International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 10 Oct 2021, pp: 707-712 www.ijaem.net ISSN: 2395-5252

Image Segmentation and Localization in Retinal Fundus Images

Soham Deshmukh

Submitted: 05-10-2021 Revised: 18-10-2021 Accepted: 20-10-2021

ABSTRACT:There is a noticeable rise in computing solutions for disease detection using Diagnostic Images. These recent scientific advances to provide computer-aided guidance using machine learning approaches, have given a chance to reach a foreseeable goal. Diabetic Retinopathy is considered as the most pervasive cause for vision deterioration affectedly mainly the working class population.The eye diseases lead either to reform retinal components or/and appearance of lesions. Those lesions differ in terms of size, shape, contrast, etc. Moreover, they always have similar characteristics to other retinal components or other pathological lesions. Therefore, eye diseases 'diagnosis is a difficult task, that takes several parameters into account. Hence, Deep Learning represents an adequate approach to resolve such problems.The aim of this paper was to identify which algorithms were more suitable in order to identify different diseases of the human eye namely Microaneurysms, Hard Exudates, Soft Exudates, Haemorrhages and Optic Discs. This paper focuses on Segmentation and Localization of these diseases.VGG-NET-16 and RES-NET50 algorithms were used for Image Segmentation whereas YOLOv4Tiny, YOLO and Faster-RCNN was used for Localization. YOLOv4Tiny gave the best results among the tested localisation methods to detect Optic Discs and Fovea.

KEYWORDS: Deep Learning, Object Detection, Neural Network, Segmentation.

I. INTRODUCTION

The need for more precise and detailed object recognition becomes crucial when we have a goal to achieve i.e complex image understanding. In this context, precisely estimating the class and location of objects contained within the images, a problem known as object detection is encountered. Undiag- nosed diabetic retinopathy (DR) leads to vision impairment and blindness and its early detection can significantly reduce this by 50%. Image classification saw an advancement when AlexNet won the 2012 ImageNet[1] competition using deep convolutional neural networks. They trained a deep CNN to classify 1.2 million high resolution images in the ImageNet contest that has 1000 categories.From this, many researchers became interested in finding a novel way to develop an efficient deep convolutional neural network [2][3][4]. This lead to many novel methods image segmentation as well as object detection, to take shape.However, manual detection of diseases like Microaneurysms,Hard Exudates, Soft Exudates, Haemorrhages and Optic Discs is dependent on the examiner and is time consuming. Automatic detection would make it possible to perform widespread screening of at-risk population, and for remote communities. Deep learning (DL) methods have an advantage of automatic feature extraction and they have achieved promising results in different computer vision and image analysis applications including automatic fundus image analysis.Its performance depends on: (a) an efficient search strategy; (b) a robust image representation; (c) an appropriate score function for comparing candidate regions with object models; (d) a multi-view representation and (e) a reliable non-maxima suppression.The paper bases its findings on the Indian Diabetic Retinopathy Image Dataset [5].

II. RELATED WORK

Diabetic retinopathy diseases are detected and also iden- tifying by image processing algorithms. Some of the novel exudate detection methods are described in Kaur and Kaur, 2015[6]; Sopharak et al., 2009[7]; Gharaibeh, 2016 [8]and many more [19] [20].

Kaur and Kaur (2015)[6] and Gharaibeh (2016)[8] used pixel based and retinal based image evaluation to describe the automatic detection of exudates. Confusion matrix was used to show sensitivity and specificity.It could not provide efficient sensitivity results.

Fuzzy c-means (FCM), a clustering based exudate detection method was used by Sopharak et al.[7] It depended on optic disc detection and blood vessel removal process. According to the obtained results the exudates could be detected, not distinguish the exudates characteristics.

III. PROPOSED WORK

A. SEGMENTATION:

The segmentation of the said four diseases [9](Microaneurysms, Hard Exudates, Soft Exudates andHaemorrhages) were done using algorithms VGG-NET16 [11] and RESNET50 [12] For a given image, this task seeks to get the probability of a pixel being a lesion(Microaneurysms, Hard Exudates, Soft Exudates or Hemorrhages). Although different retinal lesions have distinct local features, 535 for instance, MA, HE, EX, SE have a different shape, color and distribution characteristics, these share similar global features. In most DL tasks, using inadequate learning data can produce a weak and inaccurate performance. However, transfer learning paved the way to train models and acquire substantial results without the need for massive data. Hence, this work adopted this technique and used the pre-trained weights from the COCO [10] dataset to improve the model performance to detect several brain diseases. The previously learned COCO [10] features supplied the model with additional image recognition essentials needed for the detection process. Also, to further optimize the pre-trained model, the application of fine-tuning adjusted the resource allocation and prevented the depletion of memory during training and testing . The initial step to fine-tune the model was to replace the default class numbers from 80 to three, in which the three correspond to the Microaneurysms, Hard Exudates, Soft Exudates or Hemorrhages.

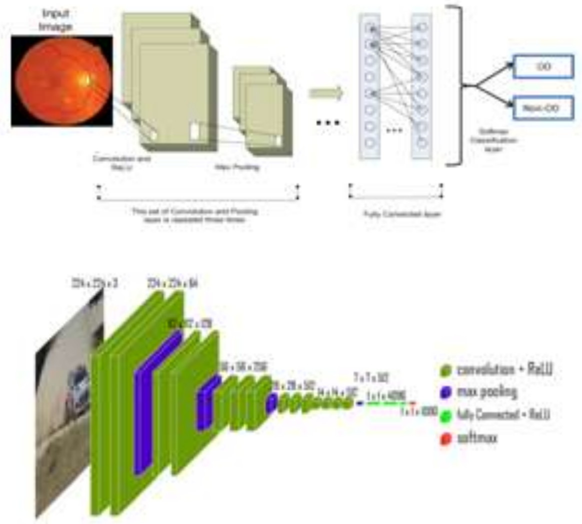

Fig 2. VGGNET Architecture

1)VGG-NET: The VGG [11] is based on convNet whose input is a 224*224 RGB image. The preprocessing layer takes the image of pixel values in range of 0–255 the after subtraction, it calculates the mean image values, calculated over the entire ImageNet training set.The input images after preprocessing are passed through these weight layers. The training images are passed through a stack of convolution layers. There are a total of 13 convolutional layers and 3 fully connected layers in VGG16[11] architecture. VGG has smaller filters (3*3) with more depth instead of having large filters. It has ended up hav- ing the same effective receptive field as if you only have one 7 x 7 convolutional layers.Another variation of VGGNet[11] has 19 weight layers consisting of 16 convolutional layers with 3 fully connected layers and the same 5 pooling layers. In both variations of VGGNet[11] there consists of two Fully Connected layers with 4096 channels each which is followed by another fully connected layer with 1000 channels to predict 1000 labels. Last fully connected layer uses softmax layer for classification purposes. Architecture walkthrough: The first two layers are convolutional layers with 3*3 filters, and first two layers use 64 filters that results in 224*224*64 volume as same convolutions are used. The filters are always 3*3 with stride of 1. After this, pooling layer was used with max-pool of 2*2 size and stride 2 which reduces height and width of a volume from 224*224*64 to 112*112*64. This is followed by 2 more convolution layers with 128 filters. This results in the new dimension of 112*112*128. After pooling layer is used, volume is reduced to

56*56*128. Two more convolution layers are added with 256 filters each followed by down sampling layer that reduces the size to 28*28*256. Two more stack each with 3 convolution layer is separated by a max-pool layer.After the final pooling layer, 7*7*512 volume is flattened into Fully Connected (FC) layer with 4096 channels and softmax output of 1000 classes.

2)RESNET: Before ResNet,[12] there had been several ways to deal the vanishing gradient issue, for instance, GoogleNet [13] (also codenamed Inceptionv1) adds an aux- iliary loss in a middle layer as extra supervision, but none seemed to really tackle the problem once and for all.The core idea of ResNet [12] is introducing a so-called ”identity shortcut connection” that skips one or more layers, as shown in the following figure: The authors of this,argue that stacking layers shouldn‘t degrade the network performance, because we could simply stack identity mappings (layer that doesn‘t do anything) upon the current network, and the resulting architecture would perform the same. This indicates that the deeper model should not produce a training error higher than its shallower counterparts. They hypothesize that letting the stacked layers fit a residual mapping is easier than letting them directly fit the desired underlaying mapping. And the residual block above explicitly allows it to do precisely that. As a matter of fact, ResNet [12] was not the first to make use of shortcut connections, Highway Network [14]introduced gated shortcut connections. These parameterized gates control how much information is allowed to flow across the shortcut. Similar idea can be found in the Long Term Short Memory (LSTM) [15]cell, in which there is a parameterized forget gate that controls how much information will flow to the next time step.Therefore, ResNet [12]can be thought of as a special case of Highway Network [14]. However, experiments show that Highway Network [14] performs no better than ResNet[12], which is kind of strange because the solution space of High- way Network [14] contains ResNet[12], therefore it should perform at least as good as ResNet[12]. This suggests that it is more important to keep these ”gradient highways” clear than to go for larger solution space. which is kind of strange because the solution space of High- way Network [14] contains ResNet[12], therefore it should perform at least as good as ResNet[12]. This suggests that it is more important to keep these ”gradient highways” clear than to go for larger solution space.which is kind of strange because the solution space of High- way Network [14] contains ResNet[12], therefore it should perform at least as good as ResNet[12]. This suggests that it is more important to keep these ”gradient highways” clear than to go for larger solution space.

B. Localization:

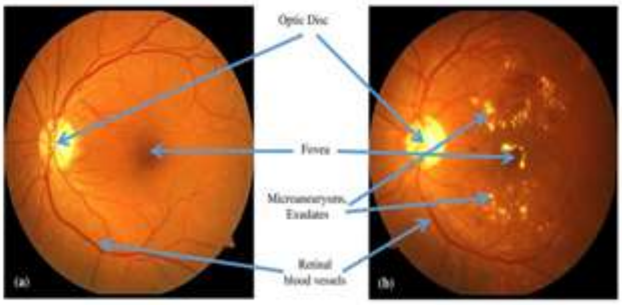

Fig 3. Eye Defects

YOLOv4-tiny [16] is based on YOLOv4 [17], which is based on YOLO [18]. Process:It starts by the model sizing the image into grids of S*S size to create a probability on an area of cells. If the centre of a probable object happens to falls into one of the cells, a preliminary bounding box is produced based on the prediction probability given by the trained model in.

P r(Object) = 0, 1 (1)

It then predicts a 3D tensor using Equation(2), using K various scaled boxes, where C is the defined number of classes, 1^{st} is confidence of prediction for each box,2^{nd} as the four t_x,t_y,t_w,t_h bounding box prediction coordinates.

S *S *(K *(4+1+C)) (2)

As given from Fig 5. , c_x&c_y are the offsets from the cluster centroid, according to the bounding box prediction based on height p_h and width p_w When the cell offset from the upper left by (c_x, c_y) and the bounding box has values of p_w and p_h, then the prediction corresponds to:

$$b_x = \sigma(t_x) + c_x \qquad (3)$$

$$b_y = \sigma(t_y) + c_y \qquad (4)$$

$$b_w = p_w e^{t_h} \quad (5)$$

$$b_h = p_h e^{t_h} \quad (6)$$

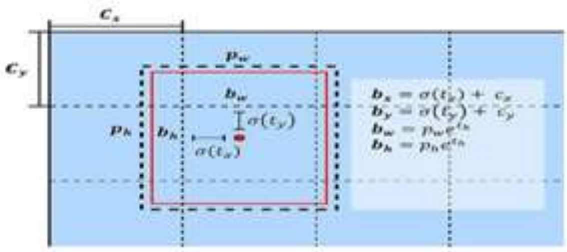

Fig 4. Bounding Box

IV. RESULTS

For Intersection over Union (IoU)and the mAP, the selected threshold is k = 0.5.Average Precision (AP) is used to determine the overall detection parameter of models instead of accuracy. It measures for a particular class instance, the number of incorrectly and correctly classified samples.Here P(k) refers to the precision at a specifically given threshold k, and Δr(k) as the shift in the Recall (RE). AP is given by:

$$AP = \frac{1}{N} \sum_{k=1}^{N} P(k)\Delta r(k) \quad (7)$$

(9)

For each category mAP computes the mean of all Average Precision.mAP usage can take part in answering which model worked best to detect lesions. mAP is given by:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (8)$$

The Intersection over Union (IoU) gives the overlap between two bounding boxes. IoU is given by:

$$IoU = \frac{AreaofIntersection}{AreaofUnion}$$

1. ACCURACY FOR THE MODELS

Model	**Microaneurysms**	**Hard Exudates**	**Soft Exudates**	**Haemorrhages**
VGG-NET(mAP)	87.1%	76.2%	78.4%	72.1%
RES-NET(mAP)	72.2%	81.1%	77.12%	75.4%

2. ACCURACY FOR THE MODELS

Model	**Optic Disc**	**Fovea**
YOLOv4Tiny(IoU)	71.22%	74.14%

1. COMPARISONS FOR LOCALIZATION ALGORITHMS

Algorithms	Results(Optic Disc)	Results(Fovea)
YOLOv4Tiny(IoU)	71.22%	74.14%
YOLO(IoU)	70.52%	69.5%
FasterR-CNN(IoU)	65.22%	70.14%

DOI: 10.35629/5252-0310707712

V. CONCLUSION

This paper showed the efficiency of various models to detect Microaneurysms, Hard Exudates, Soft Exudates, Haem- orrhages and Optic Discs in eyes using retinal fundus images. Several data pre-processing methods included the min-max normalization of pixel contrast, and generation of training labels for the optic disc and fovea coordinates. The VGG- NET-16 and RES-NET50 models also used the pre-learned weights from COCO[10] through transfer learning and the newly initialized feature sets generated by the extractor from the dataset. The segmentation was using VGG-NET16 [11] and RES-NET50 [12] models which gave different accuracy (mAP scores) for different diseases as given in the table.This work concludes that object detection models pre-trained and fine-tuned like the YOLOv4-Tiny [16] can efficiently diagnose retrieval fundus images. It has shown that YOLOv4Tiny [16] is the best among the tested localisation methods to detect Optic Discs and Fovea. Compared to classification methods, this work localized the diseases(optic disc and fovea) from the images and classified it. Unlike segmentation methods, the proposed work can run on most platforms due to the relatively small space requirement and low computational cost. Moreover, compared to existing works that employed bounding box detection methods for Microaneurysms, Hard Exudates, Soft Exudates, Haemorrhages and Optic Discs, this work prevailed as the most precise.[Table 3] .

REFERENCES

[1]. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei- Fei: ”ImageNet: A large-scale hierarchical image database”.IEEE Conference on Computer Vision and Pattern Recognition, 2009. 0.1109/CVPR.2009.5206848.

[2]. B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik: ”Simultaneous detection and segmentation”. European Conference on Computer Vision(2016).

[3]. S. Gidaris and N. Komodakis :”Object detection via a multiregion & se- mantic segmentation-aware CNN model” IEEE International Conference on Computer Vision (ICCV) 2015 10.1109/ICCV.2015.135

[4]. P. N. Druzhkov & V. D. Kustikova:”A survey of deep learning methods and software tools for image classification and object detection”.Pattern Recognit. Image Anal,2016. 10.1134/S1054661816010065

[5]. Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Fabrice Meriaudeau:”Indian Diabetic Retinopathy Image Dataset (IDRiD)”.IEEE Dataport , 2018. https://dx.doi.org/10.21227/H25W98.

[6]. Kaur, Jaskirat Mittal, Deepti. (2018). A generalized method for the segmentation of exudates from pathological retinal fun- dus images. Biocybernetics and Biomedical Engineering. 38. 27-53. 10.1016/j.bbe.2017.10.003.

[7]. Sopharak A, Uyyanonvara B, Barman S. Automatic Exudate Detec- tion from Non-dilated Diabetic Retinopathy Retinal Images Using Fuzzy C-means Clustering. Sensors (Basel). 2009;9(3):2148-61. doi: 10.3390/s90302148. Epub 2009 Mar 24. PMID: 22574005; PMCID: PMC3332251.

[8]. Gharaibeh, Nasr Al-Smadi, Ma‘moun Al-Jarrah, Mohammad. (2014). Automatic Exudate Detection Using Eye Fundus Image Analysis Due to Diabetic Retinopathy. Computer and Information Science. 7. 48-55. Benes R., Hasmanda M., & Riha K.:”Object localization in medical image”.34th International Conference on Telecommunications and Signal Processing (TSP),2011. 10.1109/tsp.2011.6043667

[9]. Benes R., Hasmanda M., & Riha K.:”Object localization in medical image”.34th International Conference on Telecommunications and Signal Processing (TSP),2011. 10.1109/tsp.2011.6043667

[10]. T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Zitnick, and P. Dolla ´r., Ross Girshick:”Microsoft COCO: Common Objects in Context”.European Conference on Computer Vision,2014. 10.1007/978-3-319-10602-148.

[11]. Karen Simonyan & Andrew Zisserman:”VERY DEEP CONVO-LUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOG-NITION”.International Conference on Learning Representations,2015. 10.1109/cvpr.2016.90

[12]. He K., Zhang, X., Ren S., & Sun J.:”Deep Residual Learning for Image Recognition”.IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016. 10.1109/cvpr.2016.90

[13]. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir

Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich:”Going Deeper with Convolutions”.IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2015. 10.1109/CVPR.2015.7298594

[14]. Rupesh Kumar Srivastava, Klaus Greff, Ju ̈rgen Schmidhuber:”Highway Networks”.ICML 2015 Deep Learning workshop,2015. arXiv:1505.00387

[15]. Hochreiter, Sepp and Schmidhuber, Ju ̈rgen:”Long short-term memory”.Neural computation.1997

[16]. Jiang, Zicong et al. ”Real-time object detection method based on improved YOLOv4-tiny.” ArXiv abs/2011.04244 (2020): n. pag.

[17]. Bochkovskiy, Alexey Wang, Chien-Yao Liao, Hong-yuan. (2020). ”YOLOv4: Optimal Speed and Accuracy of Object Detection”

[18]. Redmon, Joseph Divvala, Santosh Girshick, Ross Farhadi, Ali. (2016). ”You Only Look Once: Unified, Real-Time Object Detection”. 779-788. 10.1109/CVPR.2016.91.

[19]. Szegedy, Christian Toshev, Alexander Erhan, Dumitru. (2013). ”Deep Neural Networks for Object Detection”. 1-9.

[20]. Khojasteh, Parham Aparecido, Leandro Passos Ju ́nior, Leandro Carvalho, Tiago Rezende, Edmar Aliahmad, Behzad Papa, Joa ̃o Kumar, Dinesh. ”Exudate Detection in Fundus Images Using Deeply-learnable Features”. Computers in Biology and Medicine, 2018