

Comparison of Multiple CNNs as Feature Extractors in Faster R-CNN for Malaria Parasite Detection

Alinani Sichone

School of Computer and Software

Nanjing University of Information Science and Technology

Nanjing, China

Abstract— Deep learning algorithms, and Convolution Neural Networks in particular, have advanced significantly in the last decade and have seen ever-increasing applications to real-life problems—medical applications being some of the most promising. CNNs have been shown to perform better than traditional image processing methods and machine learning algorithms used in medical image classification or segmentation. In this paper, we adapt the Faster Region-based Convolutional Neural Network to the purpose of malaria parasite localization and classification in Giemsa-stained thin-blood smear through the process of transfer learning. Five convolutional neural networks i.e., (ResNet-50, ResNet-101, VGG16, VGG19 and EfficientNet B3) are used as feature extractors for Faster R-CNN and their impact on the accuracy of the model is compared. We use the BBBC041v1 dataset as the source of the thin-blood smear images used in the experiments and the feature extractors are compared separately in detecting and localizing the six different classes of cells contained in the dataset. The experiment results demonstrate that ResNet-101 exhibits the best performance relative to the other four models tested as feature extractors.

Keywords— *Malaria, transfer-learning, object detection, faster r-cnn.*

I. INTRODUCTION

Malaria is a deadly yet treatable illness that affects millions of people each year. Infection with a Plasmodium parasite, of which five species (*Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi*) may infect humans, causes malaria. Severe malaria is multi-syndromic and often manifests as cerebral malaria. Mortality is high if severe malaria is not promptly and effectively managed. A majority of the severe cases occur due to infection with the *falciparum* parasite [1]. These parasites are mostly transferred to humans by the bite of female *Anopheles* mosquitoes.

According to the 2021 World Malaria Report by the World Health Organization, there was an estimate of 241 million malaria cases globally in 2020, which was an increase from 227 million in 2019. Twenty-nine countries accounted for 95% percent of the cases worldwide, and six countries for 55% of the cases. The percentage of total malaria deaths in children under the age of 5 years was 77%. Most of the cases were from the African continent [2].

When malaria is detected early on, it may be efficiently treated with medicine, avoiding a mild case from becoming life-threatening. For adequate treatment and illness

monitoring, an accurate malaria diagnosis is required. Due to the constraint of resources and remoteness, access to diagnostic procedures is restricted in places where the malaria load is highest [3].

Clinical diagnosis of malaria (based on symptoms such as fever rather than a diagnostic test) uses the fewest resources and is hence still extensively used. Malaria symptoms, on the other hand, are variable and overlap with those of many other prevalent tropical illnesses, resulting in low specificity in the clinical diagnosis. False positives are widespread in highly endemic areas, resulting in not only the underlying source of symptoms remaining untreated, but also over-prescription of antimalarial medications, which causes unwanted side effects and contributes to parasite drug resistance [4]–[6].

As a result, the WHO recommends that all suspected cases of malaria be verified with a parasitological test. The percentage of instances when this occurs has risen considerably in Sub-Saharan African nations over the last decade, from 38% of all suspected cases in 2010 to 85% in 2018, owing to the advent of Rapid Diagnostic Tests (RDTs). Manual inspection of blood smears by light microscopy is the current gold-standard method for correctly identifying malaria parasites. The standard microscopic diagnostic procedure recommended by the WHO follows four steps: blood film preparation, staining, examination, and interpretation [7].

This manual inspection is slow and requires expertise; which, in resource-strained countries, can be costly and scarce. Therefore, automating the process of counting and classification of malaria-parasitized cells would be beneficial.

Our work seeks to apply recent advances in Computer Vision to the problem to the problem of malaria parasite localization and classification in thin blood smear images. We seek to implement a robust object detection model that can surmount the challenges presented in examining malaria microscopy images such as the differences in illumination, color, cell shape and density, and the insufficiency of annotated well-balanced training data.

II. METHODS

A. Conventional Malaria Image Analysis Methods

A lot of research has been proposed that has sought to leverage image processing technologies for the automation of malaria diagnosis. Many proposed methods that use traditional image processing techniques for the classification of thin

blood smear images aim to automatically count all uninfected and parasitized cells. They typically follow a process that includes preprocessing, segmentation, feature extraction, and classification [8], [9].

Figure 1 depicts a schematic representation of the typical pipeline for automated malaria image analysis using traditional image processing techniques.

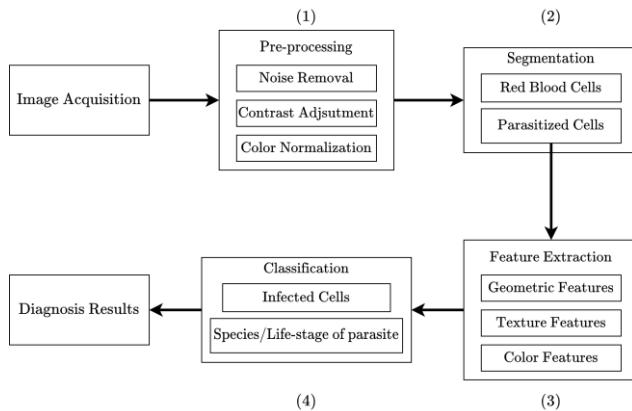


Fig. 1. Illustration depicting a schematic representation of the basic image analysis pipeline followed by most automated malaria diagnosis techniques.

B. Proposed Method

One of the main areas of research interest in Computer vision is Object Detection. In contrast to image classification, object detection not only identifies objects in an image but also localizes them. Object detection algorithms have seen increasing application in medical image processing and have shown very good results when compared to human experts.

Our method, based on Faster R-CNN [10] is shown in figure 2. We tested five different CNNs, which are ResNet50, ResNet101, VGG16, VGG19, and EfficientNet B3, as feature extractors using transfer learning. These networks have achieved excellent results in image classification tasks. The nature of Faster R-CNN allows us to use any CNN as a feature extracting backbone. We performed tests to ascertain which of the five CNNs performed best as a feature extractor in the detection of parasitized cells.

The features extracted from the CNN are shared between the Regional Proposal Network (RPN) and the Region of Interest (ROI) pooling layer. The RPN takes fixed bounding boxes, which are referred to as anchors and are placed throughout the image with different aspect ratios and sizes, and categorizes those that overlap with a ground truth object with an Intersection over union (IoU) greater than 0.5 as foreground objects and those that have an IoU less than 0.1 or have no overlap with any ground truth object as being background object. The RPN outputs a set of rectangular object proposals, each with an *objectness* score. The ROI pooling layer fixes the feature vectors of the object proposals to a uniform size and the feature vectors are finally sent to the classifier to complete the detection process.

In this paper we aim to study the performance of the above mentioned CNNs as feature extractors for Faster R-CNN in the task of malaria parasite detection.

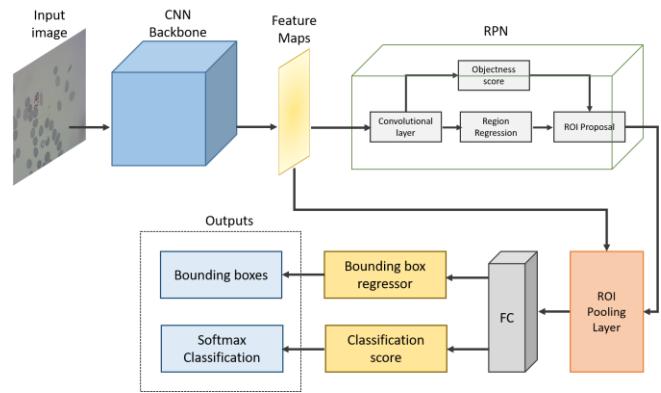


Fig. 2. Faster RCNN architecture used in this study, with the CNN Backbone being either one of the five CNNs we test.

C. Dataset

We used image set BBBC041v1, available from the Broad Bioimage Benchmark Collection [11]. This dataset consists of 1,328 images of Giemsa reagent stained thin-blood smears. The data came from ex vivo samples from *Plasmodium vivax* infected patients. For the training data, four patients were used and one was used for the testing data. There are seven classes in the dataset: red blood cells, trophozoites, rings, schizont, gametocytes, leukocytes and a “difficult” class that annotators were allowed to label as if they weren’t certain of the cell type. The object classes in the dataset have a heavy but natural imbalance, with red blood cells accounting for 97 % of the total cells in the dataset. The images are stored in .png or .jpg file formats and 24-bit color depth. There is a total of 85,985 cells in those images. 1208 of the images in the dataset are 1600×1200 pixels and the remaining 120 images are 1944×1383. 88% of the dataset is used to train the models, and the remainder to test them.

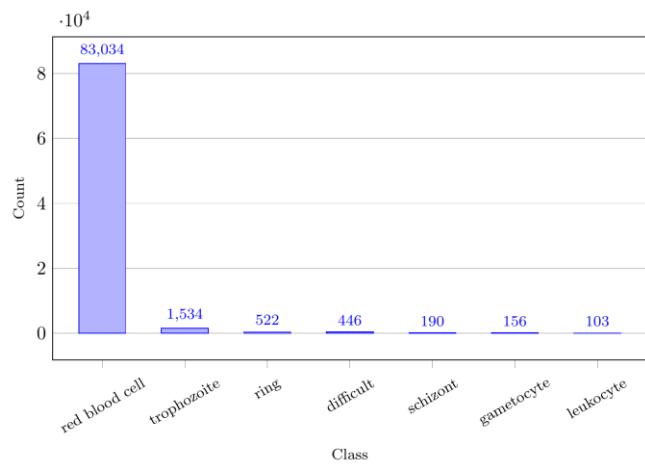


Fig. 3. Distribution of the number of cells over the seven classes.

D. Image Pre-processing

One of the most essential properties of Computer-Aided Diagnosis systems for microscopy images is robustness. However, if the systems work with multi-source images gathered under various configurations, it is difficult to assure

this property. Changes in lighting and acquisition devices modify the color of images and, in many cases, degrade the system's performance. As a result, before training and testing the models, the colors of the images in the dataset were adjusted. This normalization of colors in images is called color constancy.

The Shades of Gray algorithm [12] is used for the color transformation with a gamma correction step with a gamma set to the standard value of 2.2 [13] was applied to all the images before the color constancy was performed.

III. RESULTS

A. Parameter Settings

The models were implemented using the PyTorch framework version 1.9.1 on Python 3.7.12 and were trained on a single NVIDIA Tesla P100 with 16GB RAM in Ubuntu 20.04.

The original framework parameters of Faster RCNN are initially used in our experiments. We aim to repurpose the original object detection framework to malaria parasite detection on high-resolution blood smear images through transfer learning. We use the Albumentations library [14] to perform image augmentations during the training process. Generalized R-CNN Transform operations of normalizing the images and resizing them (bilinear mode) are also applied.

In all feature extractors, batch normalization is applied after every convolution. The parameters of the batch normalization are frozen to the parameters estimated during ImageNet pre-training.

We used stochastic gradient descent with a momentum of 0.9 as the optimizer. The initial learning rates we used depended on the particular feature extractors as detailed below. The learning rate was regulated by the ReduceLROnPlateau scheduler implemented in the Pytorch library.

- ResNet-50 [15]: We use ResNet50 with a Feature Pyramid Network. All layers from the FPN were used. The initial learning rate was set to 3e-3.
- ResNet-101 [15]: We use ResNet101 along with a Feature Pyramid Network that extracts features of 256, 512, 1024 and 2048 dimensions. The initial learning rate is set to 1e-3.
- VGG16 [16]: Features are extracted from the last MaxPool2d layer (layer 30) of the network with a stride and kernel size of 2. The initial learning rate is set to 3e-3.
- VGG19 [16]: Features are extracted from the last MaxPool2d layer (layer 36) of the network. The initial learning rate is set to 3e-3
- EfficientNet-B3 [17]: Features are extracted from the last MBConv Layer of the network. The initial learning rate is set to 3e-3.

B. Evaluation Metrics

Precision, Recall, and Average Precision are standard metrics used to evaluate the performance of object detection

models. In the context of the malaria parasite detection problem, precision is the ratio of the number of cells correctly detected to the total number of cells detected. The recall is the ratio of correctly detected cells of a class to the number of all actual cells of that particular class.

The average precision, which is used as the evaluation metric in our experiments, is the average of all the precision rates at different recall rates. The average of AP across all classes is called mean average precision (mAP). Instead, it's commonly referred to as AP. Average precision is a good measure of an object detection model's comprehensive detection capacity.

TABLE I. AVERAGE PRECISION SCALES

AP at different IoU values	
AP_{50:95}	AP averaged over interval IoU=[0.50, 0.95](primary challenge metric)
AP₅₀	AP at IoU=0.50 (PASCAL VOC metric)
AP₇₅	AP at IoU=0.75 (strict metric)

C. Evaluation Results

The models were all tested under the same model parameters and hardware environment. However, the models were trained and tested in two different ways:

1) Binarized dataset approach

In this case, we binarized the classes in the dataset into either an uninfected or infected class. The uninfected class comprised of red blood cells and leukocytes; and an infected class that includes the ring, schizont, trophozoite and gametocyte classes. Other related works on this subject have sought to classify cells in smear-images as either infected or uninfected. The results of the performance of the models tested on a binarized version of the dataset, shown in Table 2 and Figure 5, may present a useful comparison to similar works. ResNet-101 performs marginally better than the other models in this case.

TABLE II. DETECTION RESULTS FROM BINARIZED DATASET

CNN Backbone	AP _{50:95}	AP ₅₀	AP ₇₅
ResNet-50	0.669	0.849	0.792
ResNet-101	0.69	0.87	0.815
VGG16	0.64	0.85	0.767
VGG19	0.664	0.857	0.8
EfficientNet B3	0.656	0.868	0.8

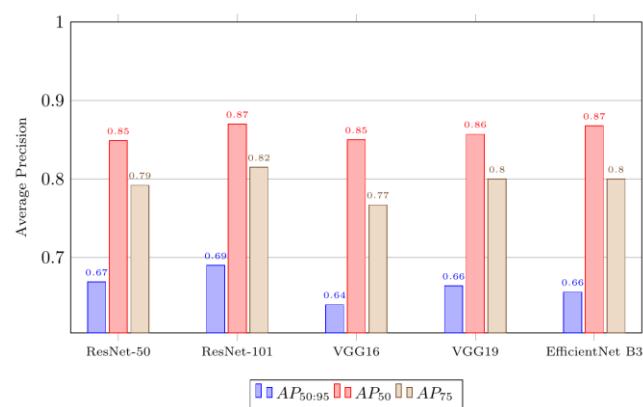


Fig. 4. Comparison of model accuracies on the binarized version of dataset

2) Standard dataset approach

In the second approach, the models were trained and tested on all classes included in the dataset except for the cells labeled as difficult. The class labeled as “difficult” was discarded from the dataset on the rationale that it would add noise to the system and make it difficult for the model to classify cells it has learned features of more confidently. As shown in the Table 3 and Figure 6, ResNet-101 is the better performing model in this case too, but with a larger margin than in the binarized approach.

TABLE III. DETECTION RESULTS FROM STANDARD DATASET

CNN Backbone	AP _{50:95}	AP ₅₀	AP ₇₅
ResNet-50	0.512	0.619	0.60
ResNet-101	0.70	0.88	0.842
VGG16	0.492	0.621	0.588
VGG19	0.505	0.649	0.611
EfficientNet B3	0.572	0.728	0.70

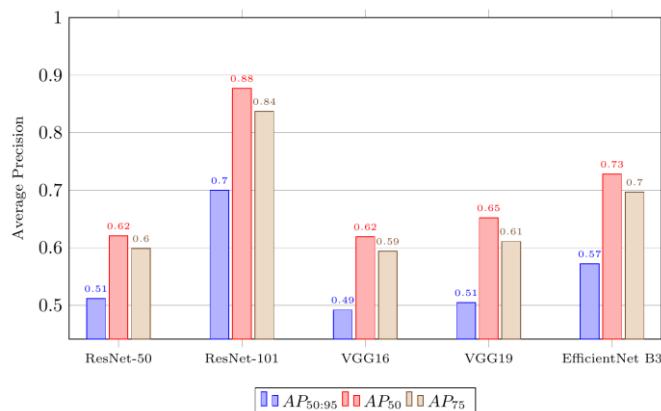
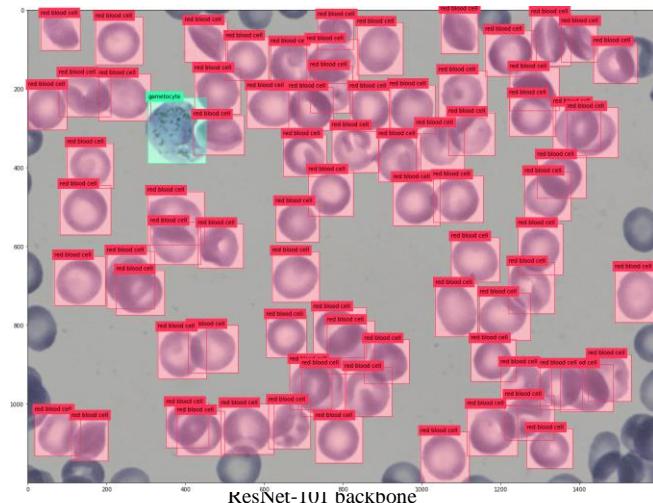


Fig. 5. Comparison of model accuracies on the standard version of dataset

D. Results Display

An example of detection results using ResNet-101 as a backbone for detection on the standard dataset are shown in the Figure 6. Due to the heavy in-balance in the dataset, the model has a high recall for the red blood cell class but performs relatively poorly on the other classes. Because of this, if there is a need to not only classify the particular development stage of the parasite and not just that the cell is infected, a two-stage detection model in which a second CNN is trained especially on the other classes in the dataset as was done in [18] would be recommended.



IV. CONCLUSION

In this paper, we used the Faster R-CNN object detection model to localize and classify uninfected and malaria-infected blood cells in giemsa-stained thin-blood smears. It is seen that a deep learning model that performs well on general image classification can also be adapted to microscopy image interpretation and have the potential to assist in the automation of malaria parasitological tests. We tested five deep learning models (ResNet-50, ResNet-101, VGG16, VGG19, and EfficientNet-B3) as feature extracting backbones for the Faster R-CNN model and compared their impact on the accuracy of the model. Of the models tested, ResNet-101 achieved the best results.

REFERENCES

- [1] Center For Disease Control and Prevention, “CDC - Parasites - Malaria.” <https://www.cdc.gov/parasites/malaria/>.
- [2] WHO, *Word Malaria Report 2021*. 2021.
- [3] M. S. ALILIO, A. MILLS, and J. G. BREMAN, “CONQUERING THE INTOLERABLE BURDEN OF MALARIA: WHAT'S NEW, WHAT'S NEEDED: A SUMMARY,” *Am. J. Trop. Med. Hyg.*, vol. 71, no. 2_suppl, pp. 1–15, Aug. 2004, doi: 10.4269/ajtmh.2004.71.2_suppl.0700001.
- [4] K. O. Mfuh *et al.*, “A comparison of thick-film microscopy, rapid diagnostic test, and polymerase chain reaction for accurate diagnosis of *Plasmodium falciparum* malaria,” *Malar. J.*, vol. 18, no. 1, p. 73, Dec. 2019, doi: 10.1186/s12936-019-2711-4.
- [5] N. Tangpukdee, C. Duangdee, P. Wilairatana, and S. Krudsood, “Malaria Diagnosis: A Brief Review,” *Korean J. Parasitol.*, vol. 47, no. 2, p. 93, 2009, doi: 10.3347/kjp.2009.47.2.93.
- [6] World Health Organization, *Guidelines for the Treatment of Malaria*, 3rd ed. 2015.
- [7] WHO, “Malaria microscopy quality assurance manual – Ver. 2,” *World Heal. Organ.*, p. 140, 2016.
- [8] S. S. Savkare, “Automatic Detection of Malaria Parasites for Estimating Parasitemia,” *Int. J. Comput. Sci. Secur.*, vol. 5, no. 3, pp. 310–315, 2011, [Online]. Available: http://www.researchgate.net/publication/229035721_Automatic_Detection_of_Malaria_Parasites_for_Estimating.
- [9] D. K. Das, M. Ghosh, M. Pal, A. K. Maiti, and C. Chakraborty, “Machine learning approach for automated screening of malaria parasites using light microscopic images,” *Micron*, vol. 45, pp. 97–106, Feb. 2013, doi: 10.1016/j.micron.2012.11.002.

- [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, 2015, vol. 28, [Online]. Available: <https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bb028a21ed38046-Paper.pdf>.
- [11] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," *Nat. Methods*, vol. 9, no. 7, pp. 637–637, Jul. 2012, doi: 10.1038/nmeth.2083.
- [12] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color and Imaging Conference*, 2004, vol. 2004, no. 1, pp. 37–41.
- [13] C. Poynton, *Digital video and HD: Algorithms and Interfaces*. Elsevier, 2012.
- [14] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020, doi: 10.3390/info11020125.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, doi: 1409.1556.
- [17] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: <http://arxiv.org/abs/1905.11946>.
- [18] J. Hung and A. Carpenter, "Applying faster R-CNN for object detection on malaria images," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 56–61.