness, but segmentation of these cells is a critical first step. Most segmentation approaches focus on cell region extraction, without directly considering cell boundary localization. This makes it difficult to track cells that have ambiguous boundaries, which result from low image contrast, anisotropic cell regions, or densely-packed cells whose boundaries appear to touch each other. These are all characteristics of the AO images that we consider here. To address these challenges, we develop an AOSegNet method that uses a multi-channel U-Net to predict the spatial probabilities of the cell boundary and obtain cell centroid and region distribution information as a means for facilitating cell segmentation. Five-color theorem guarantees the separation of any touching cells. Finally, a region-based level set algorithm that combines all of these visual cues is used to achieve subpixel cell segmentation. Five-fold cross-validation on 428 high resolution retinal images from 23 human subjects showed that AOSegNet substantially outperformed the only other existing approach with Dice coefficients [%] of 84.7 and 78.4, respectively, and average symmetric contour distances [μm] of 0.59 and 0.80, respectively.

Keywords: U-Net · Level set segmentation · Adaptive optics · Five-color theorem · Cone photoreceptor neuron

1 Introduction

Adaptive optics (AO) retinal imaging can be used to directly visualize the morphology of photoreceptor cells in the living human eye [10]. Monitoring cell morphology can enhance the understanding of disease propagation at the cellular level. Subpixel cell segmentation is a prerequisite to monitor subtle cell changes as a one pixel error can cause up to 5% error in cell size measurements [5]. Accurate cell segmentation in AO images is hindered by low image contrast that

H. Fu et al. (Eds.): OMIA 2019, LNCS 11855, pp. 86–94, 2019.
https://doi.org/10.1007/978-3-030-32956-3_11

often exists at cell boundaries (Fig. 1A). Anisotropic shading on opposite sides of cells requires special handling, and low pixel sampling necessitates subpixel segmentation in order to better monitor subtle changes in cell morphology. To date, there has only been one published method for automated cell segmentation in these AO images [10] (circularly-constrained active contour model, CCACM [5]). CCACM dynamically constructs circularly-shaped priors for each cell, which subsequently constrains active contours used in order to identify cell contours. Although CCACM achieved high accuracy when cells were loosely packed, it is prone to over-segmentation in the case of densely-packed regions where neighboring cells are very close together (cell crowding). This restricts the applicability of CCACM.

Recently, segmentation methods based on deep learning have shown substantial improvement over traditional image processing approaches [2,4,8,9,12,14]. The key challenge with AO images is separating crowded cells whose boundaries appear to touch each other. One approach would be to adaptively adjust weights at cell boundaries to train neural networks [8]. However, this approach is prone to over-segmentation of densely-packed cell regions. Contour-aware approaches [2,14] are effective ways to address cell crowding, but are prone to under-segmentation. To simplify the task of cell decrowding, joint cell segmentation and detection, combined with the use of star-convex polygons to represent cell shapes, was utilized [9]. Another possible solution is to post-process prediction results from the deep learning method by using conditional random fields [3]. However, this approach does not naturally achieve subpixel cell segmentation. Level set method [13] is an efficient means to address this issue by propagating active contours in a subpixel-level step. Five-color theorem has been combined with level sets to segment crowded cells [6]. However, identifying cell regions in terms of image intensity is unreliable, making the subsequent level set propagation in each colored region inefficient.

This paper introduces a combined approach incorporating deep learning and level sets for improving segmentation of photoreceptor cells in AO retinal images, particularly in dense regions where neighboring cell boundaries appear to touch each other. Our approach is called AOSegNet. It utilizes a multi-channel U-Net to simultaneously extract cell centroid, region, and contour visual cues, instead of only predicting cell regions as in the case of the traditional U-Net. Next, centroid and region cues from deep learning are combined to separate any clusters of touching cell regions into distinct, untouching regions, based on the five-color theorem. Finally, visual cues and colored regions are used to achieve subpixel cell segmentation. Five-fold cross-validation was performed to compare AOSegNet with existing methods. These results open up the possibility of monitoring subtle cellular changes that occur during neurodegenerative retinal diseases.

2 Methodology

AOSegNet consists of three components: a multi-channel U-Net (Fig. 1B) to extract cell cues (centroids, Fig. 1C; regions, Fig. 1D; and contours, Fig. 1E), a five-color theorem approach to separate touching cells (Fig. 1F), and a region-based level set method (Fig. 1G) to determine the final subpixel segmentation.

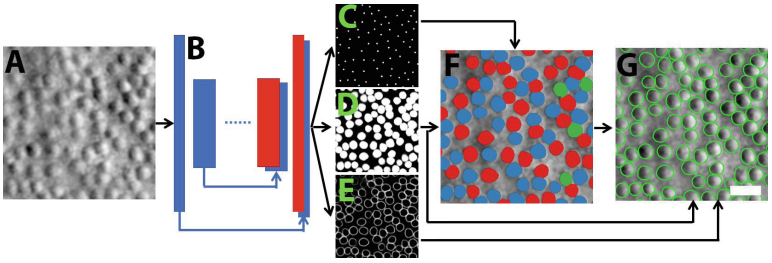

Fig. 1. Overview of AOSegNet for an example AO image of photoreceptors (A) using a multi-channel U-Net (B), which generates a set of visual cues consisting of centroids (C), regions (D), and contours (E). Cell centroid and region cues are used by the five-color theorem to separate touching cells. Red, blue and green regions represent distinct cells (F). A region-based level set segmentation is used to achieve subpixel cell segmentation (G). Scale bar, 20 μm (Color figure online)

2.1 Learning Cell Visual Cues

A multi-channel U-Net is leveraged to predict three cell visual cues: centroids, regions, and contours (Fig. 1B). Similar to the conventional U-Net [8], the multi-channel U-Net also consists of contracting and expanding paths (left and right sides, respectively). The contracting path is similar to a VGG network [11] that repeatedly applies 3×3 convolutions, followed by a rectified linear unit (ReLU) and 2×2 max pooling operations. The expanding path contains an upsampling series of the image feature map, which is a 2×2 convolution that concatenates image features from the contracting path, the upsampled feature map from the expanding path, and a ReLU.

Unlike the conventional U-Net [8] which only includes the region mask, the multi-channel U-Net contains centroid, region, and contour masks during training. A 3-channel label map is thus formulated, which improves prediction accuracy as they jointly constrain each other. It leads to a combinatorial loss function.

$$L = L_{centroid} + L_{region} + L_{contour} \tag{1}$$

Visual cues of cell centroids and regions are represented as binary masks, $I(\mathbf{x}), \mathbf{x} \in \Omega$, where Ω is the image domain. The corresponding predictions are $\hat{I}(\mathbf{x})$. The loss function for each binary mask is formulated as the combination of binary cross entropy and Dice coefficient loss. In this way, $L_{centroid}$ and L_{region} can both be defined as

$$S(\mathbf{x}) = -\frac{1}{2}\sum_{i=1}^{2} I(\mathbf{x_i}) \log \hat{I}(\mathbf{x_i}) - \frac{2\sum_{i=1}^{2} I(\mathbf{x_i})\hat{I}(\mathbf{x})}{\sum_{i=1}^{2} I^2(\mathbf{x_i}) + \sum_{i=1}^{2} \hat{I}^2(\mathbf{x_i})} \tag{2}$$

To improve the accuracy of cell contour localization, the contour mask $I_c(\mathbf{x})$ is represented as a spatial density, which assigns probability values to image points

near cell contours, assuming a Gaussian distribution, which results in

$$L_{contour} = S(\mathbf{x}) \exp\left(\left(I_c(\mathbf{x}) - \hat{I}_c(\mathbf{x})\right)^2 / \sigma^2\right) \tag{3}$$

where $\sigma = 0.5$ due to $I_c(\mathbf{x}) \in [0, 1]$. This term (Eq. 3) measures the intensity value changes between the labeled contour mask $I_c(\mathbf{x})$ and predicted contour mask $\hat{I}_c(\mathbf{x})$.

Altogether, the proposed multi-channel U-Net simultaneously predicts the probability masks of cell centroid, region, and contour cues (Figs. 1C-E) for a given AO retinal image (Fig. 1A).

2.2 Cell Decrowding

This step aims to extract cell regions that are clustered together in close proximity and separate them into groups of distinct regions. Although neighboring regions often touch, their simultaneously-learned centroids do not, which is key for efficient decrowding. Following Otsu's threshold method [7] to extract cell centroids and regions from their visual cue masks produced by the multi-channel U-Net, each centroid is used to identify its corresponding cell region through the watershed algorithm.

However, the extracted cell regions often contain segmentation errors due in large part to low pixel sampling and lack of subpixel accuracy. In order to achieve subpixel segmentation, cell regions within connected clusters of cells must first be disconnected from each another. We observe that cells within clusters can be separated based on the five-color theorem [1], which states that any 2D planar graph can be labeled with as few as five colors such that no neighbors have the same color. We can construct a planar graph with cell centroids as nodes, $V = \{v_1, v_2, \cdots, v_n\}$, with color $C(v)$ for each node. Cell centroids whose corresponding regions are connected to v_i are contained in $adj(v_i)$. The greedy coloring algorithm is used to assign a color to each cell region:

Algorithm 1. Greedy coloring

1: **for** $i = 1$ to n **do**:
2: $\quad c(v_i) := 0$
3: **end (for)**
4: **for** $i = 1$ to n **do**:
5: $\quad$ Let $c(v_i)$ be the smallest $\mathbb{Z}^+$ s.t. $c(v_i) \notin \{c(v_j) : v_j \in adj(v_i)\}$
6: **end (for)**

We can thus separate connected cells into different groups with distinct, separated cells inside, as illustrated in Fig. 1F.

2.3 Guided Level Set Subpixel Segmentation

This step achieves subpixel segmentation by combining region and contour cues from the multi-channel U-Net with identity priors from the five-color theorem. Identity priors globally constrain level set propagation, and region and contour cues locally adjust level sets. A multiphase level set segmentation framework is defined as

$$E = \int_{\Omega} \sum_{i=1}^{m} \left(E_{region} + E_{contour} + E_{identity} \right) d\mathbf{x} \tag{4}$$

where $m \leq 5$ is constrained by the five-color theorem. Let $I_r(\mathbf{x})$ be the region cue mask from the multi-channel U-Net, and $\phi : \Omega \rightarrow \mathbb{R}$ be a signed distance function that represents the level set function.

$$E_{region} = (I_r - \mu_1)^2 H(\phi) + (I_r - \mu_2)^2 (1 - H(\phi)) \tag{5}$$

Here, $H(x)$ is the Heaviside function, with $H(x) = 1$ if $x \geq 0$; otherwise $H(x) = 0$. μ_1 and μ_2 are mean values of the mask regions inside and outside of the level set ϕ, respectively.

$$E_{contour} = \mathrm{c}_1 F(I_c) |\nabla H(\phi)| \tag{6}$$

with $F(I_c) = 1 - I_c(\mathbf{x})$ because the contour cue mask is normalized to $I_c(\mathbf{x}) \in [0, 1]$ with large values at the cell boundary, where level set propagation should terminate.

Establishing signed distance functions ψ on different-colored cell regions (Fig. 1F) leads to the identity priors, which are defined as

$$E_{identity} = \mathrm{c}_2 \left(H(\phi) - H(\psi) \right)^2 \tag{7}$$

Here, $c_1 = 2$ and $c_2 = 1.5$ represent scalar weights for balancing the level set framework for all of the images in this paper. Note that the identity priors only allow level sets to propagate near the image boundary that was predicted by the multi-channel U-Net, which reduces merging of cell regions contained within a certain multicolored region, while still achieving subpixel level cell segmentation.

The level set evolution equation is derived from Eq. 4 using Eqs. 5–7.

$$\begin{aligned} \frac{\partial \phi_i}{\partial t} = \delta(\phi_i) \Bigg(& -(I_r - \mu_i)^2 + \prod_{j \neq i} (1 - H(\phi_j)) \Bigg) \\ & + \mathrm{c}_1 \mathrm{div} \left(F(I_c) \frac{\nabla \phi_i}{|\nabla \phi_i|} \right) + 2\mathrm{c}_2 (H(\phi_i) - H(\psi_i)) \Bigg), 1 \leq i \leq m \end{aligned} \tag{8}$$

Figure 1G shows the final cell segmentations computed using Eq. 8. Note that all touching clustered cells are successfully separated into individual cells with subpixel level accuracy.

2.4 Data Collection and Validation Methods

AO images of cone photoreceptors from 23 human subjects (age: 27.1 ± 8.8 years) were used to generate a total of 428 images (333×333 pixels), randomly selected from these subjects across different retinal regions. Note that AO images can vary substantially at different retinal regions of the same subject due to the variation of cone photoreceptor density, eye motion, and imaging conditions. Therefore, it is reasonable to have AO images from the same subject in both training and test datasets. Cones were manually labeled with subpixel accuracy by expert graders familiar with AO images for validation purposes. Five-fold cross-validation was performed to evaluate the accuracy and robustness of AOSegNet.

We compared segmentation results with CCACM [5], which is, to our knowledge, the only existing automated cell segmentation method for AO images of cone photoreceptors. Quantitative comparison was performed using six metrics: area overlap (AP), Dice coefficient (DC), area difference (AD), average symmetric contour distance (ASD), symmetric room mean square contour distance (RSD), and maximum symmetric absolute contour distance (MSD). Finally, cone diameters measured from our segmentation results were compared to previously published diameter measurements, including histological studies.

3 Experimental Results

3.1 Five-Fold Cross-Validation of Segmentation Accuracy and Robustness

Across each of the five folds, an average of 1343 corresponding cell regions were extracted by AOSegNet and CCACM. They were compared to each other and also to manually-labeled groundtruth. In all cases, AOSegNet performed substantially better than CCACM (Table 1). In each fold, the training time for the multi-channel U-Net was ∼6 h (2000 iterations; Microsoft Windows 7, Intel(R) core(TM) i7-6850K CPU, and dual NVIDIA GeForce GTX 1080 Ti GPUs). Following training, evaluation on each test dataset required less than 5 s per image.

Table 1. Segmentation accuracy comparison between AOSegNet and CCACM [5] over five-fold cross-validation

Method	AP (%)	DC (%)	AD (%)	ASD (μm)	RSD (μm)	MSD (μm)
AOSegNet	**74.2 ± 0.8**	**84.7 ± 0.6**	**19.9 ± 0.9**	**0.59 ± 0.02**	**0.70 ± 0.02**	**1.39 ± 0.04**
CCACM	66.0 ± 0.6	78.4 ± 0.4	26.9 ± 0.9	0.80 ± 0.02	0.98 ± 0.02	1.99 ± 0.05

Examples of segmentation results showed high cell segmentation accuracy on AO retinal images using AOSegNet (Fig. 2). Compared to CCACM, AOSegNet improved detection accuracy, and reduced both over- and under- segmentation (white arrows in Fig. 2). Our method combines spatial density contour cues as well as the five-color-theorem separation strategy, in order to accurately identify the contours of all photoreceptor cells.

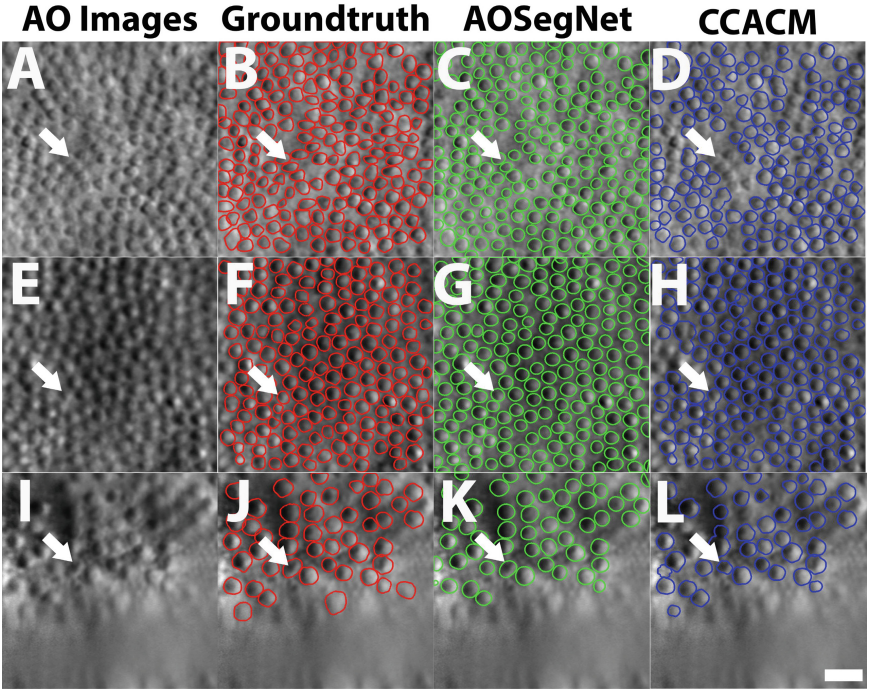

Fig. 2. Segmentation results on AO images varying in image quality and content. Compared to CCACM, AOSegNet improves detection accuracy (white arrows, top row), reduces over-segmentation (white arrows, center row), enhances under-segmentation (white arrows, bottom row), and performs well in the vicinity of image artifacts (lower portion, bottom row). Scale bar, 20 μm. (Color figure online)

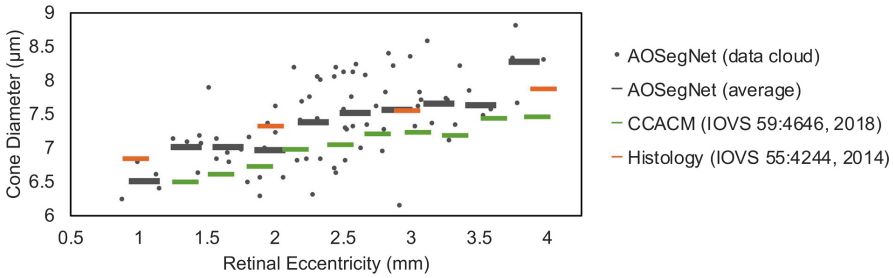

Fig. 3. Comparison of cone photoreceptor cell diameters generated with AOSegNet to those calculated with other methods. Each dot represents the average cone diameter measured within a single AO image (e.g. Fig. 1A). Since cone diameter varies depending on the location in the eye (retinal eccentricity), measured values were averaged every 0.3 mm in order to compare them to averaged values from previously-reported values. The average cone diameters measured were similar to previously-published values.

3.2 Cell Diameter Measurements

AOSegNet performed well across a test dataset consisting of 78 different AO images from healthy eyes. To demonstrate that the measurements were anatomically relevant, we computed cone photoreceptor cell diameters using the corresponding contours and compared them to those calculated with existing state-of-the-art methods (Fig. 3). Overall, AOSegNet measurements of cone diameters were similar to published values, including those measured from histological images. We also verified that the use of subpixel measurements improved accuracy: relative cell diameter differences of $7.9 \pm 0.3\%$ and $8.8 \pm 0.3\%$ were achieved for subpixel and pixel approaches, respectively, over the five folds, mean±SD.

4 Conclusion and Future Work

In this paper, we developed an AOSegNet for AO retinal images. A multi-channel U-Net was designed to simultaneously learn different types of visual cues (cell centroids, regions, and contours). These visual cues are used separately and in conjunction with each other in subsequent steps to intuitively improve segmentation performance. For example, cues integrated with the five-color theorem provide a simple solution to separate connected cell clusters, which substantially reduces segmentation errors when cells are crowded. By combining all learned and derived priors, we show that subpixel cell segmentation can be achieved. This subpixel representation is enabled in large part by the fact that cell contours were trained through a spatial density representation.

Five-fold cross-validation demonstrated that AOSegNet substantially outperforms the only existing AO photoreceptor cell segmentation method [5] across six different quantitative metrics (Table 1). These approaches will facilitate construction of normal databases of cell morphology in the living human eye, and will be useful for evaluating cell morphology in diseased eyes.

References

1. Appel, K., Haken, W.: Every planar map is four colorable. Illinois J. Math. **21**(3), 429–490 (1977)
2. Chen, H., Qi, X., Yu, L., et al.: DCAN: deep contour-aware networks for object instance segmentation from histology images. MedIA **36**, 135–146 (2017)
3. Chen, L., Papandreou, G., Kokkinos, I., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018)
4. Gu, Z., Cheng, J., Fu, H., et al.: CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging (2019, in press)
5. Liu, J., Jung, H., Dubra, A., Tam, J.: Cone photoreceptor cell segmentation and diameter measurement on adaptive optics images using circularly constrained active contour model. Invest. Ophthalmol. Vis. Sci. **59**(11), 4639–4652 (2018)

6. Nath, S.K., Palaniappan, K., Bunyak, F.: Cell segmentation using coupled level sets and graph-vertex coloring. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 101–108. Springer, Heidelberg (2006). https://doi.org/10.1007/11866565_13
7. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans Cybern. **9**(1), 62–66 (1979)
8. 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.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Schmidt, U., Weigert, M., Broaddus, C., Myers, G.: Cell detection with star-convex polygons. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 265–273. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_30
10. Scoles, D., Sulai, Y., Langlo, C., et al.: In vivo imaging of human cone photoreceptor inner segments. Invest. Ophthalmol. Vis. Sci. **55**(7), 4244–4251 (2014)
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)
12. Valen, D.V., Kudo, T., Lane, K., et al.: Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. PLoS Comput. Biol. **12**(11), e1005177 (2016). https://doi.org/10.1371/journal.pcbi.1005177
13. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. J. Comput. Vis. **50**(7), 271–293 (2002)
14. Xu, Y., Li, Y., Wang, Y., et al.: Gland instance segmentation using deep multichannel neural networks. IEEE Trans. Biomed. Eng. **64**(12), 2901–2912 (2017)