Oct 21st, 2016

Optic Cup Segmentation Using Large Pixel Patch Based CNNs

Yundi Guo  
Central South University, Changsha, China

Beiji Zou  
Central South University, Changsha, China

Zailiang Chen  
Central South University, Changsha, China

Qi He  
Central South University, Changsha, China

Qing Liu  
Central South University, Changsha, China

See next page for additional authors

Follow this and additional works at: https://ir.uiowa.edu/omia
Part of the Ophthalmology Commons

Recommended Citation
https://doi.org/10.17077/omia.1056

Rights
Copyright © 2016 the authors

Hosted by Iowa Research Online. For more information please contact: lib-ir@uiowa.edu.
Optic Cup Segmentation Using Large Pixel Patch Based CNNs

Yundi Guo\textsuperscript{1,2}, Beiji Zou\textsuperscript{1,3}, Zailiang Chen\textsuperscript{1,2} *, Qi He\textsuperscript{1,2}, Qing Liu\textsuperscript{1,2}, and Rongchang Zhao\textsuperscript{1,2}

\textsuperscript{1} School of Information Science and Engineering, Central South University, Changsha, China
\textsuperscript{2} Center for Ophthalmic Imaging Research, Central South University, Changsha, China
\textsuperscript{3} Center for Information and Automation of China Nonferrous Metals Industry Association, Changsha, China

Abstract. Optic cup (OC) segmentation on color fundus image is essential for the calculation of cup-to-disk ratio and fundus morphological analysis, which are very important references in the diagnosis of glaucoma. In this paper we proposed an OC segmentation method using convolutional neural networks (CNNs) to learn from big size patch belong to each pixel. The segmentation result is achieved by classification of each pixel patch and postprocessing. With large pixel patch, the network could learn more global information around each pixel and make a better judgement during classification. We tested this method on public dataset Drishti-GS and achieved average F-Score of 93.73\% and average overlapping error of 12.25\%, which is better than state-of-the-art algorithms. This method could be used for fundus morphological analysis, and could also be employed to other medical image segmentation works which the boundary of the target area is fuzzy.

Keywords: Optic Cup, CNNs, Pixel Patch, Segmentation

1 Introduction

Accurate optic cup (OC) segmentation on color fundus image is important for the diagnosis of glaucoma. Glaucoma is the world's first leading cause of irreversible blindness and has grown up to be a global public health problem. The clinical manifestation of glaucoma is mainly optic atrophy. The optic nerve head which usually called the optic disk (OD) passes visual information from the retina to the brain. The OD has a vertical elliptical shape\cite{19} and is divided into two separate zones: the optic cup and the neuroretinal rim. The atrophy changes the shape of the OC region. Therefore, analyzing the shape of the OC and OD quantitatively is crucial.

There are many works for OD segmentation and the precision is very high now\cite{15}\cite{1}\cite{14}. Unlike the OD, good methods for OC segmentation are still scarce

\* Corresponding author

\begin{singlespace}
and the existing results are low in precision that cannot support the diagnosis of glaucoma well. The boundary of OC is not as clear as OD, and the vascular cross point is always on it, so that OC segmentation is much harder in both manually and automatically. Some methods tried to segment OC with the detection of blood vessel[20] kinks because the 3D structure of OC makes the vessel bend at the boundary in 2D fundus images[4][3]. The challenge is to find a good initial estimation of the cup boundary and exclude vessel bends from a non-cup boundary. Statistical model-based methods tried to use regional color characteristics to find the best or an adaptive threshold to fit a boundary of the OC[18][6]. But in many fundus images, there is no obvious color difference within the disk to mark the cup boundary. Jun Cheng[2] proposed OD and OC segmentation using superpixel classification. Mittapalli[11] proposed a fuzzy c-means(FCM) based method for cup segmentation. Pixel clustering based methods could make a approximate boundary of the OC, but segmenting the OC accurately is still a challenging problem.

Deep convolutional neural networks (CNNs)[10][5] are a type of multi-layer, fully trainable models that can capture highly nonlinear mappings between inputs and outputs. These models were originally motivated from computer vision problems and thus are intrinsically suitable for image-related applications. Since the early 2000s, CNNs have been applied with great success to the detection, segmentation and recognition of objects and regions in images[9].

In this paper, we proposed an OC segmentation method using CNNs with large pixel patch to train and predict pixels belong to the OC area. Using large pixel patch, more global information around each pixel is learned by the network. With two-dimensional convolution and non-linear layers, the color features, vessel kinks or other unknown characteristics of the OC could be highly abstractly extracted. We tested this method on public dataset Drishti-GS and achieved average F-Score of 93.73% and average overlapping error of 12.25%, which is better than state-of-the-art methods. This method could also be employed to other medical image segmentation works which the boundary of the target area is fuzzy.

2 Method

We use CNNs to automatically extract features around each pixel through a large patch. The principle of the convolutional neural networks is like people’s eyes and brains to deal with what they see[8]. With large pixel patch, the networks could get as much as the global perception around each pixel.

2.1 OD Area Cut and Shading Unification

At the start we cut the square area which the OD relatively centers to, which is the region of interest[21]. This operation allows us to concentrate on OC segmentation without considering much useless background. There are numerous mature research methods for OD detection like the ASM-based method[17],
method based on approximate nearest neighbour field[12] and so on. As they are
not the key point of this work, the details are omitted here.

Because fundus images are not generally taken in consistent illumination
intensity, the variation of RGB value in different images is unfavorable for the
training. So we convert the cut images from RGB to Lab color space and apply
an equalization operation according to Formula 1 to all the three channels.

\[
P_i = \begin{cases} 
  P_{\text{mean}} + (P_i - P_{\text{mean}})/(P_{\text{max}} - P_{\text{mean}}) & P_i > P_{\text{mean}} \\
  P_{\text{mean}} - (P_{\text{mean}} - P_i)/(P_{\text{mean}} - P_{\text{min}}) & P_i < P_{\text{mean}} \\
  P_{\text{mean}} & P_i = P_{\text{mean}}
\end{cases} 
\]  

(1)

Where \( P_{\text{mean}}, P_{\text{max}}, P_{\text{min}}, P_i \) means the average value, the maximum value,
the minimum value of all the pixel values in the channel and the \( i \)th pixel value
respectively. \( i \) denotes the pixel index.

A Lab color space is based on nonlinearly compressed with dimension \( L \) for
lightness and \( a, b \) for the color-opponent dimensions. This operation will reduce
color differences from image to image to a certain level.

2.2 Data Augmentation and Training Data Extracting

In the field of medical science, the amount of available data with labels is rela-
tively small to other fields such as natural image processing. In this case, data
augmentation is needed to create more meaningful data from finite marked im-
ages. For fundus images, there are usually images from the left eye and the right
eye and the two are usually not identical. We flip the images horizontally as new
training data so that the amount of data with labels is doubled.

Theoretically, the higher the image resolution is, the more accurate the results
can be achieved. But the size of input images is proportional to the amount of
computation and memory. To make a balance, each image is scaled down to
120×120 to reduce the data scale for our CNN without obvious influence on the
results in practice.

Because the boundary of OC is fuzzy, local information like intensity gradient
or value is not sufficient to classify a pixel. We extract patches of size 51×51 for
each pixel as the input data for our CNN. This big size patch covers much area of
the whole square so that the network will learn more global information around
each pixel. The size we chose is a trade off between sufficient information and
computational burden. We also considered input patch sizes of 41×41 and 61×61
and so on.

2.3 Deep CNN Architectures

In this section we describe the architecture of our CNN with input patch size of
51×51 to explain the method we use. This CNN architecture contains three input
feature maps corresponding to RGB channels. There are three convolutional
layers, one max-pooling layer and one full connection layer in the network. Each
layer is followed with a nonlinear filtering activation function. 0.25 dropout is
applied to mitigate overfitting. A sigmoid layer is added to the end for binary classification. The structure is shown in Figure 1.

Fig. 1. CNN structure for OC pixel training and prediction. The net of CNN contains 5 layers: three convolutional layers, one max-pooling layer and one full connection layer. 51×51×3 denotes the dimension of the input image. 64@47×47 denotes 64 feature maps of size 47×47 in that layer.

2.4 Postprocessing

Pixel based prediction has a shortcoming that the predicted OC area is composed of isolated points. Some points may be apart from the unicom region and the shape of the predicted area may be not regular. Most of the isolated points are caused by prediction error. We use erosion and dilation to eliminate these isolated points. After that, as the OC area is usually like an ellipse, ellipse fitting operation is utilized to the result pixels.

3 Experiments

We use the public data set Drishti-GS[13] to evaluate our method. Though there are many fundus image datasets, most of them are not available online or publicly. Other available datasets do not incorporate challenges of OD and OC segmentation. Drishti-GS is a dataset of retinal images for OD and OC segmentation along with ground truth. There are 50 images can be directly downloaded. Open source library Keras is used for the construction of the networks. GTX TitanX×4 GPU speed-up workstation is deployed to perform the experiments. We will firstly test the performance of different patch sizes as input, and then compare our method with several other methods.

3.1 Performance on Different Patch Sizes

Different patch sizes as input are tested by CNNs. We use 40 images as training data and 10 as testing data. The overlapping error OE is used as the evaluation criterion. It is calculated by Formula 2.
\[ OE = 1 - \frac{||S \cap G||}{||S \cup G||} \]  

Where \( S \) is the segmentation result, \( G \) is the ground truth, and \( ||S \cap G|| \) is the number of shared positive pixels by \( S \) and \( G \), \( ||S \cup G|| \) is the number of positive pixels in \( S \) or \( G \). The overlapping error lies in \([0, 1]\) and smaller value indicates a higher segmentation accuracy. The average training time of each patch size is shown in Table 1. The overall overlapping error of each patch size is shown by boxplot in Figure 2.

<table>
<thead>
<tr>
<th>Patch Size</th>
<th>11×11</th>
<th>21×21</th>
<th>31×31</th>
<th>41×41</th>
<th>51×51</th>
<th>61×61</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATE(s/image)</td>
<td>5</td>
<td>11</td>
<td>32</td>
<td>73</td>
<td>95</td>
<td>113</td>
</tr>
</tbody>
</table>

Fig. 2. Box plot of the segmentation performance achieved by CNNs over 6 different patch sizes. The plot uses overlapping error to measure the performance for different patch sizes. The patch sizes are 11×11, 21×21, 31×31, 41×41, 51×51 and 61×61 respectively. The central mark represents the median, the edges of the box denote the 25% and 75% in result order.

With the increasing of patch size, the results become better and stable. With small patch, the network can only see a tiny scope around each pixel. The color does not change too much in the limited area and other pixel patches not belong to OC may be quite similar to the ones in OC area, so the network will be easily
overfitted during training. When the patch size becomes larger, more global information is contained. However, with bigger patch size, more computer memory is needed as well. In our experiments, to achieve better performance, patch size of $61 \times 61$ is acceptable. The prediction phase is fast and the computing time can be ignored.

### 3.2 Comparision With Other Methods

There are several previous methods have been proposed these years. Methods like threshold[7], r-bend[7], ASM[18], regression[16] are well compared in the latest papers like [2], [11] and so on. So we conducted experiments to compare with the superpixel based method[2] and the spatially weighted fuzzy c-means(SWFCM) based method[11]. We use the results of patch size $61 \times 61$ for the comparision, of which the test score is 0.35 and the test accuracy is 0.92. For better quantitative evaluation, we add $F$-Score to measure the segmentation results according to ground truth. They are defined as Formula 3.

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F-Score = \frac{2 \times P \times R}{P + R} \tag{3}$$

Where $TP$, $FP$, $FN$ means true positives, false positives and false negatives of the classified pixels respectively. $P$ is precision and $R$ is recall. The $P$ and $R$ are calculated based on the overlapped area between the detected contour and the groundtruth.

<table>
<thead>
<tr>
<th>Method</th>
<th>OE</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superpixel[2]</td>
<td>21.31%</td>
<td>85.98%</td>
</tr>
<tr>
<td>SWFCM[11]</td>
<td>21.21%</td>
<td>87.91%</td>
</tr>
<tr>
<td>Proposed</td>
<td>12.25%</td>
<td>93.73%</td>
</tr>
</tbody>
</table>

![Fig. 3. Sample results. From left to right columns: (a) the original images, (b) the ground truth, (c) is by superpixel[2], (d) is by SWFCM[11], and (e) is the proposed method.](image)

Table 2 and Figure 3 shows the results according to ground truth. The proposed method achieves an overlapping-error of 12.25% and $F$-Score of 93.73,
which means that the segmented cups are mostly overlapped with the ground truth. This enhanced performance is mainly because it combined the around area together of each pixel to automatically learn the image features. Each pixel whether belong to OC is deeply analysed like how people’s eye and brain work. Methods only considering local color or manually setting related features are weak to tackle with complex OC area. When the dataset is small, the generalization ability of the proposed CNN based method is better than other machine learning based methods. Because of the big size patch and CNN, which can peer the global features automatically, the grey level and the blood vessel shape are effectively combined together automatically without special design.

4 Conclusion

In this paper, we proposed an OC segmentation method using CNNs with large pixel patch to train and predict pixels belong to the OC area. With the help of GTX TitanX×4 GPU speed-up workstation, big size patch is allowed to train and the speed becomes acceptable. The CNNs make full use of the global information around each pixel and output the prediction of pixel labels. We tested this method on different pixel patch size and confirmed that bigger pixel patch makes better segmentation. Results on the Drishti-GS dataset showed that our proposed model significantly outperformed prior methods on OC segmentation. In this work, the OC segmentation problem is formulated as patch classification task that much duplicate information is reprocessed. In the future work we will improve the model to combine the whole image together for pixel prediction.

References


