IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Image Segmentation and Localization in Retinal Fundus Images

line 1: 1st Soham Deshmukh
line 2: India

***Abstract*—Computer-aided disease diagnosis in retinal image analysis could provide a sustainable approach for such large-scale screening efforts.The recent scientific advances in computing capacity and machine learning approaches provide an avenue to reach this goal.The ocular pathologies 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 than other retinal components or other pathological lesions. Therefore, ocular diseases' diagnosis seems to be a difficult task, that requires taking into account several parameters, and hence Deep Learning represents an adequate approach to resolve such problems.The aim of this project was to identify which algorithms were more suitable in order to classify different diseases of the human eye namely Microaneurysms, Hard Exudates, Soft Exudates, Haemorrhages and Optic Discs. Given the dataset there are three major sub-parts for the project which are (1)Segmentation, (2)Disease Grading and (3)Localization.**

Keywords—Deep Learning, Object Detection, Neural Network, Segmentation.

I. INTRODUCTION

Object Recognition has an important role in image processing and Computer vision field. It is the process of determining the identity of an object being observed in an image or a video sequence from a set of known tags with the help of a recognition technique. Breakthroughs in image classification started 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. They achieved more accurate prediction than the previous state of the art models. From this, many researchers became interested in finding a novel way to develop an efficient deep convolutional neural network [2] [3] [4]. 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 [6]. The segmentation of the said four diseases [7](Microaneurysms,Hard Exudates, Soft Exudates, Haemorrhages and Optic Discs) were done using algorithms VGG-NET16 and RESNET50.

A. VGG-NET:

The input to VGG [8] based convNet is a 224*224 RGB image. Preprocessing layer takes the RGB image with pixel values in the range of 0–255 and subtracts the mean image values which is 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 architecture. VGG has smaller filters (3*3) with more depth instead of having large filters. It has ended up having the same effective receptive field as if you only have one 7 x 7 convolutional layers.Another variation of VGGNet 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 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.

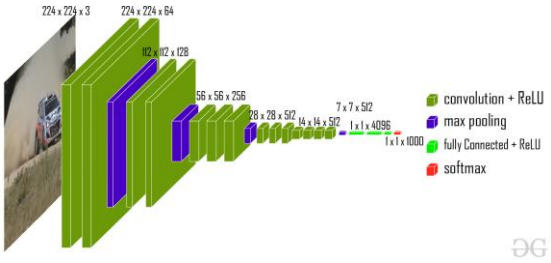


Fig. 1. VGGNET Architecture

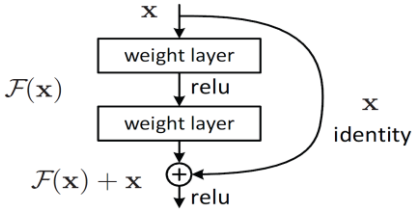
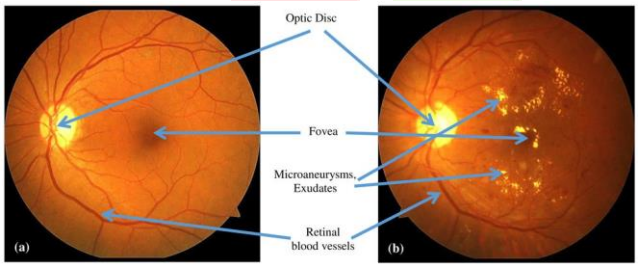
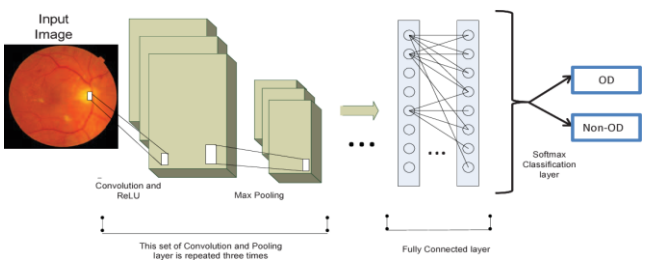


Fig. 2. RESNET

B.RESNET:

Before ResNet,[9] there had been several ways to deal the vanishing gradient issue, for instance, GoogleNet [10] (also codenamed Inceptionv1) adds an auxiliary 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 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 was not the first to make use of shortcut connections, Highway Network [11]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) [12]cell, in which there is a parameterized forget gate that controls how much information will flow to the next time step. Therefore, ResNet [9]can be thought of as a special case of Highway Network. However, experiments show that Highway Network performs no better than ResNet[9], which is kind of strange because the solution space of Highway Network contains ResNet[9], therefore it should perform at least as good as ResNet[9]. This suggests that it is more important to keep these "gradient highways" clear than to go for larger solution space. Following this intuition, the authors of [2] refined the residual block and proposed a pre-activation variant of residual block [7], in which the gradients can flow through the shortcut connections to any other earlier layer unimpededly. In fact, using the original residual block in [2], training a 1202-layer ResNet[9] resulted in worse performance than its 110-layer counterpart. All the three sub parts of the paper are done using algorithms pertaining to three different topics of image processing: Image Segmentation, Disease grading (when given train-test images) and Image Localization.The algorithm options for the first sub part were VGG-NET,U-NET [13], Mask R-CNN [14] from which VGG-NET and RES-NET was chosen and YOLO Tiny-v4 for the third part.





II. PROPOSED WORK

Fig. 3. Eye Defects
Fig. 4. Architecture

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 [15] dataset to improve the model performance to detect several brain diseases. The previously learned COCO [15] 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, as the default number of 80 corresponds to the previous classes from COCO. With the newly defined classsize, every Conv filters must also shift from the default 255 to 24, where C corresponds to the number of classes, five as the YOLO [16]coordinates, and three as the various scaled bounding boxes K. The Detection Approach of YOLO: This section briefly explains the detection process of the YOLO-based model. The process begins with the model interpreting an image using logical S*S grids and the weighted feature sets to create a probability on an area of cells. If the center of a probable object falls to one of the cells, a preliminary bounding box is produced based on the prediction probability given by the trained model in.

$$Pr(Object) = 0,1 \quad (1)$$

The model then predicts with the use of K various scaled boxes and extracts a 3D tensor based on (2), where C represents the defined number of classes, four as the t_x, t_y, t_w, t_h bounding box prediction coordinates, and one as the confidence of prediction for each bounding box.

$$S * S * (K * (4 + 1 + C)) \quad (2)$$

In Fig. , the bounding box prediction based on the width p_w and height p_h had offsets c_x and c_y from the cluster centroid. 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:

$$b_x = \sigma(t_x) + c_x \quad (3)$$
$$b_y = \sigma(t_y) + c_y \quad (4)$$
$$b_w = p_w e^{t_h} \quad (5)$$
$$b_h = p_h e^{t_h} \quad (6)$$

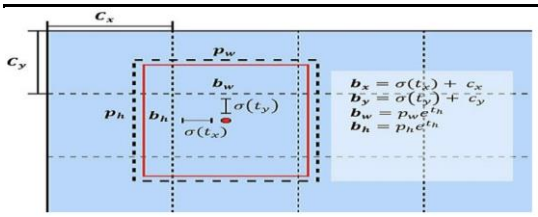


Fig.5. Bounding Box for YOLO

A. *Evaluation Metrics*

This work selected a threshold k of 0.5 to evaluate the Intersection over Union (IoU)and the mAP.In a global standard, Average Precision (AP) is the metric used to determine the overall detection prowess of object detection models rather than accuracy. This metric pertains to the number of correctly and incorrectly classified samples of a specific class instance. Where the P(k) refers to the precision at a specifically given threshold k, and Δr(k) as the shift in the Recall (RE). The following equation formally presents the AP.

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

The mAP calculates the mean of all AP for each category. Using the mAP as the primary key indicator can justify a model that worked best overall to detect brain tumors specifically.The following equation formally presents the mathematical equation for the mAP.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{8}$$

The Intersection over Union (IoU) determines the overlap between two bounding boxes.The following equation calculates the IoU by having the intersection area divided by the area of union. Microaneurysms, Hard Exudates, Soft Exudates, Haemorrhages and Optic Discs.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{9}$$

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

3. 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

III. CONCLUSION

This work presented the efficiency of employing various models to detect Microaneurysms, Hard Exudates, Soft Exudates, Haemorrhages 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 [15] through transfer learning and the newly initialized feature sets generated by the extractor from the dataset. VGG-NET16 and RES-NET50 models gave different accuracy (mAP scores) for different diseases as given in the table.VGGNet not only has a higher number of parameters and (Floating Point Operations)FLOP as compared to ResNet-152 but also has a decreased accuracy. It takes more time to train a VGGNet with reduced accuracy.This work concludes that object detection models pre-trained and fine-tuned like the YOLOv4-Tiny can efficiently diagnose retrieval fundus images. 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]

IV. 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 & semantic 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. ImageAnal,2016. 10.1134/S1054661816010065.

[6] 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.

[7] 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

[8] Karen Simonyan & Andrew Zisserman:"VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION".International Conference on Learning Representations,2015. 10.1109/cvpr.2016.90

[9] 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.

[10] 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

[11] Rupesh Kumar Srivastava, Klaus Greff, Jurgen Schmidhuber:"Highway Networks".ICML 2015 Deep Learning workshop,2015. arXiv:1505.00387

[12] Hochreiter, Sepp and Schmidhuber, Jurgen:"Long short-term memory".Neural computation.1997

[13] Olaf Ronneberger, Philipp Fischer, Thomas Brox.:"U-Net: Convolutional Networks for Biomedical Image Segmentations".International Conference on Medical Image Computing and Computer-Assisted Intervention,2015. 10.1007/978-3-319-24574-4

[14] Kaiming He; Georgia Gkioxari; Piotr Dollar; Ross Girshick:"Mask R-CNN".IEEE International Conference on Computer Vision (ICCV),2017. 10.1109/ICCV.2017.322

[15] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Zitnick, and P. Dollar., Ross Girshick:"Microsoft COCO: Common Objects in Context".European Conference on Computer Vision,2014. 10.1007/978-3-319-10602-1

[16] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao :"YOLOv4: Optimal Speed and Accuracy of Object Detection" arXiv:2004.1093.