Parasite Detection from Digital Images Using Deep Learning

Yulin Zhu and Wei Qi Yan
Auckland University of Technology, 1010 New Zealand

ABSTRACT

Parasitosis is a disease caused by parasites that could infect humans, animals, or plants. The parasites include mites, ascariasis, liver flukes, and malaria. The methods to detect parasites include pathological examination, immunological examination and imaging examination. In this paper, parasitic infections are detected from digital images acquired from a microscope, which will look for the possible infection caused by worms or eggs in a sample, such as mites and malaria. Rapid and accurate classification and detection of parasites will be very helpful for fast diagnosis and treatment. In this paper, a malaria detection method is deployed by using deep learning based on TensorFlow and achieved 0.73 mAP@0.5IOU. Even if it does not seem to be a perfect performance, in the limited time and resources, the results are still valuable. The future work could port the model to mobile phones for image detection, which would bring much more convenience and portability.

Keywords: Object Detection, Malaria Detection, TensorFlow

INTRODUCTION

Along with the popularization and development of computer vision, visual object recognition has been applied in intelligent monitoring, car plate recognition, animal and plant recognition in many aspects, and the medical field is also a hot topic for discussion. For example, during the period of COVID-19, Alibaba's team deployed image analysis technology to assist in the diagnosis of CT image of patients, and achieved satisfactory outcomes. Therefore, a fast and simple image recognition method is expected, which is able to automatically analyse digital images taken from the microscope so as to identify possible malaria infections with a well-trained object detection model. The parasite detection is expected faster and easier than ever before, thus it can serve our community better.

Malaria is able to be identified by microscopic blood tests. Usually, these tests require skilled professionals to observe and look for suspected worm or eggs under a microscope, the accuracy of results usually depends on professional laboratories or the professionalism and experience of the laboratory staff. In the leakage of professionals and equipment, diagnostic accuracy could be reduced or omitted, it could result to unpredictable consequences for patients.

Visual object recognition is the core issue in computer vision. Visual object detection is dependent on the feature expression in digital image processing. Lecun et al. proposed a convolutional neural network (CNN) in 1989 and successfully applied it to image recognition (LeCun, et al. 1989). CNN network has achieved remarkable progress in the field of object recognition, the model based on CNN in the ImageNet LSVRC-2010 has achieved the best error rate of 15.3% (Krizhevsky, Sutskever & Hinton, 2012).

CNN network is effectively classified images, because its network structure is highly optimized for 2D and 3D images, which could effectively be trained by using 2D features. The CNN includes two parts which are extractor and classifier. Each layer in the network takes use of a differentiable function to extract and transfer data from present layer to the next. The CNN net is mainly composed of three types of layers: Convolution, max pooling, and classification. By superimposing these layers, a complete CNN model will be constructed (Alom, et al., 2019). The process of object detection is to extract the feature maps through convolutions, the extracted features from the convolutional layer will reduce the problem of overfitting (Lu & Zhang, 2016).

The activation function is like a simulation of how human's brain works. A neuron receives an electrical signal and fires electrical signal. If the incoming electrical signal is not strong enough, the neuron will not react at all; but if the electrical signal is stronger than a threshold, then the neuron will react and send electrical signals to other neurons. Multilayer neural networks are usually a highly nonlinear model, the rectification usually introduces a nonlinear function (also known as activation function), which is applied to the convolution layer. Nonlinear functions mainly include logistic function, Tanh function, sigmoid function and ReLU, etc. (Hadji & Wildes, 2018). Sigmoid is a nonlinear activation function, its mathematical form is

$$Sigmoid:\ f(x) = \frac{1}{1+e^{-x}}. \tag{1}$$

It transforms the continuous value of input into the output between 0 and 1.

The tanh function is

$$tanh:\ f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}\,. \tag{2}$$

The ReLu function is

$$ReLU:\ f(x) = \begin{cases} x, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases} \tag{3}$$

or

$$f(x) = \max\,(0, x)\,. \tag{4}$$

Visual object detection is complex, because a single image usually contains multiple objects with different classes. Deep learning needs to classify and locate these targets (Lu & Zhang, 2016). The goal of object detection is to predict bounding boxes for all possible targets in an image. It is a challenge to identify the size of visual objects, especially the smaller targets. In particular, small size targets have blurred boundaries, which will lead to targets being mixed with background or noise (Li, et al., 2017). The method in object detection usually is to search and select the possible target candidate regions and then send them to a classifier.

For example, exhaustive search is to constantly change window sizes, and then scan the whole image to determine the location of the target. The selective search by segmenting the image and extracting the texture, colour, size and other features of the image will reduce the complexity of the search, and improve the search speed (Uijlings, Van De Sande, Gevers & Smeulders, 2013).

This method is applied to select the possible region that could contain the target, also called region of interest (ROI). The region CNN net selects 2,000 possible ROI bounding boxes and takes use of SVM for classification (Girshick, Donahue, Darrell & Malik, 2014).

Visual objects in an image have various shapes, scales, colours, and textures. It is very difficult to identify an object from the given image (Nair & Hinton, 2010). As we know, whilst making data prediction, the data of the training set should be cleaned to remove incomplete or redundant information, so as to improve the accuracy of the model. Similarly, in the process of object recognition, the image preprocessing is also important. The purpose of image preprocessing is to remove the useless information in the image such as noises and enhance the useful information, so as to facilitate the feature extractor to obtain the image features and improve the accuracy of image recognition (Zhang, Wu & Chang, 2020).

Image features usually encapsulate colour, texture, and shape. Texture features include orientation, contrast, and roughness. Shapes also encompass local features, such as edges and corner points. The resolution of images, the intensity of the light, the angle of the object, the shooting distance and the boundary between the objects and the background could affect the accuracy of image recognition. Therefore, the possible methods for image augmentation include geometric transformation and image enhancement.

In this book chapter, following related work, the proposed methods will be iterated. The result analysis will be explicated which leads to the final conclusion of this book chapter.

RELATED WORK

There are already examples of malaria parasite detection by using deep learning algorithms. Zhang has harnessed the feature screening method and improved KNN classifier to classify the parasitic eggs. The average accuracy has achieved 90% (Zhang, 2017). Ross, et al. pointed out that even though there are other non-microscopic methods of detecting malaria parasites, such as polymerase chain reaction (PCR), microscopic examination is still the most widely and recommended method. They have applied image preprocessing to reduce image noises. This study achieved positive predictive value (PPV) 81%.

In image segmentation, the classes of grey-level histograms were adopted to automatically select thresholds so as to separate the erythrocytes from the background (Ross, Pritchard, Rubin, Dusé & Dusé, 2006). IOU refers to intersection over union, which calculates the ratio of the intersection and union of the predicted border and the true border (Rezatofighi, et al. 2019), which needs to calculate precision and recall when eval an model.

$$IoU = \frac{Predicted\ border \cap True\ border}{Predicted\ border \cup True\ border} \tag{5}$$

$$Precision = \frac{True\ Positives}{True\ positives \cup False\ positives} \tag{6}$$

$$Recall = \frac{True\ Positives}{True\ positives \cup False\ negatives} \tag{7}$$

Using mobile phones to detect parasite is also a valuable research area. Smartphones are already ubiquitous, mobile Apps are easily to be distributed which are able to quickly upgraded. If mobile phones are employed for parasite detection, it will be possible to more quickly generalization the parasite detection methods around the world, improve the efficiency and accuracy, and thus reduce the dependence of professionals and reduce the cost. Dallet et al. demonstrated how to develop and Android App that is use of an algorithm that implemented in MATLAB environment.

An approach has been implemented to detect the centroid every red blood cell (RBC) in the image using Annular Ring Ratio (ARR) (Dallet, Kareem & Kale, 2014). A special device that is able to take a blood sample using a smartphone was recommended. Although today's smart phones have the fast computing ability and more memory than ever before, but we still need to consider how to train a lightweight model if we expect developing an object detection App on smartphones.

The model training for object classification requires high computing power. In many cases, the use of GPU with excellent performance could significantly improve the speed of training and recognition. There are a slew of platforms that provide the ability to use both CPU and GPU for computing, such as TensorFlow and Keras (Gulli & Pal, 2017). Keras could receive an unlimited number of channels, biomedical experiments often involve several dozens of fluorescent labels in multispectral imaging (Hung, et al. 2020).

For visual object detection, multiple meta architectures and feature extractors are compared. Different variables were also applied to test the model, such as image size, target size, and the number of proposals (Huang, et al. 2018).

THE METHODOLOGY

In this paper, we select a lightweight model and train an image detector to mark out the plasmodium from digital images. The reason is related to the time limitation and equipment performance, we would like to keep a possible to deploy the model to a smartphone.

The resources in this study are laptop equipped with Intel i7 6700 CPU, Nvidia GTX960M GPU. Operating system based on Ubuntu 20.04 LTS, installed Anaconda 3 with Python 3.6, TensorFlow 2.2.0 which already contain Keras, and installed Object Detection API. The CUDA Drive and cuDNN has been installed to accelerate the training.

Visual object detection usually has two directions, single-stage detection and two-stage detection. Single-stage target detection is faster than two-stage target detection, but the accuracy will be reduced. There are two methods in single level target detection: SSD and YOLO. SSD (Single Shot MultiBox Detector) is a one-stage general object detection algorithm, Single Shot indicates that SSD algorithm belongs to one-stage method, MultiBox indicates that SSD is multibox prediction. SSD adopts pyramid structure, it will use feature maps from every convolution layer, which improved the ability to detect different objects. Feature maps of different objects are adopted in multiple scales.

In CNN network, the feature map is generally large, the convolution or pooling layer is gradually adopted to reduce the size of the feature map. Both a larger feature map and a smaller feature map are employed for object detection, the advantage is that the larger feature map is utilized for detection of relatively small targets, while the smaller feature map is harnessed for detection of large objects. The dataset contains 1,182 thick blood smear images with bounding boxes of 7,245 parasites (Quinn, et al., 2016).

In the image processing, both the original and enhanced datasets are randomly split 70% as the training set and 30% as the test set, the file names are same for both datasets. This is to compare whether the sample with image enhancement is more easily identified than the original sample. We process he images and label file to TFRecord files. In addition, it is helpful to explore whether pre-trained model could accelerate the model convergence (He, Girshick & Dollár, 2019).

For the model training and evaluations, RGB images are resized to 300×300 pixels, the batch size has been set to 18, warmup_learning_rate is assigned to 0.005, warmup_steps are given as 100, learning_rate_base is taken 0.025, total_steps are shown as 3000, and max_number_of_boxes are assigned as 100.

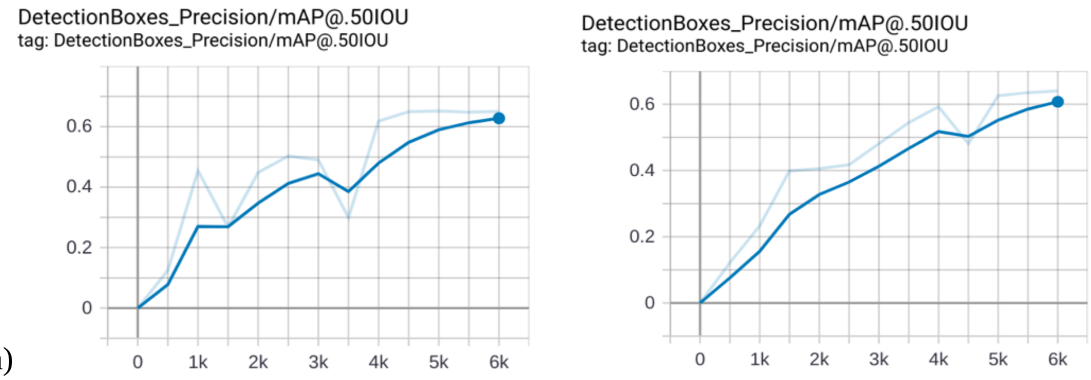
Comparing with pre-trained model, we take use of model trained on COCO 2017 dataset using SSD MobileNet v2. VGG-16 network is composed of 13 convolutional layers + 3 full connection layers. In this case, it will have 6 layers to generate feature map, each point in every feature map is applied to structured default box, 8732 detection boxes will be generated for each image. SSD is implemented from every feature map that will keep shallow feature map which contains more detailed information, and is more suitable for different size of object detection, especially small object detection.

Based on the performance, we need to resize the image to 300×300 pixels and batch size 20. We take advantage of ReLu as active function, which could solve vanishing gradient problem. ReLU6 is ordinary ReLU but is limited to the maximum output value as 6, this is for the low precision of the mobile Float16, also can have very good numerical resolution. If the scope of the activation of ReLU has not limit, the output range is 0 to infinite; if the activation value is very large, distributed in a large range, the low precision of Float16 cannot accurately describe the well such a large range of values that bring accuracy loss.

RESULT ANALYSIS

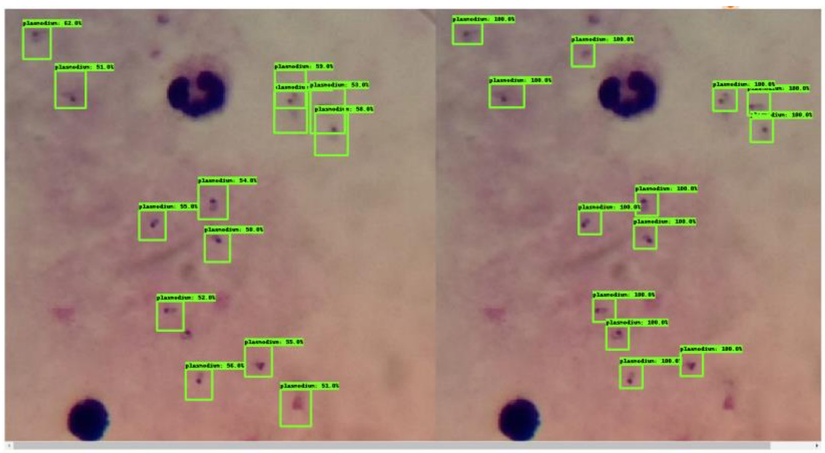
From the training of automatic malaria detector, we came up that automated detector is able to save time, and the accuracy are acceptable. As shown in Figure 1, the shadow line is the actual value, blue line shows the trend.

Whilst comparing the same model with enhanced images, the accuracy dropped significantly. As image enhancement, the idea is to make the edges mpre clearer. But from the result, the default box is larger than except. This was caused by the image sharpening made with less noises. We could see the image, the localization loss does not drop.



(a) (b)

Figure 1: Comparison between (a) SSD_MobileNet_V1 and (b) SSD_MobileNet_V2



(a) (b)

Fig. 2. Samples of parasite detection (a) is the detection results, (b) is grand truth)

CONCLUSION

In this paper, the lightweight model is treated as the proposed solution for parasite detection. The model achieved 0.73 mAP@0.5IOU. Rapid and efficient image recognition could assist laboratory staff to improve their working efficiency, test more samples, and reduce the possibility of missing any parasite objects.

The resolution of this dataset is not particularly high, the boundaries of malaria objects are often blurred and blended with the background. Due to lack of professionals, the model train was processed at the resolution 300×300, which is less than half of the original image. The bounding box of malaria samples in original image has the resolution 40×40. These conditions could be a problem if SDD is employed for classification, even if the algorithm is very fast. But what if the dataset not only contains just one class of parasite, in real laboratory testing environment, the majority samples will definitely include different parasite with different shape, size, and color. The algorithm needs to adapt to different size of samples, but also needs to ensure the picture is relatively clear.

On another hand, we need to consider that the parasites to be detected in the future are not only malaria. Meanwhile, the advantage of SSD algorithm is the rapid detection with various sizes and types, so different sizes and shapes of parasite samples should be introduced into the experiment.

More materials could assist model training for the purpose to get more accurate results. The pretrained model could speed up the training process and improve the precision. Whilst using pretrained model, the result is better.

REFERENCES

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan. M., Essen. B., Awwal. A. & Asari, V. K. (2019) A state-of-the-art survey on deep learning theory and architectures. Electronics, 8(3), 292

Dallet, C., Kareem, S., & Kale, I. (2014) Real time blood image processing application for malaria diagnosis using mobile phones. In: IEEE International Symposium on Circuits and Systems (ISCAS), 2405–2408

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587

Gulli, A., & Pal, S.: Deep learning with Keras: Implement neural networks with Keras on Theano and TensorFlow. Pack (2017)

Hadji, I., & Wildes, R. P. (2018) What do we understand about convolutional networks? arXiv:1803.08834

He, K., Girshick, R., & Dollár, P. (2019) Rethinking ImageNet pre-training. In: IEEE International Conference on Computer Vision, pp. 4918-4927

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., & Le, Q. V. (2019) Searching for MobileNetv3. In: IEEE International Conference on Computer Vision, pp. 1314-1324

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., & Murphy, K. (2017) Speed/accuracy trade-offs for modern convolutional object detectors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3296–3297

Hung, J., Goodman, A., Ravel, D., Lopes, S. C. P., Rangel, G. W., Nery, O. A., Malleret, B., Nosten, F., Lacerda, M. V. G., Ferreira, M. U., Renia, L., Duraisingh, M. T., Costa, F. T. M., Marti, M., & Carpenter, A. E. (2020) Keras R-CNN library for cell detection in biological images using deep neural networks. BMC Bioinformatics, 21(1)

Ji, H., Liu, Z., Yan, W., Klette, R. (2019) Early diagnosis of Alzheimer's disease based on selective kernel network with spatial attention. Asian Conference on Pattern Recognition

Ji, H., Yan, W., Klette, R. (2019) Early diagnosis of Alzheimer’s disease using deep learning. ACM ICCCV

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012) ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25(2)

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989) Backpropagation applied to handwritten zip code recognition. Neural Computation, 1(4), 541-551

Li, J. (2013) Study on image Feature Extraction and Recognition technology of human parasitic worm eggs. Inner Mongolia Traditional Chinese Medicine, 32(31), 121-121

Li, J., Liang, X., Shen, S., Xu, T., Feng, J., & Yan, S. (2017) Scale-aware Fast R-CNN for pedestrian detection. IEEE Transactions on Multimedia, 20(4), 985-996

Lu, H & Zhang, Q. (2016) Applications of deep convolutional neural network in computer vision. Journal of Data Acquisition and Processing, 31(1), 1-17

Nair, V., & Hinton, G. (2010) Rectified linear units improve restricted Boltzmann machines. In: International Conference on Machine Learning

Quinn, J. A., Nakasi, R., Mugagga, P. K., Byanyima, P., Lubega, W., & Andama, A. (2016) Deep convolutional neural networks for microscopy-based point of care diagnostics. In: Machine Learning for Healthcare Conference, pp. 271-281

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019) Generalized intersection over union: A metric and a loss for bounding box regression. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 658-666

Ross, N., Pritchard, C., Rubin, D., Dusé, A., & Dusé, A. (2006) Automated image processing method for the diagnosis and classification of malaria on thin blood smears. Medical & Biological Engineering & Computing, 44(5), 427–436

Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013) Selective search for object recognition. International Journal of Computer Vision, 104(2), 154-171

Wang, Y. (2021) Colorizing Grayscale CT Images of Human Lung Using Deep Learning. Master’s Thesis, Auckland University of Technology, New Zealand.

Wang, Y., Yan, W. (2022) Colorising grayscale CT images of human lungs using deep learning methods. Springer Multimedia Tools and Applications.

Yan, W. (2019) Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics. Springer.

Yan, W. (2021) Computational Methods for Deep Learning: Theoretic, Practice and Applications. Springer.

Younas, F., Usman, A., Yan, W. (2022) A deep ensemble learning method for colorectal polyp classification with optimized network parameters. Applied Intelligence.

Zhang, L. (2020) Virus Identification from Digital Images Using Deep Learning. Master’s Thesis, Auckland University of Technology, New Zealand

Zhang, L., Yan, W. (2020) Deep learning methods for virus identification from digital images. International Conference on Image and Vision Computing New Zealand

Zhang, S., Wu, Y., & Chang, J. (2020). Survey of image recognition algorithms. IEEE Information Technology, Networking, Electronic and automation control conference (ITNEC), pp. 542-548

Zhang, Z. (2017) Automatic identification and classification of parasite eggs in microscopic faecal images. Master's thesis, University of Electronic Science and Technology of China