Faster R-CNN classification for the recognition of glaucoma

Ajitha S and M V Judy

Cochin University of Science and Technology, Pin- 682022, Kerala, India
E-mail:ajithas@cusat.ac.in,judy.nair@gmail.com.

Abstract. Glaucoma is an optic neuropathy characterized by progressive degeneration of retinal ganglion cells. The early identification of Glaucoma is extremely important as it is detrimental to one’s blindness. In this paper, we present the identification of glaucoma using faster R-CNN which is one of the most well-known object detection neural networks. The proposed method uses artificial intelligence and enhanced deep learning to detect Glaucoma. Faster R-CNN comprises two modules, the region proposal network (RPN), in which the region of object is distinguished on the picture, and a network that enables to classify the objects in the proposed region. We have accomplished the finest output by applying a transfer learning scheme with ResNet50 and VGG16. Using ResNet we have detected Glaucoma with up to 96% accuracy. The test results obtained by making use of two unique publicly available data sets DRISHTI_GS and ORIGA with 751 images demonstrate that this arrangement can be a significant alternative for the computer design aid framework for the large-scale screening programs of glaucoma detection.

Keywords: Faster R-CNN, Optic Nerve, Glaucoma, Transfer Learning, Resnet 50, VGG16

1. Introduction

Nowadays, Blindness affects the aged people due to their neglected concern over health conditions. Glaucoma is the second most common cause of blindness, declared by world health organisation(WHO). According to research, the frightful fact about this disease is that people affected, started to increase day by day from 5.2 million and it may reach 80 million in 2020 and 111.8 million in 2040. Glaucoma arises when drainage canal in the eye is partly or completely blocked which leads to the increase intraocular pressure which damages the opticnerve. If this damage left untreated, may lead to total blindness. Therefore, the early detection of glaucoma is necessary. One of the commonly used imaging modality to diagnose glaucoma are digital fundus images [11]. Two main approaches available for glaucoma diagnosis are Optic cup-to-disc ratio (CDR) measurement and Nerve fiber layer (NFL) valuation. Recent studies show that glaucoma screening can be effective using a test with initial automated classification and suspicious cases are then referred to an ophthalmologist who performs additional tests and examinations for final confirmation of the diagnosis. In this work we used a different approach using Faster R-CNN algorithm for glaucoma detection problem.
2. Related Works

The study of automatic detection of Glaucoma has been going on for several years. Most of the research is made on the diagnosis of glaucoma. It is also important to detect the suspected glaucoma by analysing haemorrhages, whether it is present or absent in a particular region near the optical disc, in fundus image. Region of Interest segmentation is performed on optical disc and haemorrhage is segmented in particular region by means of adaptive thresholding technique. The detection of intermediate grade of glaucoma is one of the method of glaucoma screening system. The classification accuracy of 93.57% is obtained with the classification of suspected glaucoma [12]. A method proposed by Wong et al.2008 [14] to calculate the CDR after obtaining the optic cup and optic disc masks using level-set techniques on 104 images and found that their method produced results with a variation of up to 0.2 CDR units from the ground truth.

Hybrid feature set consists of functional and structural features that are used to detect glaucoma. Visual field testing and Inter ocular pressure measurement examine the functional features. The structural features (cup to Disc ratio) are detected by OCT imaging technology. K Nearest Neighbour classifier showed 96% accuracy using fundus images in the classification of accurate automated glaucoma systems [7].

Alam Diptu et al.2018 analysed the presence of glaucoma using adaptive Neuro Fuzzy inference system, followed by OCT and tonometry. ANFIS model of artificial intelligence performs the prediction of presence of glaucoma, the absence of glaucoma and whether the patient is suspected of glaucoma. Intraocular pressure (IOP) and inferior quadrant thickness are the two important parameters that influence glaucoma. ANFIS provides an accuracy of 81.25% in the recognition of glaucoma [5]. In glaucoma, the optic Nerve head can be destroyed and it leads to loss of vision. The image based features and disease related features are the two essential features introduced by Mousa et al. (2019) to identify glaucoma. Support vector machine classifier produced an accuracy of 87% and artificial Neural Network produced an accuracy of 98% in the automatic detection of glaucoma. These SVM and ANN classifiers can be performed with and without data normalisation.

Mamta Juneja et al. 2019 introduced glaucoma automatic detection by deep convolutional neural network. Due to the increase in intraocular pressure in the eye, the optic nerve gets affected and causes irreversible blindness. By means of convolutional Neural Network, hierarchical information can be computed from the digital fundus image, through which are able to differentiate non glaucomatous and glaucomatous image patterns. Artificial intelligence expert systems segment both optic cup and optic disc. On analysis of over 50 digital fundus images, an accuracy of 95.8% was obtained for optic disc and an accuracy of 93% was obtained for cup segmentation. [6]

To professionally track the redundancy of glaucoma, Liu et al presented glaucoma detection, which is based on attention. An attention based large scale database is established which consists of about positive glaucoma (4,878) digital fundus images and negative glaucoma (6,882) digital fundus images. The total number of fundus images are 11,760 out of which the attention maps of about 5,824 images are acquired through an eye-tracking experiment. The prominent regions for glaucoma detection are emphasized, in the attention prediction subnet, attention maps are prophesied. The localised pathological areas envision the features, which are added to AG-CNN structure to improve the performance in glaucoma detection [1].

Mohammed et al. 2018 developed the localization of optic disc with entropy-based algorithm. Since the vision loss mainly depends on optic disc. The Optical disc Area contains enormous information. Therefore, entropy is calculated over several patches of the optical disc by sliding window technique. Even more features can be calculated from the optical disc in order to improve robustness. It achieves an accuracy of 89% in optical disc localization.[9]

Yuan Gao et al. 2019 proposed automatic optic disc segmentation. Optic Disc segmentation and localization play a vital role in detection of eye diseases. The method of contour extraction which has saliency detection as well as threshold techniques has been applied. The excellent oval
shaped, local image fitting model is developed in order to extract the complete precise optic disc boundary. This LIFO model not only considers intensity, but also shape information which is based on optic disc boundary, that avoids the PPA influence, noise and blood vessels. The F Score of 0.951 is obtained with LIFO

Andres et al 2019, formed a fresh clinical database ACRIMA, comprising 705-labelled images. It is composed of 309 normal images and 396 glaucomatous images which has been made publicly accessible for the analysis of glaucoma.[15]. For automatic glaucoma valuation with fundus images, the five models such as VGG16, ResNet50, VGG19, Xception and InceptionV3 are calculated. By cross validation as well as cross testing validation strategy, high specificity and sensitivity was obtained compared to various methods.

Baida Bander et al introduces a technique for Optic Disc and Cup Segmentation in Colour Fundus image for the diagnosis of Glaucoma. Glaucoma is characterised by Vertical Cup to Disc ratio and is mainly considered for detecting glaucoma. Delineation of Optic Disc features influences the analysis of glaucoma. Dense Net is the deep learning method with fully CNN used for optic Disc and optic cup segmentation and it also facilitates pixel wise classification with its U-shaped architecture. Optic Disc, Optic cup boundary values are used to validate CDR for glaucoma diagnosis.

To our knowledge very few deep learning approaches have been proposed for glaucoma detection. In this study we used Faster R-CNN pre trained convolutional neural network model and region proposed network.

3. Methodology

3.1 Materials Used

Below listed are the materials and tools used while developing the proposed system using the python 3.6.8 platform. In this work, the classifications tasks are considered for 2 classes because the most of the data sets were defined with two classes (glaucoma and normal).

1. Keras Version 2.0.3 - the library we have used for deep learning throughout the development of the system. Keras is integrated as the front-end part of Tensor flow.
2. Tensor flow Version 1.8.0 - A deep learning library provided open source by Google, along with its different packages facilitates deep learning.
3. Open CV Version 4.0 - Open CV denotes open source computer vision library that facilitates the real-time computer vision and is used throughout the development of image segmentation and basic image operations such as read and write.
4. PyCharm IDE - This integrated development Environment act as the platform for building a python application

3.2 Dataset

3.2.1 DRISHTI - GS - The DRISHTI-GS dataset has retinal fundus images (101) for segmentation of optic plate. It is separated into preparing and testing sets of 50 and 51 images, individually. 4 ophthalmologists have annotated all the images with clinical encounters of 3,5, 9 and 20 years, independently.

3.2.2 ORIGA – ORIGA is a freely accessible dataset which contains 650 images of retinal fundus for the purpose of benchmarking computer aided segmentation and also the classification figured a strategy based on reconstruction for localization as well as optic plates classification

3.3 Architecture of R-CNN

Region proposals were used by the family of object detectors to detect objects within images. The Time taken to detect objects in an image completely depends on the number of proposed regions. Faster R-CNN detectors are developed so as to improve the detection performance with a large number of regions.
3.3.1 VGG 16

VGG16 architecture comprises of twelve convolution layers, trailed by pooling layers that are supreme and afterward 4 fully-connected layers lastly, a SoftMax classifier of 1000-way. 

**First and Second Layers**: The input RGB image of AlexNet is of size 224x224x3 that are fed into the first two convolutional layers with feature maps 64 or filters size 3×3 and same pooling with 14 Stride. The dimensions of the image will change to 224x224x64. At that point, VGG16 applies pooling layer (maximum) or channel size 3×3 and a stride of two for sub-sampling layer. It will be diminished to the subsequent image dimensions of 112x112x64.

**Third and Fourth Layer**: The 2 convolutional layer with feature maps of 128 of size 3×3 and stride 1. At that point there is again a pooling layer with maximum of 3×3 channel size and a stride 2. It is similar as past pooling layer aside from it has feature maps 128 and so the diminished output will be 56x56x128

**Fifth and Sixth Layers**: These layers also convolutional layers of channel size 3×3 and stride one. Here, both utilized feature maps of 256. These convolutional layers are trailed by a pooling layer maximum with channel size 3×3, a stride 2 and have feature maps of 256

**Seventh to Twelveth Layer**: Two arrangements of 3 convolutional layers pursued by a pooling layer maximum. All of the convolutional layers have 3×3 size of 512 filters and a stride 1. At last, the diminished size will be 7x7x512.

**Thirteenth Layer**: Through a fully connected layer the convolutional layer output is straighten with 25088 feature maps that have the size 1×1 for everyone.

**Fourteenth and Fifteenth Layers**: These layers are fully connected layers of two with 4096 units.

**Output Layer**: At last, there is a output layer SoftMax ŷ with the 1000 number of potential values.

3.3.2 RESNET 50

ResNet meant for residual network, however what is residual learning? Deep CNN have accomplished the classification result of human level image. Deep networks capable to extract high, middle and low level features and classifiers are in an end to end multi layer fashion, and the feature "levels" of are advanced by the quantity of stacked layers. By taking a gander at the ImageNet result, the stacked layer is of crucial significance. When a deeper network begins to converge, a issue of degradation has been uncovered: accuracy gets saturated (obvious) with the network depth expanding and afterward quickly degraded. It prompts higher training error when such degradation isn't brought about by over fitting by or adding more layers to a deep network. The decay of training accuracy shows that not all frameworks are anything but difficult to optimize. This is eradicated by, Microsoft presented a deep residual learning framework. Straightforwardly fit a desired underlying mapping, rather than trusting each of the stacked layers, they explicitly let these layers fit a residual mapping. By utilization of the residual network, there are numerous issues that can be fathomed, for example,

- ResNets are anything but difficult for optimization, however “plain" networks (that are essentially stack layers) when the depth increments, it shows higher training error.
- ResNets, without much of a stretch gain accuracy from extraordinarily expanded depth, delivering results, which are superior to previous networks.

Plain Network: Plain baselines are for the most part enlivened by the philosophy of VGG nets (Figure 1, left).
Convolutional layer for the most part have 3×3 filters and follow two straightforward principles:
1. For a similar output feature map, the layers with number of filters will be the same;
2. If the size of the features map becomes halved, the quantity of filters is multiplied to save the each layer’s time complexity. ResNet model consist of few filters and less complication than of VGG nets.

Residual Network: The right image in figure1 exemplifies basic residual block of Resnet. Depending upon the above plain network, an input shortcut connection there will be a transformation of network in case of counterpart residual version. They identity the shortcuts F(x{W}+x) can be legitimately utilized when there are similar dimensions for the input and output At the point when the dimensions increment it thinks about two options:
1. Identity mapping is performed with the shortcut, with extra zero passages padded for dimensions expanding. This choice presents no extra parameter.
2. The projection shortcut in F(x{W}+x) helps in dimension matching (done by 1×1convolutions). With both of options, when the shortcuts go crosswise over two size feature maps, it performed with 2 stride. Every ResNet block is either 2 layers deep or 3 layers deep. 50-layer ResNet: Every 2-layer block is supplanted in the net of 34-layer with this block of 3-layer bottleneck, provides a ResNet of 50-layer. They used the second choice for expanding the dimensions. This model has 3.8 billion FLOPs. A 224×224 harvest is randomly sampled from an image otherwise with its horizontal flip, with the per-pixel mean is subtracted. The learning rate begins from 0.1 and is partitioned by 10 for up to 60×10000 iterations, when the error plateaus and the models are trained. They utilize a weight decay of 0.0001 and a momentum of 0.9.

Steps pursued by a Faster R-CNN algorithm shown in figure 2 to detect objects in an image:

**Figure 1.** Plain network Residual network

**Figure 2.** Diagram of FR-CNN
1. It returns the feature maps for the image when an input image is taken and passes it to the ConvNet.
2. Get object proposals by applying region Proposal Network on these feature maps.
3. In order to cut down every one of the proposals to a similar size, apply ROI pooling layer.
4. So as to group any, predict the bounding boxes for the image. At last, pass these proposals to a fully connected layer.

4. Results and Discussion

This study used 751 fundus retina images from Drishti and Origa data base. Out of the total 751 images, 80% (601) images are chosen for training the model and the remaining 20% (150) images are used to test the accuracy of classifier. VGG16 and Resnet 50 are used as backbone of the Faster R-CNN and is initialized with an ImageNet pretrained model. The entire network is fine-tuned end to end with a training set of 601 images for 1000 iterations about 500 epochs shown in Table 2 & figure 4 on a NVIDIA TITAN XP GPU. Training time and model size for both VGG16 and Resnet 50 depicts in Table 1 and figure 3.

Our model achieved 97% of overall training accuracy. It describes that 583 out of 601 images were classified correctly. The validation accuracy narrates the accuracy achieved on the test set. The proposed system attains 92.5% on Drishti and 90% accuracy on Origa database. Our model achieved 95% sensitivity, 90% specificity and 90% PVV on Drishti database illustrated in Table 3 and figure 5. Experimental result proved that the present system achieved 95% sensitivity, 85% specificity and 86.56% PVV on Origa database. Figure 6 to figure 13 shows examples of the classification by the proposed model in the Drishti and Origa database.

Table 1. Training time and Model size

| Network     | Training Time | Model size |
|-------------|---------------|------------|
| Resnet 50   | 10.7 hours    | 113.9 MB   |
| VGG16       | 9.6 hours     | 106 MB     |

Figure 3 Comparison of training time and model size for Resnet 50 and VGG 16

Table 2. After monitoring training process, Number of epochs selected by validation set.

| Group  | Epochs Number | Number of Iterations |
|--------|---------------|----------------------|
| RESNET | 500           | 1000                 |
| VGG16  | 500           | 1000                 |
Figure 4 Comparing Resnet and VGG16 with epochs and iterations

Table 3. Classification of Drishti and Origa using Resnet

|     | Sensitivity (%) | Specificity (%) | PVV (%) | Accuracy (%) |
|-----|-----------------|-----------------|---------|--------------|
| Drishti | 95              | 90              | 90      | 92.5         |
| Origa   | 95              | 85              | 86.5    | 90           |

Table 4. Classification of Drishti and Origa using VGG 16

|     | Sensitivity (%) | Specificity (%) | PVV (%) | Accuracy (%) |
|-----|-----------------|-----------------|---------|--------------|
| Drishti | 94              | 93              | 93      | 92           |
| Origa   | 92              | 87              | 88      | 90           |

Figure 5 Comparison of Drishti and Origa Dataset using Resnet
Figure 6. Glaucomatous image using RESNET pre-trained on the dataset DRISHTI

Figure 7. Non-Glaucomatous image using RESNET pre-trained on the dataset DRISHTI

Figure 8. Glaucomatous images using RESNET pre-trained on the dataset ORIGA
Figure 9. Non-Glaucomatous image using RESNET pre-trained on the dataset ORIGA

Figure 10. Glaucomatous image using VGG16 pertained on the dataset DRISHTI

Figure 11. Non-Glaucomatous image using VGG16 pre-trained on the dataset DRISHTI
5. Conclusion
In this paper, we exploited and evaluated the application of Faster R-CNN approach to detect glaucoma accurately in colour fundus images. We tested two publicly available Drishti-GS and OrigA data set on VGG16 and Resnet 50 architecture. Experimental result showed that Resnet 50 was offered better performance. The test results obtained by making use of two unique data sets with 751 images demonstrate that this arrangement can be a significant alternative for the computer design aid framework for the large-scale screening programs of glaucoma detection.
In the future we will continue our research to develop new architectures for detecting glaucoma on large database.

REFERENCES
[1] Li L, Xu M, Liu H, Li Y, Wang X, Jiang L, Wang Z, Fan X, Wang N. “A Large-scale Database and a CNN Model for Attention-based Glaucoma Detection”. IEEE trans on med imaging. Vol,pp 1-11, Jul 2019
[2] Diaz-Pinto, Andres & Morales S, Naranjo V, Köhler T, Mossi J M, Navea A. “CNNs for automatic glaucoma assessment using fundus images: an extensive validation”. Biomed engg online, Vol 18, no.1,pp 29, Dec2019,
[3] Raghavendra U, Hamido Fujita, Sulatha V Bhandary, Anjan Gudigar, Jen Hong Tan, Rajendra acharya U, “Deep Convolution Neural Network for Accurate Diagnosis of Glaucoma Using Digital Fundus Images”, Information Sciences, Vol 441, pp41-49, May2018.
[4] Dey A, Dey KN. “Automated glaucoma detection from fundus images of eye using statistical feature extraction methods and support vector machine classification”. Industry Interactive Innovations in Science, Engg and Tech, Springer, Singapore (pp. 511-521), 2018

[5] Diptu NA, Khan MA, Debnath S, Al Imam A, Ador K A, Rahman RM. Early “Detection of glaucoma Using Fuzzy Logic in Bangladesh Context”. In 2018 International Conference on Intelligent stems IEEE (pp. 87-93) Sep 2018.

[6] Juneja M, Singh S, Agarwal N, Bali S, Gupta S, Thakur N, Jindal P. “Automated detection of glaucoma using deep learning convolution network”. Multimedia Tools and Applications, pp.1-23, 2019

[7] Khalil T, Khalid S, Muhammad Usman Akram, Amina Jameel. “Review of Machine Learning techniques for glaucoma detection and prediction”. IEEE Computing Conference pp. 620-632, Aug 2017.

[8] Ahilan, Jarin T, et. Al, “Segmentation by Fractional Order Darwinian Particle Swarm Optimization based Multilevel Thresholding and Improved Lossless Prediction based Compression Algorithm for Medical Images”, July 2019, IEEE Access PP(99):89570-89580. DOI: 10.1109/ACCESS.2019.2891632.

[9] Muhammed LA. “Localizing Optic Disc in Retinal Image Automatically with Entropy Based Algorithm”. Int journal of biomed imaging pp.1-7, 2018.

[10] Gao Y, Yu X, Wu C, Zhou W, Lei X, Zhuang Y. “Automatic Optic Disc Segmentation Based on Modified Local Image Fitting Model with Shape Prior Information”. Journal of healthcare engg. pp.1-10, 2019.

[11] Salam AA, Khalil T, Akram MU, Jameel A, Basit I. “Automated detection of glaucoma using structural and non structural features”. Springer plus. Vol 5, no.1, pp.1519. Dec 2016.

[12] Sengar N, Dutta MK, Burget R, Ranjoha M. “Automated detection of suspected glaucoma in digital fundus images”. In 2017 40th Intl Conf on Telecommunications and Signal Processing, IEEE. , pp. 749-752, Jul 2017.

[13] Al-Akhras M, Alawirdhi M, Habib M. “Using soft computing techniques to diagnose Glaucoma disease”. Journal of infection and public health, pp 1-8, Oct 2019.

[14] Wong DWK, Liu J, Lim JH, Jia X, Yin F, Li H, Wong TY. Level-set based automatic cup-to-disc ratio determination using retinal fundus images in ARGALI. In: 30th Annual International IEEE EMBS Conference, vol. 30. 2008, p. 2266–. https://doi.org/10.1109/IEMBS.2008.4649648.