

Malaria Parasite Detection in Thick Blood Smear Microscopic Images Using Modified YOLOV3 and YOLOV4 Models

Fetulhak Abdurahman (✉ afetulhak@yahoo.com)

Jimma University <https://orcid.org/0000-0002-5670-0319>

Kinde Fante Anlay

Jimma University

Mohammed Aliy

Jimma University

Research

Keywords: Malaria, Plasmodium falciparum, Thick blood smear, Deep Learning, Object detection, YOLOV3, YOLOV4, Feature map

Posted Date: September 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-74079/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Malaria Parasite Detection in Thick Blood Smear Microscopic Images Using Modified YOLOV3 and YOLOV4 Models

Fetulhak Abdurahman^{1*}, Kinde Anlay² and Mohammed Aliy³

*Correspondence:

afetulhak@yahoo.com

¹Faculty of Electrical and Computer Engineering, Jimma University, 378 Jimma, Ethiopia
Full list of author information is available at the end of the article

Abstract

Background Information: Manual microscopic examination is still the "golden standard" for malaria diagnosis. The challenge in the manual microscopy is the fact that its accuracy, consistency and speed of diagnosis depends on the skill of the laboratory technician. It is difficult to get highly skilled laboratory technicians in the remote areas of developing countries. In order to alleviate this problem, in this paper, we propose and investigate the state-of-the-art one-stage and two-stage object detection algorithms for automated malaria parasite screening from thick blood slides.

Methods: YOLOV3 and YOLOV4 are state-of-the-art object detectors both in terms of accuracy and speed; however, they are not optimized for the detection of small objects such as malaria parasite in microscopic images. To deal with these challenges, we have modified YOLOV3 and YOLOV4 models by increasing the feature scale and by adding more detection layers, without notably decreasing their detection speed. We have proposed one modified YOLOV4 model, called YOLOV4-MOD and two modified models for YOLOV3, which are called YOLOV3-MOD1 and YOLOV3-MOD2. In addition, we have generated new anchor box scales and sizes by using the K-means clustering algorithm to exploit small object detection learning ability of the models.

Results: The proposed modified YOLOV3 and YOLOV4 algorithms are evaluated on publicly available malaria dataset and achieve state-of-the-art accuracy by exceeding the performance of their original versions, Faster R-CNN and SSD in terms of mean average precision (mAP), recall, precision, F1 score, and average IOU. For 608 × 608 input resolution YOLOV4-MOD achieves the best detection performance among all the other models with mAP of 96.32%. For the same input resolution YOLOV3-MOD2 and YOLOV3-MOD1 achieved mAP of 96.14% and 95.46% respectively.

Conclusions: The experimental results demonstrate that the performance of the proposed modified YOLOV3 and YOLOV4 models are reliable to be applied for detection of malaria parasite from images that can be captured by smartphone camera over the microscope eyepiece. The proposed system can be easily deployed in low-resource setting and it can save lives.

Keywords: Malaria; Plasmodium falciparum; Thick blood smear; Deep Learning; Object detection; YOLOV3; YOLOV4; Feature map

1 Background

Malaria is the leading public health burden. Its prevalence is expressively high, especially in low-income countries. In 2018 only, an estimated number of 228 million

cases were recorded worldwide, and most of the cases were in Africa (comprising 93%) followed by the South-East Asia region (3.4%). The cause of malaria is a unicellular parasite called plasmodium. The parasite is transmitted from the infected person to the healthy through the bites of the female anopheles mosquito. Generally, there are five different species of plasmodium (*p. falciparum*, *p. vivax*, *p. ovale*, *p. malariae*, and *p. knowlesi*) of which, *p. falciparum* is the most familiar species found in clinical diagnosis. Following *p. falciparum*, the next species that commonly infects humans is the *p. vivax*, while the remaining three are rare in clinical settings. Depending on the severity of the infection, all the species pass through four life stages: ring, trophozoite, schizont, and gametocyte [70].

Malaria is curable, but late diagnosis and treatment will cause serious health complications even the loss of life. Microscopy-based visual examination of stained thick and thin blood slides is the golden standard for malaria diagnosis [44]. However, the accuracy of microscopy-based diagnosis heavily depends on the individual's slide reading experience and attentiveness during the procedure. Besides, the shortage of well-trained personnel imposes discrepancies in the effectiveness of microscopy-based diagnosis results in malaria-endemic and resource-constrained areas, especially in rural parts of Africa. Inaccurate diagnosis results will lead to morbidity and socio-economic problems within society. It will also lead to poor decision making and planning in the eradication process of malaria.

To overcome the drawbacks of manual microscopy-based diagnosis, an alternative low cost, fast and accurate computer-aided diagnosis system is required. The recent advancements in the field of computer vision after the introduction of deep learning algorithms have been showing promising results for the detection of the malaria parasite and, in general, for the detection of abnormalities in medical images.

For the past few decades, a substantial number of studies have been conducted on the detection of malaria parasites and other disease pathogens from microscopic images using different image processing techniques. The traditional image processing techniques based on image segmentation to identify the target object within an image has a high computational cost and low-performance [47] [26] [46] [57]. The classical machine learning approaches based on manually extracted features for the specified object type had also been widely studied as object classification and detection technique [56] [65] [11]. The limitation of such classical machine learning-based techniques is their inability to cope-up with the inherent variability of images captured in different environments.

In recent years, deep learning based object detection and classification techniques have gained popularity because of their ability to overcome the limitations of the traditional image processing techniques, including the classical machine learning algorithms. Convolutional neural networks (CNNs) are the backbone of such deep learning algorithms. One of the advantages of CNNs is their capability to extract visual features automatically from labeled or annotated images and learn from different representations of a single object acquired at different environmental conditions. Moreover, they show high performance and accuracy in detecting and classifying objects from raw images.

The deep learning techniques used for object detection can be categorized into three groups. The first type is the use of conventional CNN architectures to classify

objects by taking image patches that are cropped using different preprocessing steps such as segmentation or by applying sliding window technique over the input image. These techniques are computationally expensive due to the massive amounts of patches generated and the application of convolution operation for all the image patches [75]. The second type is based on proposing region of interest from the image and classifying those regions as an object of interest or not e.g R-CNN [18], Fast R-CNN [17], Faster R-CNN [55], and SPP-Net [21]. These types of algorithms are called two-stage object detection algorithms since they have a region proposal part and the object detection and classification unit separately. The third types are called one-stage object detection algorithms such as YOLO decedents (V1 [52], V2 [53], V3 [54], V4 [3]), SSD [36] and others which take the object detection process as one complete regression problem rather than proposing region of interests separately.

The motivation of this study is to investigate the applicability of the state-of-the-art one-stage and two-stage object detection algorithms for detection of malaria parasite from microscopic images captured using mobile phone camera. Despite their success, the state of the art deep learning based one-stage and two-stage object detection algorithms are not investigated by researchers for detection of malaria parasite. There is one study that attempted to use modified YOLOV3 architecture for the detection of malaria parasite [6]. The model was proposed to make the detection process computationally lightweight so that it can be deployed on mobile devices. This lightweight model was obtained by modifying the backbone of YOLOV3 and the reported performance of the model is about 90.2% mean average precision.

In the current work, we propose malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models. We have modified these models to improve their capability to detect small objects such as malaria parasites. The modified models achieve the state-of-the-art accuracy. The contribution of this study are as follows:

- 1 To improve malaria parasite detection capability of YOLOV3 and YOLOV4 based models we modify the network architecture of the original models by extending the feature scales and by adding the detection layer size. The modified models have higher small object detection capabilities.
- 2 Comprehensive experiments are done to evaluate the detection and speed performance of the original and improved YOLOV3 and YOLOV4 models using publicly available malaria dataset.
- 3 We have done a comprehensive comparative study to evaluate the performance of state of the art two-stage and one-stage object detection algorithms such as Faster RCNN, SSD, YOLOV3, and YOLOV4 for the detection of malaria parasite.

The remainder of this paper is organized as follows. Section 2 describes in detail the existing studies for malaria parasite detection. In section 3 the results are discussed in details by comparing the performance of state of the art one-stage and two stage object detectors used in this study. In section 4 the conclusions are made based on the experimental results. Section 5 describes in detail the methodology used in this study including the dataset 5.1, the detection architectures 5.2 and the training process and hyper-parameter optimization techniques 5.2.4.

2 Related Studies

Researchers have proposed various techniques of image processing and machine learning algorithms for automated detection of malaria parasite from digital microscopic images. The existing studies on this topic can be grouped into three: 1) conventional image processing and machine learning based techniques, 2) patch based deep learning algorithms, and 3) one-stage and two-stage deep learning object detection algorithms.

Most of the existing works for automated detection of malaria parasites use the conventional image processing and machine learning techniques. However, only a few research works has been reported in literature on detection of malaria parasite using the state-of-the-art deep learning algorithms. In this section, the comprehensive survey of the existing works is presented.

2.1 Conventional Image Processing and Machine Learning Based Techniques

Before the advancement of deep learning models in solving object detection problems, algorithms based on image processing were widely employed for image segmentation and detection. Several research studies have been proposed using conventional image processing and machine learning methods for automatic detection of malaria parasite from both thin and thick blood smears. A study by [42] uses an image processing approach called voxel (volume pixel) on which temporal standard deviation (STD) was used to obtain a single image from a video sequence that recorded the movement of an infected RBC cell. The bi-modal histogram of the greyscale image was calculated to differentiate between the two classes in the proposed model which are pixels with strong movement and higher STD and pixels with softer movement. A high STD means that there is a change in pixel values indicating the presence of an object while low temporal STD reports low activity indicating there is no object. An automated end-to-end *p. falciparum* detection and quantization system is developed using thin blood smear of human and mouse [47]. In this work first, they segmented the RBC cells using multiscale Laplacian of Gaussian (LoG) detection followed by an active contours-based segmentation scheme named coupled edge profile active contours (C-EPAC). Then by using the combination of color and texture information to distinguish infected and uninfected RBCs. However, the detection and the segmentation process is computationally expensive and they used a limited number of training images which will make their system unable to generalize in big data sets.

The study by [43] tries to use an estimation based on their assumption on isolated and clustered RBCs to avoid the time-consuming segmentation algorithms using region growing, template matching, active contour tracing and morphological operation, etc. In this work, they develop a local maxima of intensities based algorithm to detect the parasite signals from the input image composed of three different channels for WGA-AlexaFlour488 fluorescence (green), DAPI fluorescence (blue), and for Mitotracker fluorescence (red). In their detection process, they also identify the different stages of *p. falciparum* using a complete image processing technique and they did not use machine learning classifiers. The study by [26] tries to design a motorized automated device to move the microscope stage in a 3-directional plane to get high-quality microscopic images for further image processing to detect *p.*

falciparum and *p. vivax* from thick blood film. Their Image processing technique is very high level and did not analyze the detail morphology of the parasite and they use only the size of the chromatin to classify the detected object as *p. falciparum*, *p. vivax* or not. In their work, they did not describe detailed image processing techniques used in the process of segmentation which is the essential component in object detection.

Rosado et al. [56] used a two-class support vector machine (SVM) classifiers to classify *p. falciparum* trophozoites and white blood cells (WBC) from mobile phone camera captured thick blood smear microscopic images. In this work, they extracted 314 different features for every candidate trophozoites and 152 feature for every candidate WBC images using the color, texture, and geometry information. They use the adaptive thresholding methods to segment both WBC and trophozoite candidates before extracting the features mentioned above. Based on the collected features the SVM classify the segmented candidate as WBC or not and trophozoite or not. They achieved 98.2% of sensitivity and 72.1% specificity for WBC detection and achieved a sensitivity of 80.5% and a specificity of 93.8% for *p. falciparum* trophozoites detection. A work done by [46] use linear discriminant classification (LDC), k-nearest neighbor (NNC) and linear regression (LR) to detect malaria parasite using 23 morphological descriptors extracted from quantitative phase images of live, unstained red blood cells. [7] used bayesian learning and support vector machine (SVM) as a classification algorithm after collecting 96 morphological and textural features from different stages of *p. vivax* and *p. falciparum* infected erythrocytes. They used leishman stained thin blood smear slides.

Another study by [65] investigated the possibility of detecting and identifying the four species of malaria, *p. falciparum*, *p. vivax*, *p. ovale*, *p. malariae* from thin blood smear microscopic images. They have used the KNN classifier with a relative distance metric which is based on a normalized version of L1 norm. They extract 83 features based on color histogram, local area granulometry and shape measurement vector. Their proposed model achieved better results compared with fisher linear discriminant (FLD) and custom neural network model. Devi et al. [11] proposed prediction error, LBP–GLCM, chrominance channel histogram, R–G channel difference histogram and Gabor features to combine them with existing features such as green channel histogram, saturation channel histogram, Shanon entropy, Renyi entropy, Havarda and Charvat entropy and Kapur entropy to discriminate infected erythrocytes from microscopic images of thin blood smear slide. In their work a hybrid machine learning classifiers SVM, k-NN and ANN achieved better results when compared with individual classifiers.

A study by [48] uses a modified K-means clustering algorithm based on the pixel intensity information of the pre-processed image. To not loss the pixel intensity information of stained parasites that constitute the smallest cluster group, they retained the information by imparting higher weights to the small clusters. The pixel intensity information only cannot be enough to detect and identify malaria parasites in this study as textural and morphological information is not included. Similar work is done by [1] using an unsupervised modified K-means clustering algorithm to segment malaria parasites from the saturation and intensity component of HSI color space of thin blood film microscopic image. In this work only the *p. vivax* trophozoite and gametocyte stages of the malaria parasite are considered.

A study by [57] used the average eccentricity calculated from the binary image, color, shape, and average radius of the erythrocytes calculated using granulometry as a feature to detect malaria parasites after segmentation of the background, erythrocyte, and parasite using image histogram-based threshold. In this work they used two separate back propagation feed-forward neural networks one for classifying the segmented erythrocytes as infected or not and the second to classify the life stages. The threshold based segmentation did not work well for *p. malariae* parasites. Bibin et al. [2] proposed a deep belief network (DBN) to differentiate malaria parasites from non-parasites using color histogram, color coherence vector, and gray level run length matrix features to train the DBN.

Recently, a study done on performance analysis of feature sets used for malaria-infected erythrocyte detection shows morphological, intensity, and texture features have greater discrimination capability [10]. The review study done by [39] indicates morphological operators are the most successful and commonly used techniques in malaria parasite detection and segmentation process in conventional image processing techniques for malaria parasite detection.

2.2 Patch Based CNN Approaches for Detecting Malaria Parasite

Machine learning models based on hand-crafted features achieve poor generalization capability in classification tasks. This is because hand-crafted features are sub-optimal for a classification task. Besides, the hand-engineered feature extraction, ranking, and selection is computationally expensive and demands for expertise to extract those features. In contrast, deep-learning based classification models avoid the difficulties of the conventional hand-crafted feature based malaria parasite detection methods with their capability to learn robust features from diversified training data automatically. Since LeNet's first introduction [29] and AlexNet's [27] remarkable performance at the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [58], CNNs have been widely used for various image classification and object detection tasks. In patch-based CNN object detection techniques first, the region of interests (ROIs) are cropped from the input image using either image processing techniques or sliding window approach, then the ROIs will be an input to the CNN for classification. In many applications, the patch extraction process generated a large number of patches for a single image. The convolution operation is done for the individual patches which makes the computational cost too high.

The comparative analysis of the performance of three deep CNN models was performed for the detection of malaria-infected cells in [12]. In this work, the Red blood cells (RBC) were segmented from thin blood film whole slide images using thresholding operation and Hough Circle transform [13]. The performances of LeNet-5, AlexNet, GoogLeNet, and SVM were compared for the classification of the patches which were obtained from the segmented image regions. Among the models, GoogLeNet achieved the best performance of 98.13% from the given test data set whereas SVM scored the lowest accuracy. A customized and pre-trained CNN based deep learning models for classifying parasitized and uninfected thin blood cells to screen malaria cases were reported in [51]. The pre-trained CNN models used feature extractors including AlexNet, VGG-16, Xception, ResNet-50, and DenseNet-121 by optimizing their parameters and hyper-parameters. From the

reported statistical analysis pre-trained ResNet-50 achieved the best performance in discriminating infected erythrocytes as compared to the other models.

Gopakumar et al. [19] used focus stack of images taken as frames of video containing focus stack of multiple Field-of-Views (FoV) of Leishman stained slide images to detect *p.falciparum* parasites and different life-cycle stages of the parasite. After detection of suspected parasite location using local intensity minima, a 32 X 32 patch is taken surrounding the area. The patches are given as input to the SVM classifier based on intensity and textural information of the patch. CNN classification based on a single patch and focus stack of patches as input is also used to classify between healthy and infected patches of malaria parasite. In this work, the focus stack of patches with CNN as a classifier achieved the best performance with sensitivity and specificity of 97.06% and 98.50% respectively. Liang et al. [31] reported a custom CNN model to classify infected erythrocytes from uninfected ones using single-cell thin blood smear images. The study by Torres et al. [66] developed a prototype digital device Autoscope based on a CNN algorithm for identification and counting of malaria parasite from giemsa stained thick blood film. The images are captured in a very controlled environment and this makes it difficult for their device to work on images captured in real-time clinical settings.

Another pre-trained VGG19 CNN model for feature extraction and SVM classifier was proposed for malaria microscopic identification [69]. The study reported in [72] describes the application of CNN algorithms to identify malaria parasites from thick blood smear. The intensity based iterative global minimum screening (IGMS) is used as a pre-processing technique to detect candidate parasite locations to be given as input to the CNN for further classification. Their mobile phone based application achieves an accuracy of 97.26% and AUC of 97.34% on patch level with a promising correlation between ground truth and detected parasite locations in image level.

Two classification techniques for malaria parasite detection were reported in [49]. The first technique extracts connected component and moment features from pre-selected patches based on a sliding window technique and apply extremely randomized trees for the classification purpose. The second technique applies a custom CNN algorithm to classify between infected and uninfected patches of malaria taken from thick blood microscopic images. The CNN based technique outperforms the classical decision tree based system having an average precision of 97%. This method uses patch extraction using a sliding window technique which takes longer processing time during inference. Rahman et al. [50] demonstrated three techniques to classify malaria parasite from segmented thin blood smear images. The first one is custom CNN with 19 layers. The second one is a transfer learning approach in which they use VGG16 pre-trained on ImageNet by freezing layer L1 to layer L16 in the VGG16 network architecture. In their third technique features are extracted using the pre-trained VGG16 network model and feed into the SVM classifier for detecting the malaria parasite as a final output. They also demonstrated the ensemble of the above three approaches in their detection model but the transfer learning model achieves better performance compared with the other techniques.

Pan et al. [45] showed a deep convolutional neural network model based on LeNet5 architecture for detection and classification of malaria parasite from thin blood film whole slide images. Since their dataset was too small to avoid the problem of over-fitting they use an image interpolation algorithms in the spatial domain and feature

domain as a data augmentation technique. The advantage of visualizing features extracted in the convolution and activation layers to better understand the learning strategy of deep learning models for classification of malaria parasite was discussed in [61]. The study by Delahunt et al. [8] proposed a patient-level training/validation results using two different CNN architectures for the ring and late-stage malaria parasites. Yang et al. [73] developed a mobile phone based malaria parasite screening from thick blood smear microscopic images using patch based custom CNN algorithm. In their work, they used intensity based histogram analysis technique to preselect parasite locations within the image during the detection process. They achieved an accuracy of 97.26% accuracy on patch level and above 98% correlation coefficient for both image and patient-level evaluation of their system.

2.3 The State-of-the-art One-stage and Two-stage Object Detection Algorithms

After the success of AlexNet in 2012, image classification using deep learning methods has become a hot research topic. Researchers have developed different deep neural network architectures such as ZFNet [74], VGG [59], GoogleNet(Inception)[62], Resnet [20], Densenet [22] which has decreased the top 5 error rate at ILSVRC image classification challenge with record-breaking performance. However such new architectures come up with a large number of trainable parameters which makes the use of patch based and sliding window based CNN techniques to be abandoned due to their high computational cost.

Then, a research team at UC Berkley developed a CNN based object detection architecture, which they called it regions with CNN (R-CNN). This was the first two-stage object detection algorithm based on deep learning. Even though R-CNN shows dramatic improvement in object detection tasks as compared to the classical hand-crafted feature based systems it has its own limitations. To further improve the performance of deep learning based models, other researchers come up with different two-stage and one-stage deep learning based object detection algorithms such as Fast R-CNN, Faster R-CNN, MASK R-CNN, SPPNet, YOLO, SSD, RetinaNet and others [25].

Lately, state of the art one-stage and two-stage object detection algorithms are being used for medical image analysis: detection of cancer ([68], [34], [67], [16]), detection of organs and their abnormalities([30], [14], [37], [64]), for detection of pulmonary diseases ([40],[71], [60]), for detection and segmentation of intracranial hemorrhages ([4]), for classifying and segmenting microscopy images ([28], [14], [38]).

Despite their success in several applications, the state-of-the-art deep learning based one-stage and two-stage object detection algorithms have not been extensively studied for the detection of malaria parasite in microscopic images. The work reported in [6] uses a modified YOLOv3 architecture to detect *p.falciparum* parasite from thick blood smear microscopic image taken with a digital microscope and smartphone camera. The modified model was obtained by replacing the feature extraction part with convolutional bottleneck residual blocks and removing some of the convolutional layers at the detection part of the original YOLOv3 architecture. This work has lower computational complexity and high inference speed as compared to the original YOLOv3 version; at the cost of lower detection accuracy. Automated *p.vivax* detection system from microscopic images of thin blood

smear was reported in [5]. The model uses the transferred AlexNet classifier in cascade with YOLOv2 to reduce the false-positive prediction of their model. This cascaded model achieves 8% improvement as compared to the original YOLOv2 model. Faster Region-based Convolutional neural network (Faster RCNN) model, which was pre-trained on ImageNet, was applied to classify red blood cells (RBC's) and other non-RBC objects [23]. After detecting the non-RBC objects (including infected cells and leukocytes), the model uses the AlexNet classifier to classify the detected cell categories into leukocyte, gametocyte, ring, trophozoite, and schizont. This model achieved improved accuracy as compared to the baseline model which employs conventional hand-engineered feature for classification of each cell after segmentation. Careful observation of malaria parasites from microscopic images reveals that the objects to be detected by the deep learning models are very small. The above algorithms are not optimized for the detection of very small objects. In our work, we optimize the hyper-parameters of the existing object detection models to make them specialized in detecting small objects in an image.

3 Result and Discussion

In this study we have done detailed experiments to evaluate the performance of state of the art one-stage and two-stage deep learning based object detectors for detection of *p.falciparum* from thick blood smear microscopic images captured using smartphone camera. During evaluation, the dataset is randomly split into 90/10 for training/testing for all the detection models. We evaluated the detection models used in this study based on the evaluation metrics including mean average precision(mAP), precision, recall, F1 score, and inference time in frames per second (FPS).

3.1 Comparison of Modified YOLOV3 and YOLOV4 Models with Their Original Versions

For the YOLO based models (original YOLOV3, YOLOV4, and their modified versions) we have used an IOU threshold of 0.3, which is selected experimentally, to calculate the mAP during performance evaluation at inference time. As shown in Figure 1 the performance of the detection models at IOU value of 0.3 is optimal and it decreases while increasing the value to 0.4 and 0.5.

During the model evaluation, we did not rely on a single evaluation metric to say the detection model is good enough to be used for *p.falciparum* detection in microscopic images. In addition to the above evaluation metrics, we visualize in detail the number of detections as true positive (TP), false positive(FP), and false negative(FN) as shown in Table 2. The major challenge in clinical settings when screening microscopic images is the ability to separate between artifacts and real parasites. We experimentally select the best confidence threshold values during inference to obtain the best-balanced detection output between the false positives(artifacts) and false negatives for all the YOLO based detection models at IOU value 0.3.

According to our experimental results presented in Table 1 the YOLOV4-MOD achieves the highest detection performance with mAP of 96.32% and 96.20% at input resolutions of 608 x 608 and 416 x 416 respectively. The original YOLOV4 achieves mAP of 95.84% and 95.44% at 608 x 608 and 416 x 416 respectively.

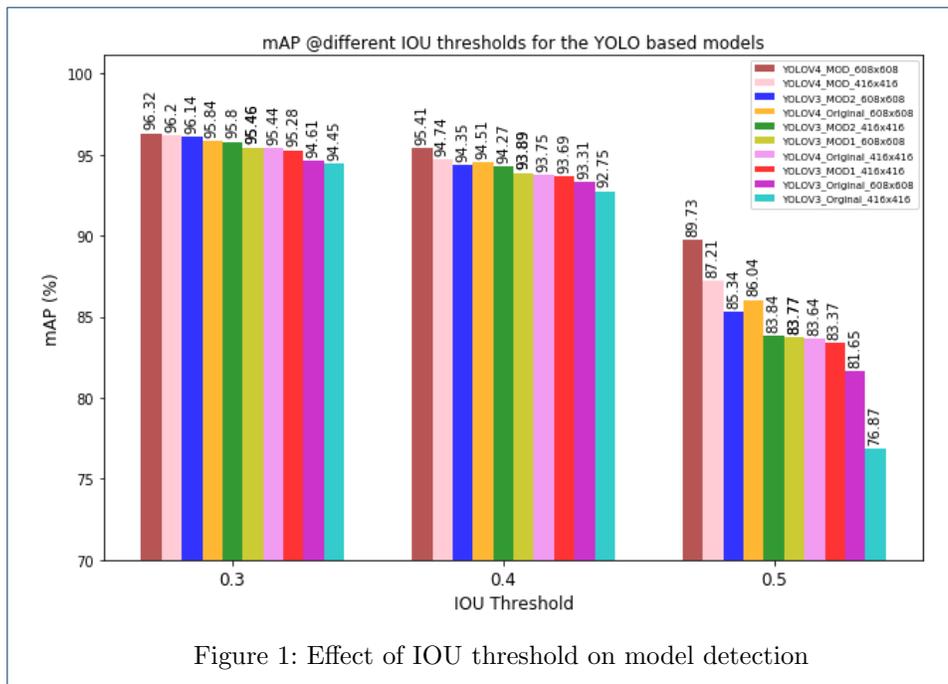


Figure 1: Effect of IOU threshold on model detection

The results indicate that the YOLOV4-MOD shows better *p.falciparum* detection performance compared with its original version. YOLOV4-MOD also provides the highest precision (95%), recall(94%) and F1 score(94%) for our test set data. As shown in Table 2 the YOLOV4-MOD at 608 x 608 resolution also shows less FP(50) and FN(60) predictions compared with all the other models.

All the YOLO based models provide higher *p.falciparum* detection performance at an input size of 608 x 608, but at the same time made the detection slower than the 416 x 416 input resolution. The modified YOLOV3-MOD2 with an input size of 608 x 608 has ranked third with mAP of 96.14% compared with the other YOLO based models and it has ranked first among the other YOLOV3 based models. It also has a balanced precision(92%) and recall(93%) rate indicating the model can discriminate well the artifacts from parasites. The YOLOV3-MOD1 has achieved mAP of 95.46% and 95.28% at input resolution of 608 x 608 and 416 x 416 respectively. The YOLOV3-MOD1 shows lower detection performance among the modified YOLOV3 versions but it still has shown better detection performance compared with the original YOLOV3 models.

Besides we have seen the average IOU of the different models to compare their ability for precise localization of the parasites during inference time. As shown in Table 2 both the modified YOLOV3 and YOLOV4 models show better average IOU results compared with their Original counterparts. The detection speed comparison of the different models is shown in Table 1 represented by frames per second (FPS). From the table, the YOLOV4-MOD shows small increase in its inference time as compared with the original YOLOV4 model. However, the inference time for YOLOV3-MOD2 is much lesser than its original version. This is due to the large convolution computation at shallow feature maps having a bigger feature scale.

When we compare the detection performance between the YOLOV3 and YOLOV4 models both in the original and modified versions, the YOLOV4 models achieve

better *p.falciparum* detection performance. The YOLOV4-MOD is the best models selected for the *p.falciparum* detection compared with the other YOLO based models. In general for small object detection such as *p.falciparum* our modified YOLOV3 and YOLOV4 models with fine-grained features at high-resolution feature maps show better detection performance compared with the original versions of YOLOV3 and YOLOV4. It implies that the modified network structure learns more geometric and semantic information for small-objects compared with their original versions.

3.2 Comparison of Modified YOLOV3 and YOLOV4 Models with Other State-of-the-art Detection Methods

We also compare the performance of the modified YOLOV3 and YOLOV4 models with recent state-of-the-art one-stage detector SSD and two-stage detector Faster R-CNN as shown in Table 1. From the table the modified as well as the original versions of YOLOV3 and YOLOV4 achieve better *p.falciparum* detection result compared with Faster R-CNN and SSD models. The Faster R-CNN achieves 71.0% mAP while the SSD achieves 71.4%. The SSD model have shown highest inference time compared with the other models but it has the worst detection accuracy as compared to the YOLO based models. The Faster R-CNN model is very slow in its detection speed and has lower detection performance compared with the other models.

Table 1: Comparisons of accuracy and speed for the detection models using our malaria dataset. The highlighted rows indicate the best performing model.

Models	mAP@0.3	Precision	Recall	F1-Score	FPS
YOLOV4-MOD @608 × 608	96.32%	95%	94%	94%	29.60
YOLOV4-MOD @416 × 416	96.20%	93%	93%	93%	30.56
YOLOV3-MOD2 @608 × 608	96.14%	92%	93%	92%	15.30
YOLOV3-MOD2 @416 × 416	95.80%	92%	92%	92%	17.83
YOLOV3-MOD1 @608 × 608	95.46%	92%	92%	92%	21.40
YOLOV3-MOD1 @416 × 416	95.28%	92%	92%	92%	26.75
YOLOV4 @608 × 608	95.84%	92%	92%	92%	30.77
YOLOV4 @416 × 416	95.44%	92%	92%	92%	33.89
YOLOV3 @608 × 608	94.61%	91%	92%	92%	28.67
YOLOV3 @416 × 416	94.45%	91%	91%	91%	30.43
Faster R-CNN	71.0%	92.7%	86.9%	89.71%	8
SSD @300x300	71.4%	91%	84%	87%	41

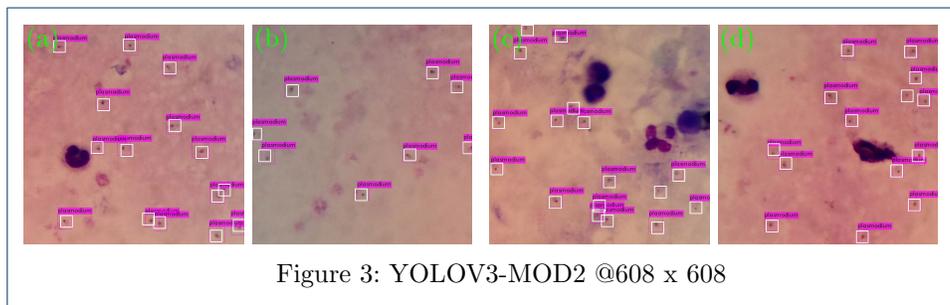
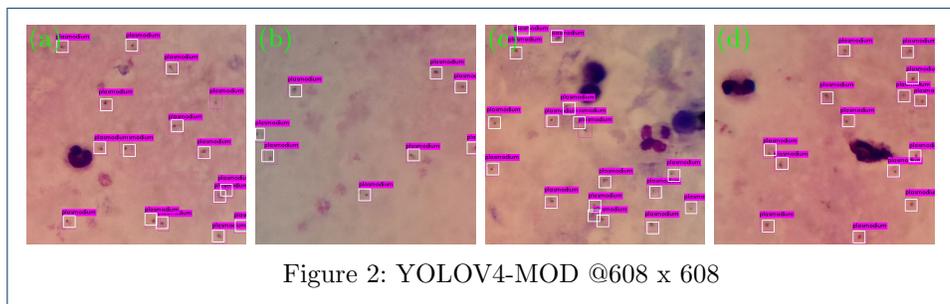
Table 2: Comparison of average IOU and detection results in terms of TP, FP and FN predictions for the YOLO based models using our test set data

Models	TP	FP	FN	Avg. IOU
YOLOV4-MOD @608 × 608	946	50	63	62.12%
YOLOV4-MOD @416 × 416	941	71	68	61.84%
YOLOV3-MOD2 @608 × 608	938	84	71	61.77%
YOLOV3-MOD2 @416 × 416	928	83	81	61.03%
YOLOV3-MOD1 @608 × 608	927	86	82	61.03%
YOLOV3-MOD1 @416 × 416	924	78	85	60.64%
YOLOV4 @608 × 608	930	81	79	61.15%
YOLOV4 @416 × 416	925	80	85	60.67%
YOLOV3 @608 × 608	932	88	77	59.98%
YOLOV3 @416 × 416	916	95	93	58.85%

3.3 Visualization Test Results

To further evaluate the *p. falciparum* detection performance of the different detection models used in this study partial detection results of the models are shown in Figures 2-8. In the figures, we have used the 608 x 608 resolution for the YOLO based models since all the models achieve better detection results at this resolution. The figures indicate the performance of the different models under the same four test images. The ground truth and the predicted bounding boxes are shown in white and violet respectively. As shown from Figure 2 the YOLOV4-MOD detects all the parasites with only one false positive case in the first image and one false negative and one false positive prediction results in the third image. The YOLOV4-MOD has shown better detection results compared with the original YOLOV4 shown in Figure 5 with minimal added computation cost. This is mainly due to rich feature information learned by concatenating additional shallow level features with semantic rich high-level features in the modified YOLOV4 architecture.

In Figures 3 and 4 the YOLOV3-MOD2 and YOLOV3-MOD1 models correctly recognize the *p.falciparum* with only few incorrectly detected parasites compared with the original YOLOV3 in Figure 6 which misses more parasites. For the Faster R-CNN and SSD models as shown in Figures 7 and 8 the models predicts a lot of parasites as non-parasite as compared to the modified models. In general, from the figures we can see that the detection performance of the improved YOLOV3 and YOLOV4 models shown better detection results with good localization as compared to their original versions and other detection models.



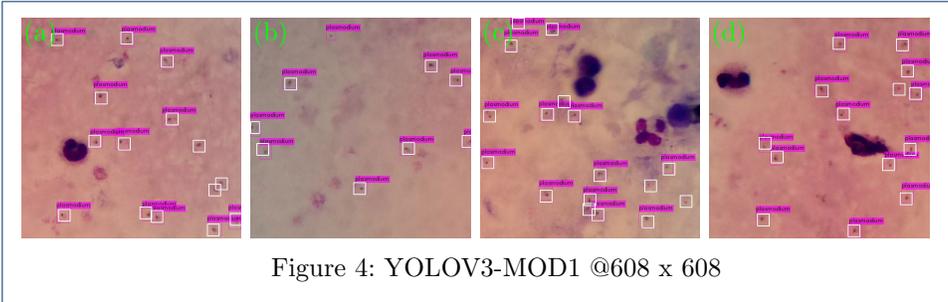


Figure 4: YOLOV3-MOD1 @608 x 608

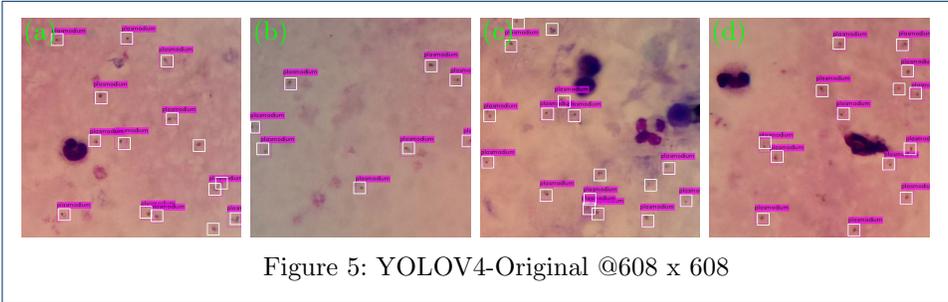


Figure 5: YOLOV4-Original @608 x 608

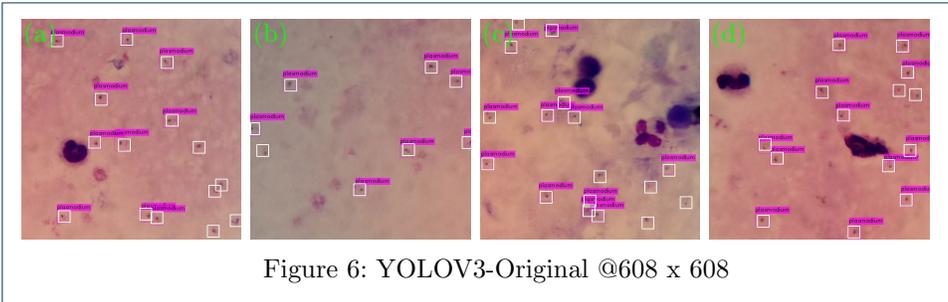


Figure 6: YOLOV3-Original @608 x 608

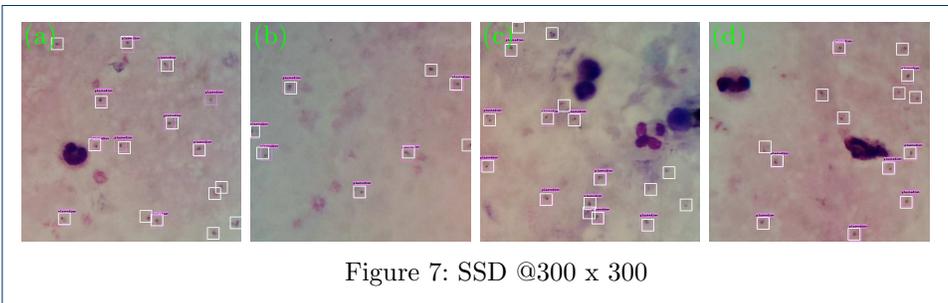


Figure 7: SSD @300 x 300

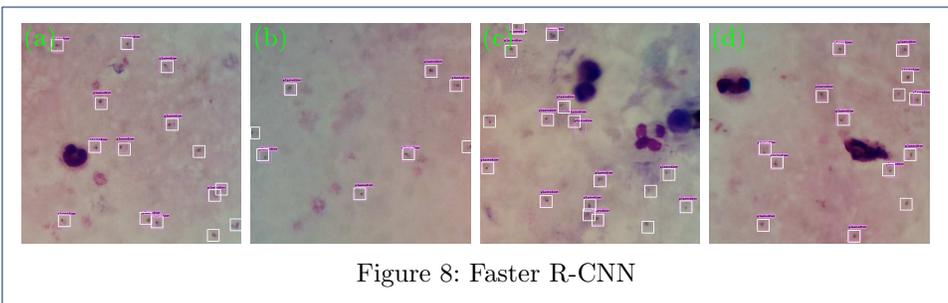


Figure 8: Faster R-CNN

4 Conclusions

Malaria poses significant health challenges in developing countries. The diagnosis and treatment of the disease is an important step to cure patients. The current manual microscopic examination of malaria depends on the skills and experience of the microscopist. The shortage of skilled microscopists in developing countries has motivated the researchers to develop a computerized diagnosis system. Due to the advancement of deep learning in image classification and object detection tasks, automated microscopic diagnosis of malaria can be made more effective, efficient and reliable.

In this paper, the performance of the state-of-the-art one-stage and two-stage object detection algorithms are thoroughly investigated in the detection of malaria parasite from thick blood sample microscopic images. The original versions of YOLOV3 and YOLOV4 models are not optimized to detect small objects since these models have bigger receptive fields at their detection heads. We have modified the network architecture of the YOLOV3 and YOLOV4 object detection algorithms to enhance their performance in small-object detection. For the YOLOV3 model, we have modified the network architecture 1) by extending the feature map scale at the third detection head of the original YOLOV3(YOLOV3-MOD1) and 2) by extending the detection heads from three to four (YOLOV3-MOD2). From the two modified versions of the YOLOV3 architecture, YOLOV3-MOD2 have shown better small object detection accuracy. Similarly, we have modified the original YOLOV4 model by extending the detection head from three to four (YOLOV4-MOD). These modifications have enabled the models to obtain more detailed spatial and geometric information to detect small objects. We have also adjusted the scales of the anchor boxes and generated new anchor sizes based on clustering results from publicly available malaria dataset. These enable the detection models to precisely predict the location of the malaria parasite in microscopic images.

Several experiments were conducted to compare the performance of our modified YOLOV3 and YOLOV4 based models with existing models such as SSD, and Faster R-CNN. The performance of the proposed modified YOLOV3 and YOLOV4 algorithms are evaluated on publicly available malaria dataset. For 608 x 608 input image resolution, YOLOV4-MOD has achieved the best detection performance among all the other models with mAP of 96.32%. For the same input image resolution, YOLOV3-MOD2 and YOLOV3-MOD1 have achieved mAP of 96.14% and 95.46% respectively. The modified models outperform the original versions of YOLO3 and YOLOV4, Faster R-CNN and SSD in terms of mean average precision (mAP), recall, precision, F1 score, and average IOU. In this study, we have demonstrated the feasibility and effectiveness of our proposed YOLOV3 and YOLOV4 modified architectures for the detection of *p. falciparum* in microscopic images captured using mobile phone camera. The future work will focus on demonstrating the applicability of these algorithms for detection of various plasmodium species and their stages on both thin and thick blood smear microscopic images.

5 Methods and Materials

Automated malaria diagnosis will greatly help to control malaria endemic, especially in low resource setting areas by providing fast and accurate diagnosis results. The

question is, can the state-of-the-art deep learning based object detection algorithms accurately detect malaria parasites, to deploy them in remote areas of developing countries where there is critical shortage of laboratory technicians? In this study, one-stage and two-stage object detection algorithms are investigated for detection of *p. falciparum* from thick blood smear microscopic images. We used Faster R-CNN model from the two-stage object detectors. Similarly, the SSD, YOLOV3 and YOLOV4 from the one-stage object detection algorithms. YOLOV3 and YOLOV4 are the most powerful object detection algorithms in the one-stage category both in terms of detection accuracy and speed.

Since the size of malaria parasite in thick blood smear microscopic images is very small direct use of the detection algorithms, which are good at natural object detection, will not work well. In this study, we modified the parameters and the network architecture of the above detection algorithms to increase their performance for small object detection. To improve the small object detection performance of the YOLOV3 and YOLOV4 models we modified their network architecture to include more fine grained features from low resolution feature maps. The performance of the detection algorithms in terms of detection speed and detection accuracy were investigated for detection of *p. falciparum*.

5.1 Dataset

The malaria parasite detection dataset used in this study was taken from [49] which is available for researchers publicly. The dataset was collected using smartphone camera attached to the eyepiece of the microscope using a special attachment device developed for this purpose. The dataset contains a total of 1182 RGB microscopic images of thick blood smear slides which were stained with Field stain at x1000 magnification level, and all the images have a resolution of 750 x 750 pixels. The images are annotated by expert laboratory technicians. The dataset contains 948 infected images with 7628 *p.falciparum* parasites and 243 normal images with artifacts due to impurities. Since the dataset contains only *p.falciparum* our models were trained to detect this malaria parasite species only. All the detection models in this study were trained and evaluated using 90% of the dataset for training, among which we took 10% of it for validation. The remaining 10% was used to test the performance of the detection models. Figure 9 shows sample microscopic images which contain *p.falciparum* parasites and their corresponding bounding boxes and Table 3 shows the number of images and their statistics in both training and testing set. In the original dataset, there is no information about the number of field of views taken for a single patient or any other patient level information. Thus, we evaluated the performance of our detection models on image level.

Table 3: Statistics of of the dataset

	Training	Testing	Total
Number of images	1075	107	1182
Negative samples	214	20	243
Positive samples	861	87	948
Number of parasites	6619	1009	7628

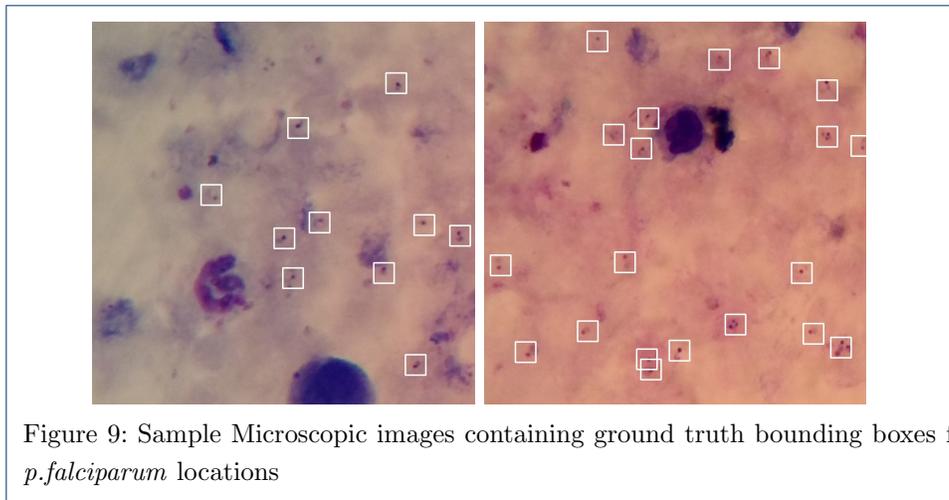


Figure 9: Sample Microscopic images containing ground truth bounding boxes for *p.falciparum* locations

5.2 Detection Architectures

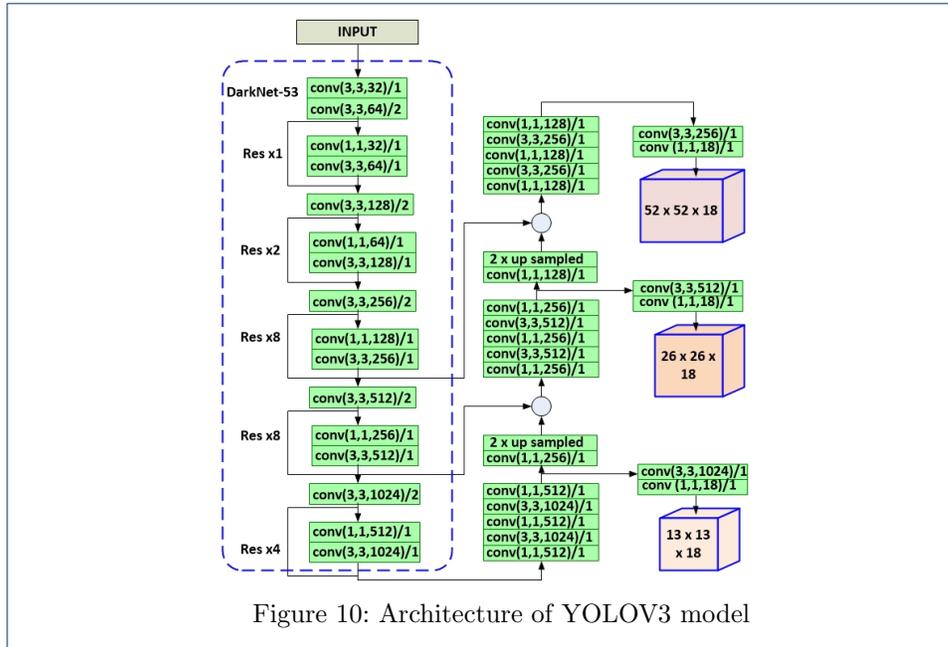
In this sub-section brief description of the detection models used in this study are provided. Next, the modified versions of the YOLOv3 and YOLOv4 models are presented. Lastly, the training process and hyperparameter optimization of the detection networks are discussed in detail.

5.2.1 YOLOV3: *You Only Look Once*

Currently, YOLOV3 [54] is one of the most powerful one-stage object detection algorithms in various applied research application areas. YOLOV3 remarkably improves the object detection accuracy, well positioning of the object bounding boxes and fast detection speed compared with its previous versions YOLOV1 [52], YOLOV2 [53]. YOLOV3 predicts the candidate object class probabilities and their bounding box coordinates from the input image using a single convolutional neural network. YOLOV3 benefited from applying new features such as the concept of Feature Pyramid Networks(FPN) [33] for multi-scale prediction, new backbone network called Darknet53 which uses residual networks for better feature extraction.

The network structure of YOLOV3 is modeled as a single regression problem having one backbone CNN and three object detection heads. Darknet53 contains one 3 x 3 convolution layer, 23 residual units (5 Resnet Blocks) having a 1 x 1 and 3 x 3 convolution layers. There are five stride by two convolution layers, at the beginning of each of the five Resnet blocks, to down-sample the input image by a factor of 32. Before each convolution layer there is batch normalization (BN) [24] followed by Leaky ReLU [41] activation function. Figure 10 shows the network architecture of YOLOV3.

In the YOLOV3 model, there are three detection heads called yolo layers. These three detection heads divide the input image into three different grids of size SxS. Each grid cell is responsible for the detection of objects whose center falls on that grid cell. A grid cell in a detection header is assigned with three anchor boxes whose center is the center of the grid cell. The bounding boxes have different aspect ratios and scales which is computed using the K-means clustering algorithm using the ground truth bounding box information. Small, medium, and large scale anchor

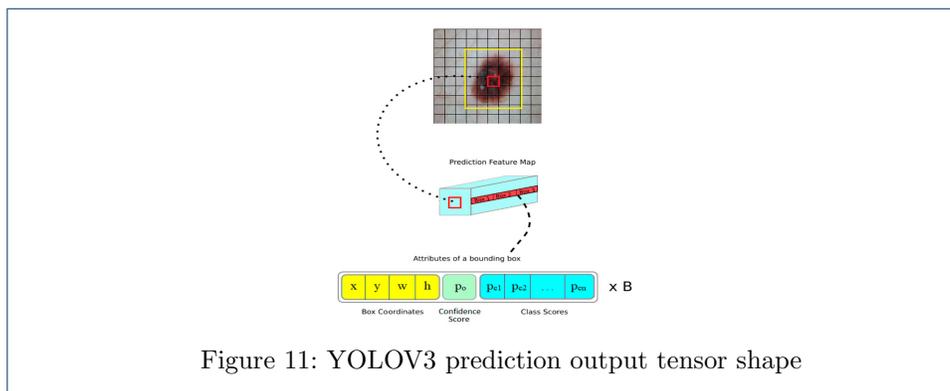


boxes are assigned to feature maps having larger, medium and small grid sizes respectively to make specific feature maps to detect particular sized objects.

To detect objects of different size YOLOV3 use a modified version of FPN to extract both spatial and semantic information-rich feature maps at three different scales. The first feature map is generated as the final output of the Darknet53 backbone network. To generate the second feature map, an additional 1×1 , 3×3 , 1×1 , 3×3 , 1×1 , 1×1 convolution operations are done on the first feature scale and up-sampled by a factor of two. Then the up-sampled feature map is concatenated with a feature map generated from the fourth strided by 2 convolution. The third feature map is generated from the second feature map by performing the same number of convolution operations with different kernel sizes and up-sampled by a factor of two. Then the up-sampled feature map is concatenated with the feature map generated from the third strided by 2 convolution output.

In the YOLOV3 model for an input image of resolution 416×416 the feature map size will be 13×13 , 26×26 , and 52×52 for the first, second and third detection layers respectively. After generating the three feature maps YOLOV3 performs 6 (1×1 and 3×3) convolution operations before the detection layer to further refine the features. At each of the three detection layers, YOLOV3 generates a feature tensor of shape $N \times N$ ($B \times (4 + 1 + C)$) as shown in Figure 11. $N \times N$ indicates the grid size of that particular detection layer feature map. B represents the number of bounding boxes predicted in a single grid cell, which is 3 in the case of the YOLOV3 model. Four offset values of coordinates (center point, width, and height) t_x, t_y, t_w, t_h for each of the three bounding boxes and their confidence score. The confidence score tells how probable an object exists $P_r(Object)$ (which is either 1 when there is an object or 0 when there is no object) and the confidence of its precision IOU_{Pred}^{Truth} (which is the intersection over union (IOU) ratio between the predicted bounding box and the matched anchor box). C indicates the conditional class probabilities

$P_r(Class_i|Object)$ of the bounding boxes predicted by the grid cell. In this study $C=1$ since we detect one malaria parasite species which is *p. falciparum* from the input image. YOLOV3 predicts offset values relative to the grid cell, rather than



the actual dimensions of the bounding boxes. Assuming the center of the grid cell responsible for predicting the bounding box is offset from the top-left corner of the image by (C_x, C_y) and the assigned anchor box has width and height as p_w and p_h then the actual bounding box dimensions (b_x, b_y, b_w, b_h) are calculated by the formula given below:

$$\begin{aligned}
 b_x &= \sigma(t_x) + C_x \\
 b_y &= \sigma(t_y) + C_y \\
 b_w &= p_w \exp(t_w) \\
 b_h &= p_h \exp(t_h)
 \end{aligned}
 \tag{1}$$

Where σ indicates a sigmoid function is used to make the predicted bounding box center to be within the grid cell. b_x and b_y represent the center of the predicted bounding box. b_w and b_h represent the width and height of the anchor box.

During training, a binary cross-entropy loss function is used for the confidence score, and the class probabilities while a mean squared error loss is used for the prediction of the coordinates of the bounding box. The classification and localization error loss will be calculated only for those anchor boxes responsible for detecting objects. The general loss function in the YOLOV3 model has three parts and is defined as follows:

$$\begin{aligned}
\text{Total loss} &= \text{Loss}_{loc} + \text{Loss}_{obj} + \text{Loss}_{cls} \\
&= \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\
&\quad + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\
&\quad + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} l(C_i - \hat{C}_i) + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} l(C_i - \hat{C}_i) \\
&\quad + \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in \text{classes}} l(p_i(c) - \hat{p}_i(c))
\end{aligned} \tag{2}$$

Where 1_{ij}^{obj} indicates that the predicted bounding box is represented by the i_{th} grid cell using the j_{th} anchor box. 1_i^{obj} represents if object is found in the i_{th} grid cell. λ_{coord} and λ_{noobj} are optimization parameters to increase the localization error and to decrease the confidence error for anchor boxes with no matched object respectively. S represents the grid size of the detection feature map and B represents the number of anchor boxes assigned for a given grid cell. $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$ represent the four predicted bounding box coordinates and x_i, y_i, w_i, h_i represent the ground truth anchor box coordinates. \hat{C}_i is the predicted confidence score whereas C_i is the actual confidence score value. $\hat{p}_i(c)$ indicates the predicted class probabilities of the box and $p_i(c)$ is the actual class probability for the ground truth object.

5.2.2 YOLOV4: You Only Look Once

YOLOV4 proposed by [3] comes with the state-of-the-art performance compared with other current deep learning based object detection techniques such as EfficientDet [63] and YOLOV3 [54]. To achieve such high detection speed and high detection accuracy in real-time, the authors combine different features investigated by other researchers which can enhance the performance of deep learning based object detectors with little or no computation cost during inference. The features added in the YOLOV4 architecture are classified into two categories of methods called Bag of Freebies(BOF) and Bag of Specials(BOS). The BOF and BOS are applied both at the backbone and the detector modules. In the BOS part, they tightly coupled SPP with the backbone network for improving the receptive field size of the detection layer feature maps. The backbone of YOLOV4 is called CSP-Darknet53 which is based upon the cross-stage-partial connection for enhancing feature robustness by connecting feature maps at the different stages of the backbone and head network with minimal computation cost. Modified PANet [35] is used for enriching the semantic and geometric information of extracted features by concatenating lower-level features maps with higher-level feature maps in contrast to the FPN used by YOLOV3. A new activation technique called Mish is used in contrast to YOLOV3's Leaky ReLU with minimal computation coast within the BOS. Most of the BOF comes from data augmentation techniques such as Mosaic, CutMix, and self-adversarial training (SAT). The authors of YOLOV4 use Cross Mini-Batch Normalization(CmBN) for normalizing batches of training images and

5.2.3 Modified YOLOV3 and YOLOV4 Architectures

The authors in YOLOV3 and YOLOV4 train and evaluate the network architectures using Imagenet [9], Pascal VOC [15], and MS COCO [32] datasets. These datasets containing natural images that are far bigger than malaria parasite, and without modifying the network architecture and parameter optimization the original versions of YOLOV3 and YOLOV4 have a very low performance on the detection of such small objects. In this study, we modify the network architecture of YOLOV3 into two separate architectures which we call them YOLOV3-MOD1 and YOLOV3-MOD2. In the original YOLOV3 model feature maps at a scale of 13 x 13, 26 x 26, and 52 x 52 are used at the three detection layers. However, the feature maps in the original YOLOV3 model have bigger receptive fields than the *p. falciparum* size.

In YOLOV3-MOD1 we change the shallow feature map scale to 104 x 104 which is better in the detection of *p. falciparum* compared with the original YOLOV3. The third detection layer feature map in our modified YOLOV3-MOD1 is generated by concatenating the second strided by 2 convolution in the Darknet-53 backbone network and four times up-sampled feature from the second detection layer feature map. Figure 13 shows the network structure of YOLOV3-MOD1. As shown in the figure the modified YOLOV3-MOD1 has detection layer feature maps at 13 x 13, 26 x 26, and 104 x 104. The modified multi-scale scale feature map extracts better information for the detection of *p. falciparum* with a minimal added computational cost.

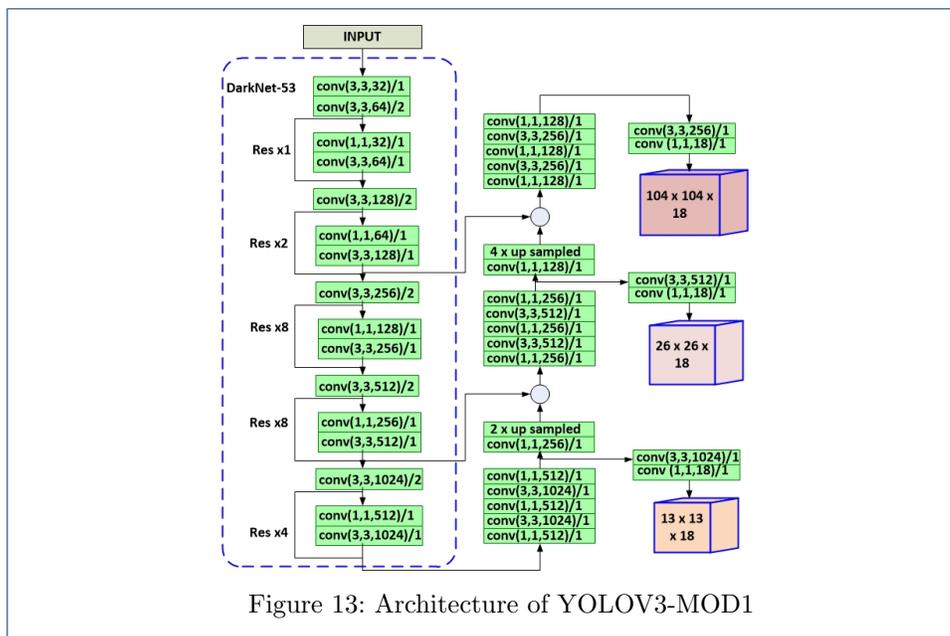


Figure 13: Architecture of YOLOV3-MOD1

In the second modification to YOLOV3 which is YOLOV3-MOD2, we add a fourth detection layer in addition to the already existing three detection layers in the original YOLOV3 model. The added fourth detection layer increases the performance of YOLOV3 for small object detection since a short connection of deeper features with shallow features will enhance the fine grained information of the added detection layer. The proposed YOLOV3-MOD2 model is shown in Figure

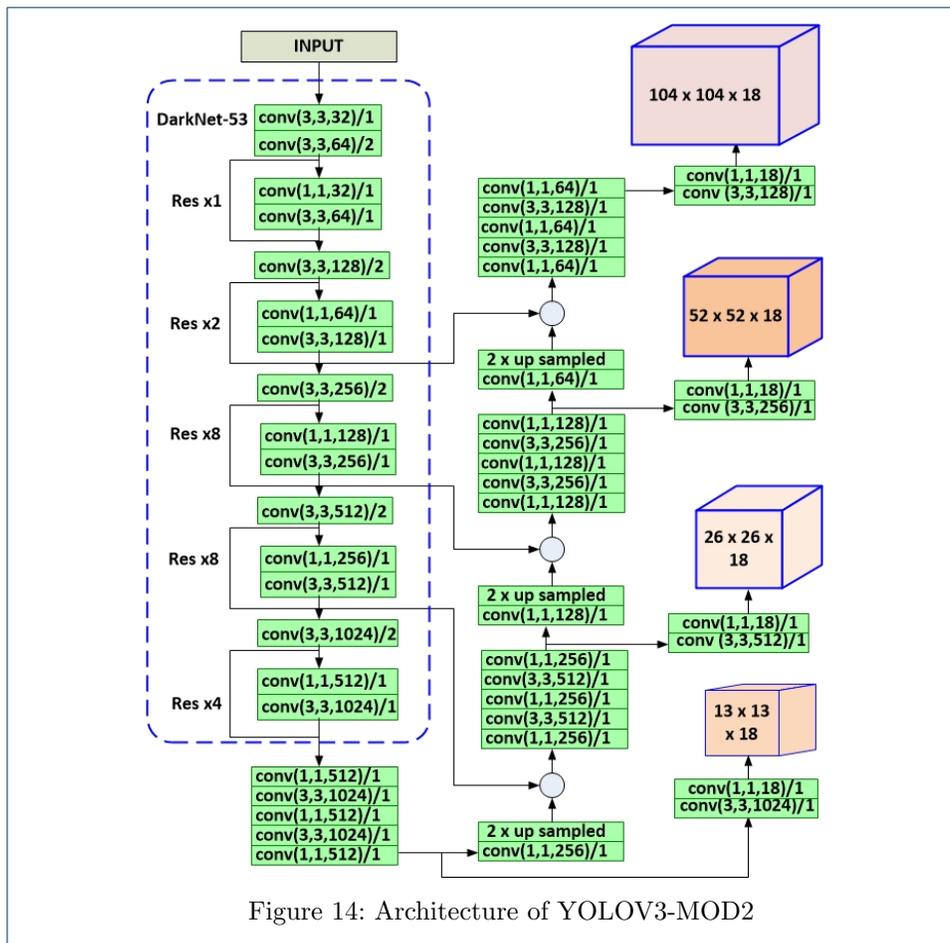
14. YOLOV3-MOD2 has four detection feature maps with size 13 x 13, 26 x 26, 52 x 52, and 104 x 104. We have also added three additional anchor box sizes for the newly added detection layer. We have used the K-means clustering algorithm to generate the 9 new anchor box sizes for YOLOV3-MOD1 and 12 new anchor box sizes for YOLOV3-MOD2 based on the ground truth bounding box coordinate information from our malaria dataset. These two modified YOLOV3 models improve the performance of *p.falciparum* detection compared with the original YOLOV3 version. The results are discussed in detail in section 3. YOLOV3-MOD2 has better detection accuracy compared with YOLOV3-MOD1 but it has less detection speed during inference.

For the modified YOLOV4 which we call it YOLOV4-MOD in the rest of the paper, we added a fourth detection layer based on the original YOLOV4 model. The added detection layer enables the modified YOLOV4 model to obtain more geometric features that are concatenated with deeper level features using the PANet architecture. By adding a fourth detection layer we can obtain more comprehensive features that enhances the performance of YOLOV4-MOD to detect small objects compared with the original YOLOV4 mode. For the modified YOLOV4 architecture we have generated 12 new anchor box sizes which are evenly distributed to the four detection layers based on their size. YOLOV4-MOD shows better detection performance in detecting *p.falciparum* in our experimental results.

5.2.4 *Training and Hyper-parameter Optimization of the Detection Models*

In this study, we investigated the applicability of three deep learning based object detectors for malaria parasite detection including Faster R-CNN, SSD, and two variants of YOLO: YOLOV3 and YOLOV4. During the training phase, we used a pre-trained weight for each of the detection models by experimentally selecting the best pre-trained weight using our malaria dataset. Afterwards we re-train the detection networks to adapt to our malaria parasite detection task using fine-tuning.

To train the YOLO based algorithms (original YOLOV3 and YOLOV4, YOLOV3-MOD1, YOLOV3-MOD2, YOLOV4-MOD) we have used the Darknet Framework. We train all the YOLO based algorithms for 4,000 iterations using stochastic gradient descent (SGD) optimization with Adam optimizer. We have used an initial learning rate of 0.001, a batch size of 64 with subdivision 32, momentum of 0.9 with weight decay of 0.0005. The learning rate was reduced at 3200 and 3600 iterations by multiplying it by 0.1 respectively. A weight file pre-trained using ImageNet [9] dataset for YOLOV3 based models and MS COCO [32] dataset for YOLOV4 based models is fine-tuned using our malaria dataset. We have trained the YOLO based algorithms with input image sizes of 416 x 416 and 608 x 608 to achieve better detection performance as the size of the input image increases. During YOLOV3 and YOLOV4 training multi-scale training is enabled by changing the input resolution in the range (320 x 320) to (896 x 896) every 10 training steps on the fly. This enables YOLO to detect objects at different image resolutions. We use the default data augmentation parameters used in YOLOV3 configuration file with saturation value = 1.5, hue = 0.1 and exposure = 1.5. For YOLOV4 models mosaic data augmentation technique is used as a default parameter. At training time a 0.5 IOU



threshold and 0.25 confidence threshold values are used for updating the regression and class probability loss during training. We have used modified anchor box sizes generated using K-means clustering algorithms based on ground truth bounding box coordinate values in our malaria dataset. For original YOLOV3, YOLOV4, and YOLOV3-MOD1 we use 9 new anchor sizes based on our malaria dataset. For YOLOV3-MOD2 and YOLOV4-MOD we generated 12 new anchor sizes and evenly distribute them to the four detection layers according to their size.

For Faster R-CNN and SSD, we did not modify the network architecture rather we modify the network parameters to increase the performance of the detectors in our malaria parasite detection task. We have changed different hyper-parameters empirically for Faster R-CNN and SSD models based on their default configuration in Tensorflow Object detection API. The main hyperparameter for these two detectors is their backbone feature extractor network. We experimentally selected Inception-v2 as a feature extractor for both Faster R-CNN and SDD which is pre-trained on MS COCO dataset.

For Faster R-CNN we select aspect ratios of 1:1, 1:2 and 2:1, and scales of [0.1, 0.15, 0.2, 0.25] since the default anchor box sizes are much bigger than the malaria parasite size. We fine-tune a pre-trained Faster R-CNN model using our malaria dataset with an initial learning rate of 0.001 which is reduced by 0.1 factor at training iterations of 65,000 and 85,000 and optimized by momentum SGD. We

train the Faster R-CNN model for a total of 100K iterations. For SSD we use a modified minimum scale of 0.1 and a maximum scale of 0.9 to generate suitable anchor sizes for our malaria dataset. We trained SSD with 300 x 300 input size. We made additional experiments to select the optimal initial learning rate of 0.001 with batch size of 24, and RMS Prop optimizer to train both the SSD models for 100K iterations. For both SSD and Faster R-CNN models, we adopt a drop out unit with a value of 0.5 to avoid the problem of over-fitting. For all the detection models the experimental environment was done on cloud service with Tesla T4 GPU and 12GB GDDR5 VRAM.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Fetulhak Abdurahman: Conceptualization of the study, Methodology, Experimental design and Analysis, Document Preparation. **Kinde Anlay:** Methodology, Writing Result and Discussion, Revision of the document. **Mohammed Aliy:** Data Analysis and Revision of the document. All the authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

No informed consent or ethical approval is required for the purpose of this study.

Availability of data and materials

The malaria parasite microscopic dataset used in this study are included in [49] and this is the link for the dataset [<http://air.ug/downloads/plasmodium-phoncamera.zip>].

Funding

Not applicable.

Author details

¹Faculty of Electrical and Computer Engineering, Jimma University, 378 Jimma, Ethiopia. ²Faculty of Electrical and Computer Engineering, Jimma University, , 378 Jimma, Ethiopia. ³School of Biomedical Engineering, Jimma University, 378 Jimma, Ethiopia.

References

1. Abdul Nasir, A.S., Mashor, M.Y., Mohamed, Z.: Segmentation based approach for detection of malaria parasites using moving k-means clustering. In: 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, pp. 653–658 (2012). doi:[10.1109/IECBES.2012.6498073](https://doi.org/10.1109/IECBES.2012.6498073)
2. Bibin, D., Nair, M.S., Punitha, P.: Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access* **5**, 9099–9108 (2017)
3. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: Yolov4: Optimal speed and accuracy of object detection. *ArXiv abs/2004.10934* (2020)
4. Chang, P.D., Kuoy, E., Grinband, J., Weinberg, B.D., Thompson, M., Homo, R., Chen, J., Abcede, H., Shafie, M., Sugrue, L.P., Filippi, C.G., Su, M.-Y., Yu, W., Hess, C.P., Chow, D.S.: Hybrid 3d/2d convolutional neural network for hemorrhage evaluation on head ct. *AJNR. American journal of neuroradiology* **39** **9**, 1609–1616 (2018)
5. Yang, F.-c., Quizon, N., Silamut, K., Maude, R.J., Jäger, S., Antani, S.K.: Cascading yolo: Automated malaria parasite detection for plasmodium vivax in thin blood smears. (2019)
6. Chibuta, S., Acar, A.C.: Real-time malaria parasite screening in thick blood smears for low-resource setting. *Journal of Digital Imaging*, 1–13 (2020)
7. Das, D.K., Ghosh, M., Pal, M., Maiti, A.K., Chakraborty, C.: Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron* **45**, 97–106 (2013). doi:[10.1016/j.micron.2012.11.002](https://doi.org/10.1016/j.micron.2012.11.002)
8. Delahunt, C.B., Jaiswal, M.S., Horning, M.P., Janko, S., Thompson, C.M., Kulhare, S., Hu, L., Ostbye, T., Yun, G., Gebrehiwot, R., Wilson, B.K., Long, E., Proux, S., Gamboa, D., Chiodini, P., Carter, J., Dhorda, M., Isaboke, D., Ogotu, B., Oyibo, W., Villasis, E., Tun, K.M., Bachman, C., Bell, D., Mehanian, C.: Fully-automated patient-level malaria assessment on field-prepared thin blood film microscopy images, including supplementary information. *CoRR abs/1908.01901* (2019). [1908.01901](https://arxiv.org/abs/1908.01901)
9. Deng, J., Dong, W., Socher, R., Li, L., Kai Li, Li Fei-Fei: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
10. Devi, S.S., Herojit Singh, N., Hussain Laskar, R.: Performance analysis of various feature sets for malaria-infected erythrocyte detection. In: Das, K.N., Bansal, J.C., Deep, K., Nagar, A.K., Pathipooranam, P., Naidu, R.C. (eds.) *Soft Computing for Problem Solving*, pp. 275–283. Springer, Singapore (2019)

11. Devi, S.S., Laskar, R.H., Sheikh, S.A.: Hybrid classifier based life cycle stages analysis for malaria-infected erythrocyte using thin blood smear images. *Neural Computing and Applications* **29**(8), 217–235 (2018). doi:[10.1007/s00521-017-2937-4](https://doi.org/10.1007/s00521-017-2937-4)
12. Dong, Y., Jiang, Z., Shen, H., David Pan, W., Williams, L.A., Reddy, V.V.B., Benjamin, W.H., Bryan, A.W.: Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In: 2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), pp. 101–104 (2017). doi:[10.1109/BHI.2017.7897215](https://doi.org/10.1109/BHI.2017.7897215)
13. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM* **15**, 11–15 (1972)
14. El-Melegy, M., Mohamed, D., ElMelegy, T.: Automatic detection of tuberculosis bacilli from microscopic sputum smear images using faster r-cnn, transfer learning and augmentation. In: Morales, A., Fierrez, J., Sánchez, J.S., Ribeiro, B. (eds.) *Pattern Recognition and Image Analysis*, pp. 270–278. Springer, Cham (2019)
15. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html> (2012)
16. Ezhilarasi, R., Varalakshmi, P.: Tumor detection in the brain using faster r-cnn. In: 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)-I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference On, pp. 388–392 (2018). doi:[10.1109/I-SMAC.2018.8653705](https://doi.org/10.1109/I-SMAC.2018.8653705)
17. Girshick, R.: Fast r-cnn. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448 (2015)
18. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
19. Gopakumar, G.P., Swetha, M., Siva, G.S., Subrahmanyam, G.R.K.S.: Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. *Journal of biophotonics* **11** 3 (2018)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
21. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *Computer Vision – ECCV 2014*, pp. 346–361. Springer, Cham (2014)
22. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
23. Hung, J., Goodman, A., Lopes, S., Rangel, G., Ravel, D., Costa, F., Duraisingh, M.T., Marti, M., Carpenter, A.E.: Applying faster r-cnn for object detection on malaria images. *ArXiv abs/1804.09548* (2017)
24. Ioffe, S., Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift (2015). [1502.03167](https://arxiv.org/abs/1502.03167)
25. Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R.: A survey of deep learning-based object detection. *IEEE Access* **7**, 128837–128868 (2019)
26. Kaewkamnerd, S., Uthaiipibull, C., Intarapanich, A., Pannarut, M., Chaotheing, S., Tongsim, S.: An automatic device for detection and classification of malaria parasite species in thick blood film. *BMC Bioinformatics* **13** (2012). doi:[10.1186/1471-2105-13-S17-S18](https://doi.org/10.1186/1471-2105-13-S17-S18)
27. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)
28. Kutlu, H., Avci, E., Özyurt, F.: White blood cells detection and classification based on regional convolutional neural networks. *Medical Hypotheses* **135**, 109472 (2020). doi:[10.1016/j.mehy.2019.109472](https://doi.org/10.1016/j.mehy.2019.109472)
29. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. (1998)
30. Lemay, A.: Kidney recognition in ct using yolov3. *ArXiv abs/1910.01268* (2019)
31. Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., Guo, P., Hossain, M.A., Antani, S.K., Maude, R.J., Huang, X., Jaeger, S., Thoma, G.R.: Cnn-based image analysis for malaria diagnosis. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 493–496 (2016)
32. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *Computer Vision – ECCV 2014*, pp. 740–755. Springer, Cham (2014)
33. Lin, T., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J.: Feature pyramid networks for object detection. *CoRR abs/1612.03144* (2016). [1612.03144](https://arxiv.org/abs/1612.03144)
34. Liu, M., Jiang, J., Wang, Z.: Colonic polyp detection in endoscopic videos with single shot detection based deep convolutional neural network. *IEEE Access* **7**, 75058–75066 (2019). doi:[10.1109/ACCESS.2019.2921027](https://doi.org/10.1109/ACCESS.2019.2921027)
35. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. *CoRR abs/1803.01534* (2018). [1803.01534](https://arxiv.org/abs/1803.01534)
36. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: Ssd: Single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *Computer Vision – ECCV 2016*, pp. 21–37. Springer, Cham (2016)
37. Liu, W., Cheng, L., Meng, D.: Brain slices microscopic detection using simplified ssd with cycle-gan data augmentation. In: Cheng, L., Leung, A.C.S., Ozawa, S. (eds.) *Neural Information Processing*, pp. 454–463. Springer, Cham (2018)
38. Lo, Y.-C., Juang, C.-F., Chung, I.-F., Guo, S.-N., Huang, M.-L., Wen, M.-C., Lin, C.-J., Lin, H.-Y.: Glomerulus detection on light microscopic images of renal pathology with the faster r-cnn. In: Cheng, L., Leung, A.C.S., Ozawa, S. (eds.) *Neural Information Processing*, pp. 369–377. Springer, Cham (2018)
39. Loddó, A., Di Ruberto, C., Kocher, M.: Recent advances of malaria parasites detection systems based on

- mathematical morphology. *Sensors (Basel, Switzerland)* **18** (2018). doi:[10.3390/s18020513](https://doi.org/10.3390/s18020513)
40. Ma, J., Li, X., Li, H., Menze, B.H., Liang, S., Zhang, R., Zheng, W.-S.: Group-attention single-shot detector (ga-ssd): Finding pulmonary nodules in large-scale ct images. In: Cardoso, M.J., Feragen, A., Glocker, B., Konukoglu, E., Oguz, I., Unal, G., Vercauteren, T. (eds.) *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning. Proceedings of Machine Learning Research*, vol. 102, pp. 358–369. PMLR, London, United Kingdom (2019). <http://proceedings.mlr.press/v102/ma19a.html>
 41. Maas, A.L.: Rectifier nonlinearities improve neural network acoustic models. (2013)
 42. Mas, D., Ferrer, B., Cojoc, D., Finaurini, S., Mico, V., Garcia, J., Zalevsky, Z.: Novel image processing approach to detect malaria. *Optics Communications* **350**, 13–18 (2015). doi:[10.1016/j.optcom.2015.03.064](https://doi.org/10.1016/j.optcom.2015.03.064)
 43. Moon, S., Lee, S., Kim, H., Freitas-Junior, L.H., Kang, M., Ayong, L., Hansen, M.A.E.: An image analysis algorithm for malaria parasite stage classification and viability quantification. *PLOS ONE* **8**(4), 1–12 (2013). doi:[10.1371/journal.pone.0061812](https://doi.org/10.1371/journal.pone.0061812)
 44. O'Meara, W., Barcus, M., Wongsrichanalai, C., Sinuon, M., Maguire, J., Jordan, R., Prescott, W., McKenzie, F.: Reader technique as a source of variability in determining malaria parasite density by microscopy. *Malaria journal* **5**, 118 (2006). doi:[10.1186/1475-2875-5-118](https://doi.org/10.1186/1475-2875-5-118)
 45. Pan, W.D., Dong, Y., Wu, D.: Classification of malaria-infected cells using deep convolutional neural networks. In: Farhadi, H. (ed.) *Machine Learning. IntechOpen, Rijeka* (2018). Chap. 8. doi:[10.5772/intechopen.72426](https://doi.org/10.5772/intechopen.72426). <https://doi.org/10.5772/intechopen.72426>
 46. Park, H.S., Rinehart, M.T., Walzer, K.A., Chi, J.-T.A., Wax, A.: Automated detection of p. falciparum using machine learning algorithms with quantitative phase images of unstained cells. *PLOS ONE* **11**(9), 1–19 (2016). doi:[10.1371/journal.pone.0163045](https://doi.org/10.1371/journal.pone.0163045)
 47. Poostchi, M., Ersoy, I., McMenamin, K., Gordon, E., Palaniappan, N., Pierce, S., Maude, R.J., Bansal, A., Srinivasan, P., Miller, L., Palaniappan, K., Thoma, G., Jaeger, S.: Malaria parasite detection and cell counting for human and mouse using thin blood smear microscopy. *Journal of Medical Imaging* **5**(4), 1–13 (2018). doi:[10.1117/1.JMI.5.4.044506](https://doi.org/10.1117/1.JMI.5.4.044506)
 48. Purwar, Y., Shah, S., Clarke, G., Almugairi, A., Muehlenbachs, A.: Automated and unsupervised detection of malaria parasites in microscopic images. *Malaria journal* **10**, 364 (2011). doi:[10.1186/1475-2875-10-364](https://doi.org/10.1186/1475-2875-10-364)
 49. Quinn, J., Andama, A., Munabi, I., Kiwanuka, F.: *Automated Blood Smear Analysis for Mobile Malaria Diagnosis*, pp. 115–132 (2014)
 50. Rahman, A., Zunair, H., Rahman, M.S., Yuki, J.Q., Biswas, S., Alam, M.A., Alam, N.B., Mahdy, M.R.C.: Improving malaria parasite detection from red blood cell using deep convolutional neural networks. *ArXiv abs/1907.10418* (2019)
 51. Rajaraman, S., Antani, S.K., Poostchi, M., Silamut, K., Hossain, M.A., Maude, R.J., Jaeger, S., Thoma, G.R.: Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ* **6**, 4568 (2018). doi:[10.7717/peerj.4568](https://doi.org/10.7717/peerj.4568)
 52. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
 53. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. *CoRR abs/1612.08242* (2016). [1612.08242](https://arxiv.org/abs/1612.08242)
 54. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. *CoRR abs/1804.02767* (2018). [1804.02767](https://arxiv.org/abs/1804.02767)
 55. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(6), 1137–1149 (2017)
 56. Rosado, L., da Costa, J.M.C., Elias, D., Cardoso, J.S.: Automated detection of malaria parasites on thick blood smears via mobile devices. *Procedia Computer Science* **90**, 138–144 (2016). doi:[10.1016/j.procs.2016.07.024](https://doi.org/10.1016/j.procs.2016.07.024). 20th Conference on Medical Image Understanding and Analysis (MIUA 2016)
 57. Ross, N.E., Pritchard, C.J., Rubin, D.M., Dusé, A.G.: Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Medical and Biological Engineering and Computing* **44**(5), 427–436 (2006). doi:[10.1007/s11517-006-0044-2](https://doi.org/10.1007/s11517-006-0044-2)
 58. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., Li, F.-F.: Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* **115**, 211–252 (2015)
 59. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2015)
 60. Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., Ibragimov, B.: Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers & Electrical Engineering* **78**, 388–399 (2019). doi:[10.1016/j.compeleceng.2019.08.004](https://doi.org/10.1016/j.compeleceng.2019.08.004)
 61. Sivaramkrishnan, R., Antani, S., Jaeger, S.: Visualizing deep learning activations for improved malaria cell classification. In: Fodeh, S., Raicu, D.S. (eds.) *Proceedings of The First Workshop Medical Informatics and Healthcare Held with the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining. Proceedings of Machine Learning Research*, vol. 69, pp. 40–47. PMLR, ??? (2017). <http://proceedings.mlr.press/v69/sivaramkrishnan17a.html>
 62. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. *CoRR abs/1409.4842* (2014). [1409.4842](https://arxiv.org/abs/1409.4842)
 63. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. *ArXiv abs/1911.09070* (2019)
 64. Tang, W., Zou, D., Yang, S., Shi, J.: Dsl: Automatic liver segmentation with faster r-cnn and deeplab. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds.) *Artificial Neural Networks and Machine Learning – ICANN 2018*, pp. 137–147. Springer, Cham (2018)
 65. Tek, F.B., Dempster, A.G., Kale, I.: Parasite detection and identification for automated thin blood film malaria diagnosis. *Computer Vision and Image Understanding* **114**(1), 21–32 (2010). doi:[10.1016/j.cviu.2009.08.003](https://doi.org/10.1016/j.cviu.2009.08.003)
 66. Torres, K., Bachman, C.M., Delahunt, C.B., Alarcon Baldeon, J., Alava, F., Gamboa Vilela, D., Proux, S., Mehanian, C., McGuire, S.K., Thompson, C.M., Ostbye, T., Hu, L., Jaiswal, M.S., Hunt, V.M., Bell, D.: Automated microscopy for routine malaria diagnosis: a field comparison on giemsa-stained blood films in peru.

- Malaria Journal **17**(1), 339 (2018). doi:[10.1186/s12936-018-2493-0](https://doi.org/10.1186/s12936-018-2493-0)
67. Ünver, H.M., Ayan, E.: Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm. In: *Diagnostics* (2019)
 68. van Rijthoven, M., Swiderska-Chadaj, Z., Seeliger, K., van der Laak, J., Ciampi, F.: You only look on lymphocytes once (2018)
 69. Vijayalakshmi, Kanna, R.: Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications* (2019). doi:[10.1007/s11042-019-7162-y](https://doi.org/10.1007/s11042-019-7162-y)
 70. WHO: World malaria report 2019. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/publications-detail/world-malaria-report-2019>. [Online; accessed 30-April-2020] (2019)
 71. Xiao, J., Zhang, Y., Bian, K., Zhou, G., Yan, W.: Denxfpn: Pulmonary pathologies detection based on dense feature pyramid networks. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1234–1238 (2019). doi:[10.1109/ICASSP.2019.8683021](https://doi.org/10.1109/ICASSP.2019.8683021)
 72. Yang, F., Poostchi, M., Yu, H., Zhou, Z., Silamut, K., Yu, J., Maude, R.J., Jaeger, S., Antani, S.: Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics*, 1–1 (2019). doi:[10.1109/JBHI.2019.2939121](https://doi.org/10.1109/JBHI.2019.2939121)
 73. Yang, F., Poostchi, M., Yu, H., Zhou, Z., Silamut, K., Yu, J., Maude, R., Jaeger, S., Antani, S.: Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics* **PP**, 1–1 (2019). doi:[10.1109/JBHI.2019.2939121](https://doi.org/10.1109/JBHI.2019.2939121)
 74. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. *CoRR* **abs/1311.2901** (2013). [1311.2901](https://arxiv.org/abs/1311.2901)
 75. Zhang, C., Chang, C., Jamshidi, M.: Bridge damage detection using single-stage detector and field inspection images. *CoRR* **abs/1812.10590** (2018). [1812.10590](https://arxiv.org/abs/1812.10590)

Figures

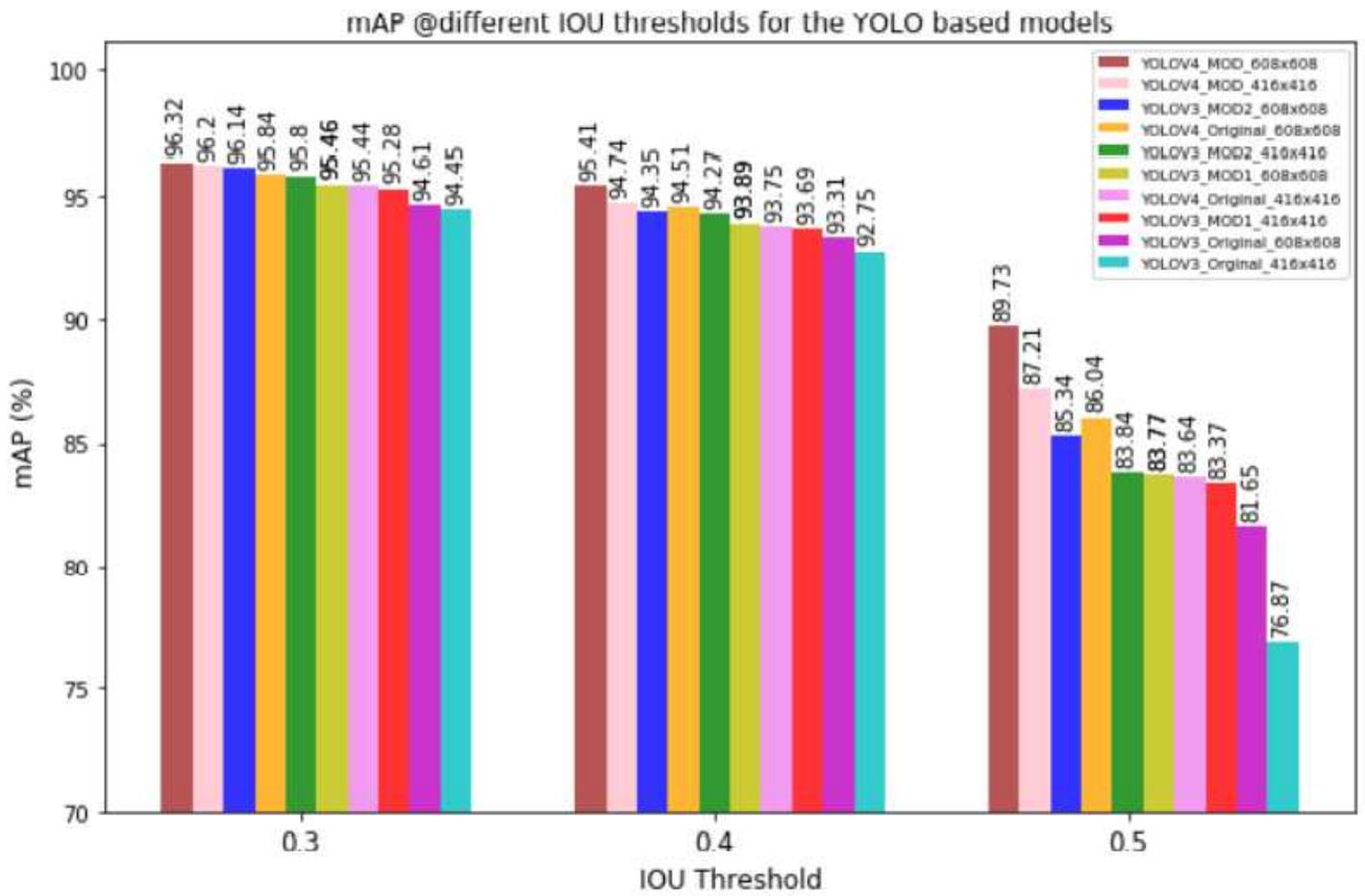


Figure 1

Effect of IOU threshold on model detection

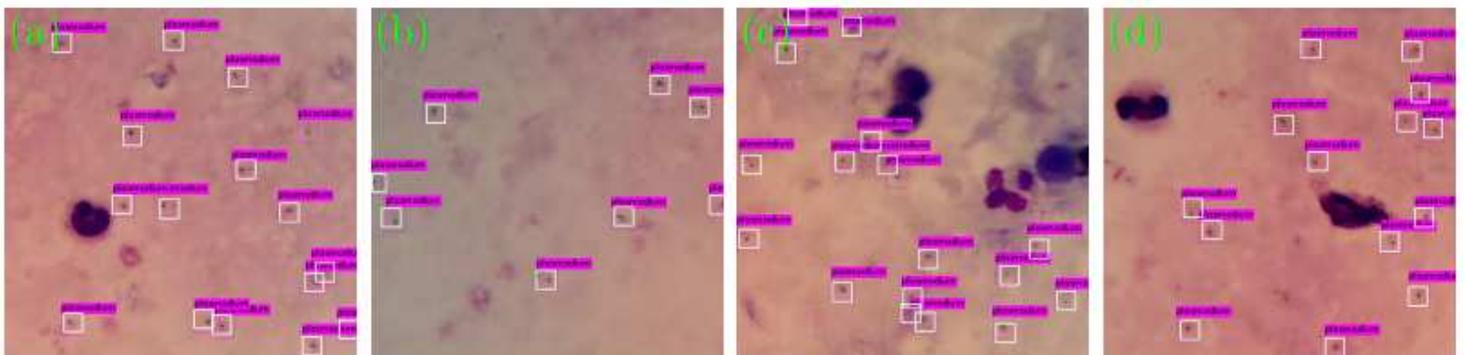


Figure 2

YOLOV4-MOD @608 x 608

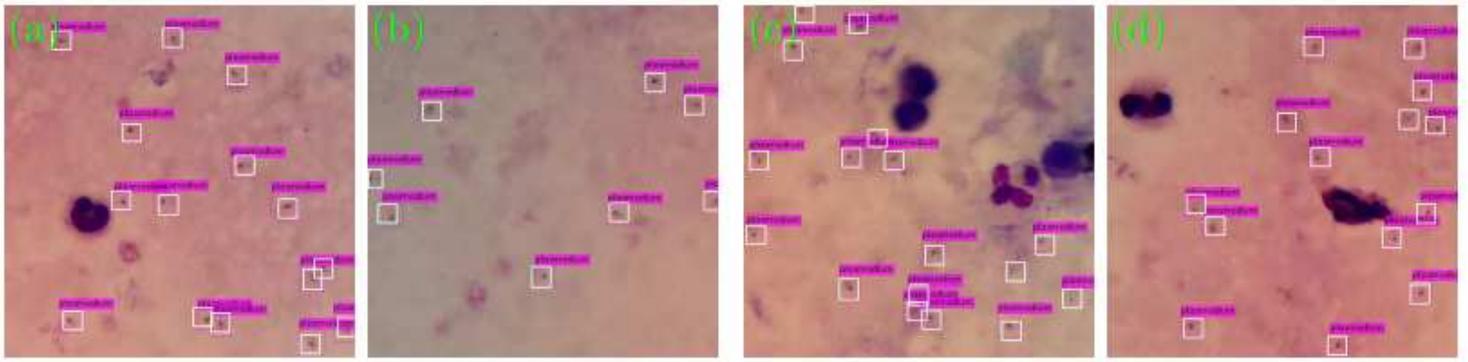


Figure 3

YOLOV3-MOD2 @608 x 608

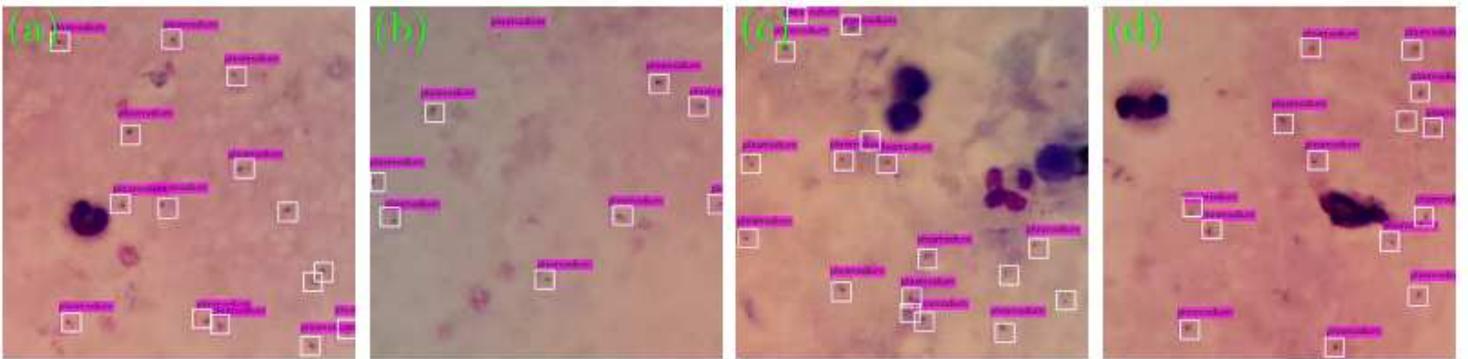


Figure 4

YOLOV3-MOD1 @608 x 608

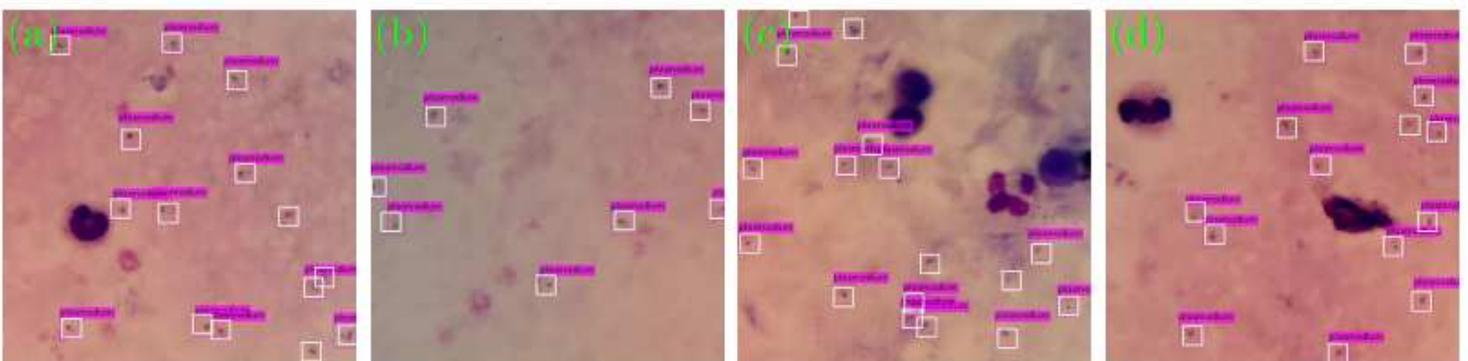


Figure 5

YOLOV4-Original @608 x 608

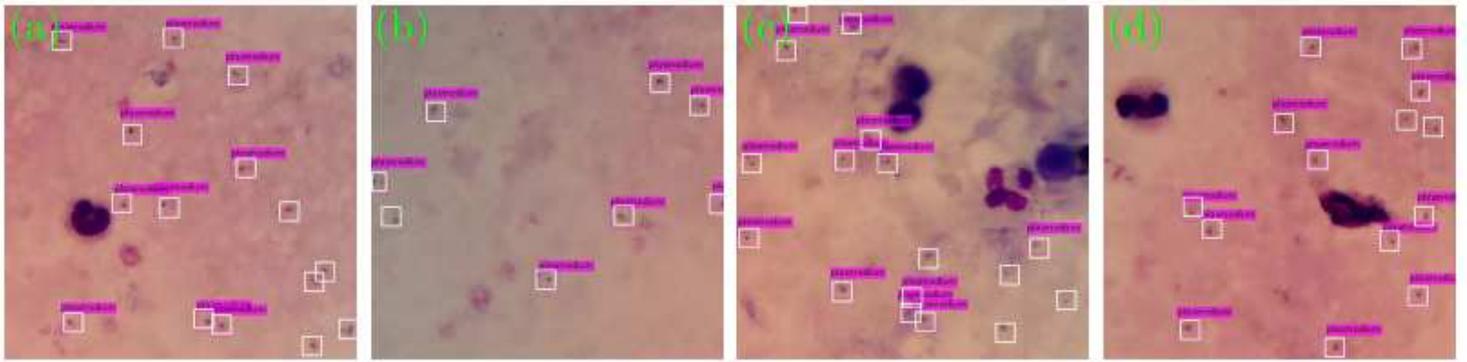


Figure 6

YOLOV3-Original @608 x 608

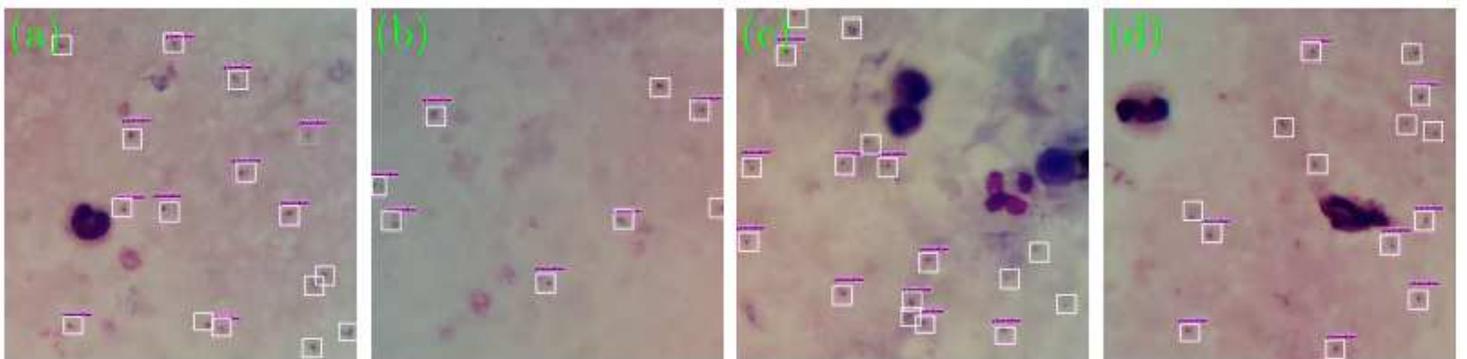


Figure 7

SSD @300 x 300

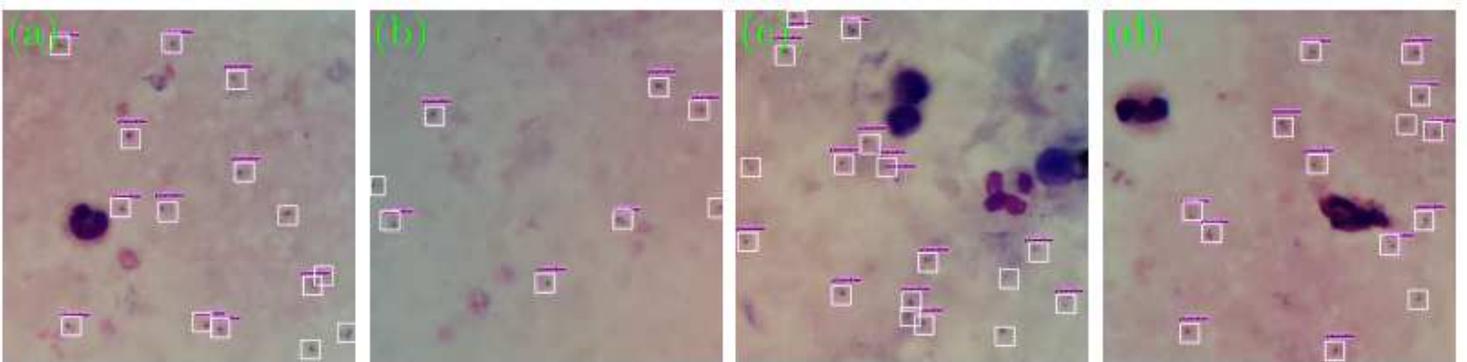


Figure 8

Faster R-CNN

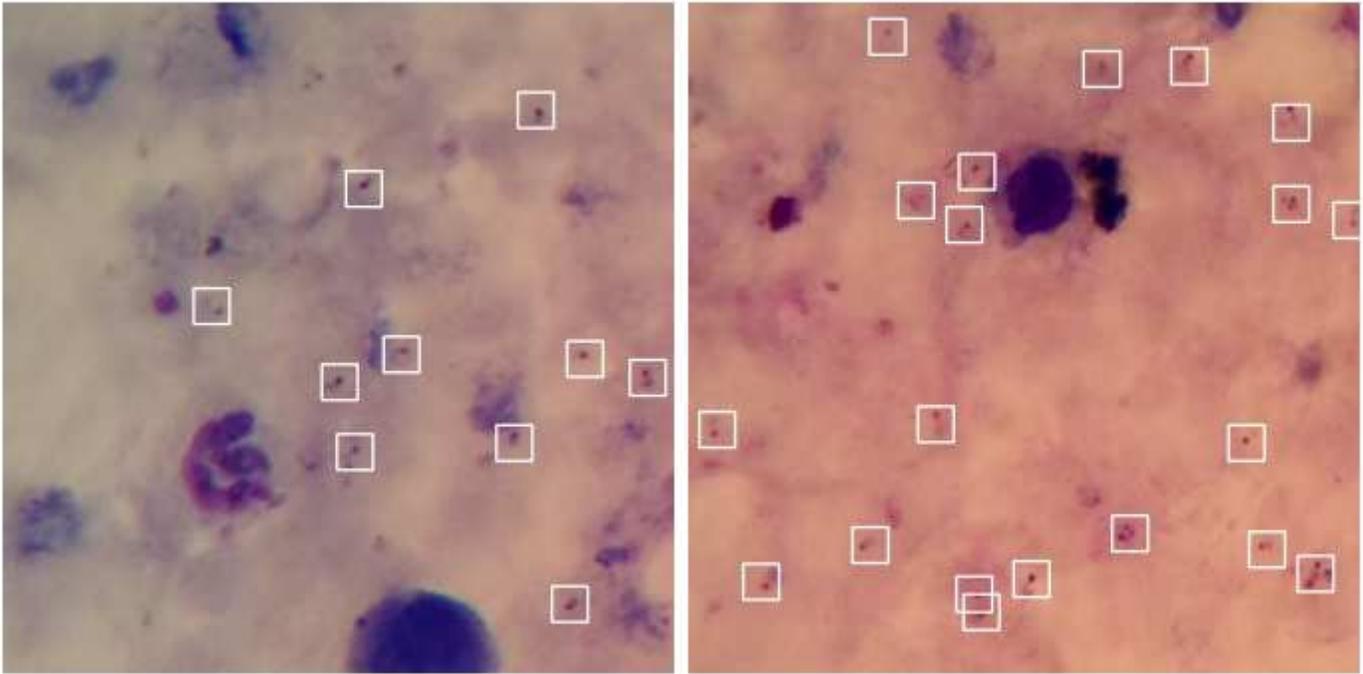


Figure 9

Sample Microscopic images containing ground truth bounding boxes for p.falciparum locations

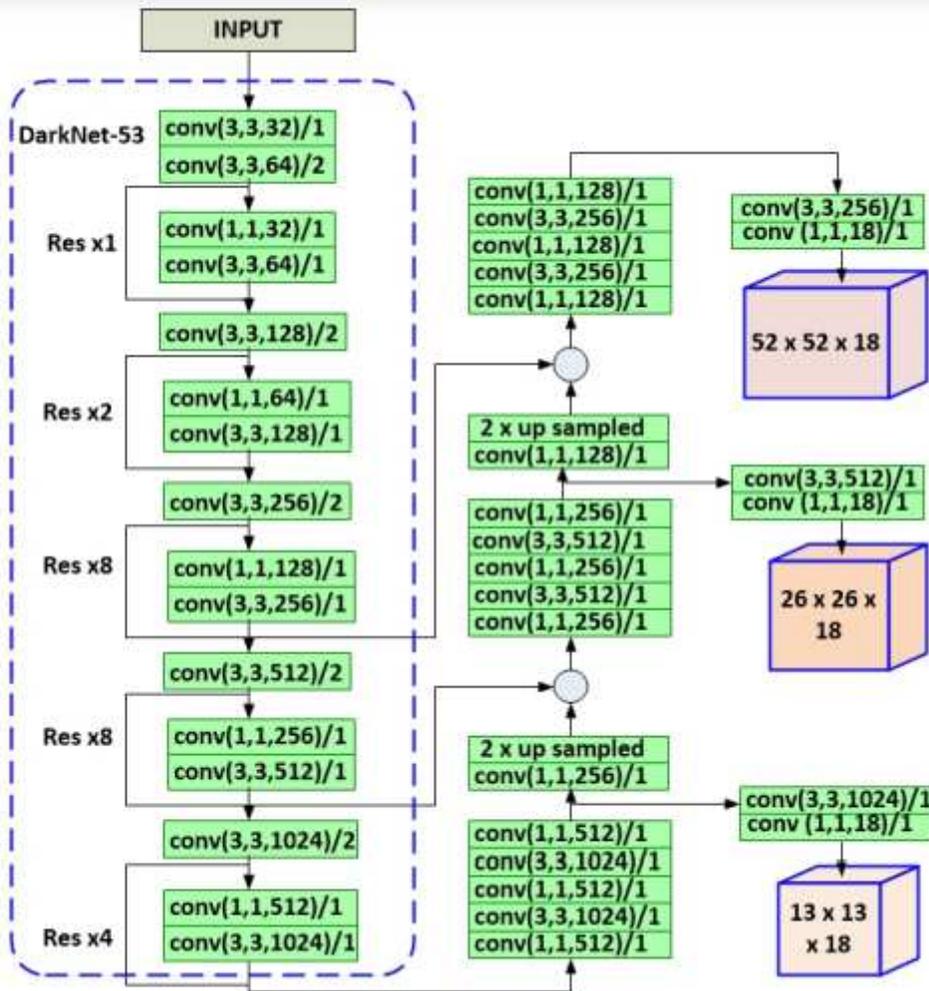


Figure 10

Architecture of YOLOV3 model

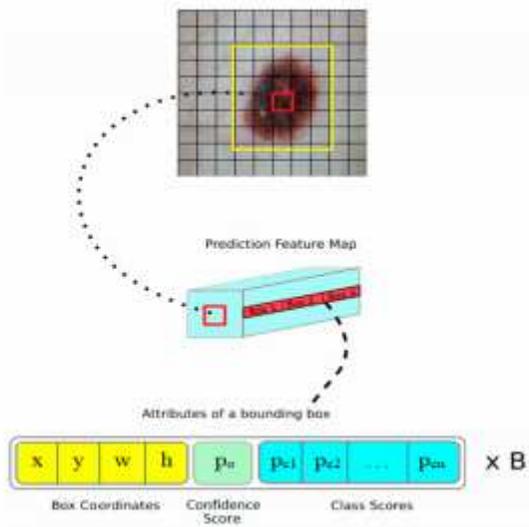


Figure 11

YOLOV3 prediction output tensor shape

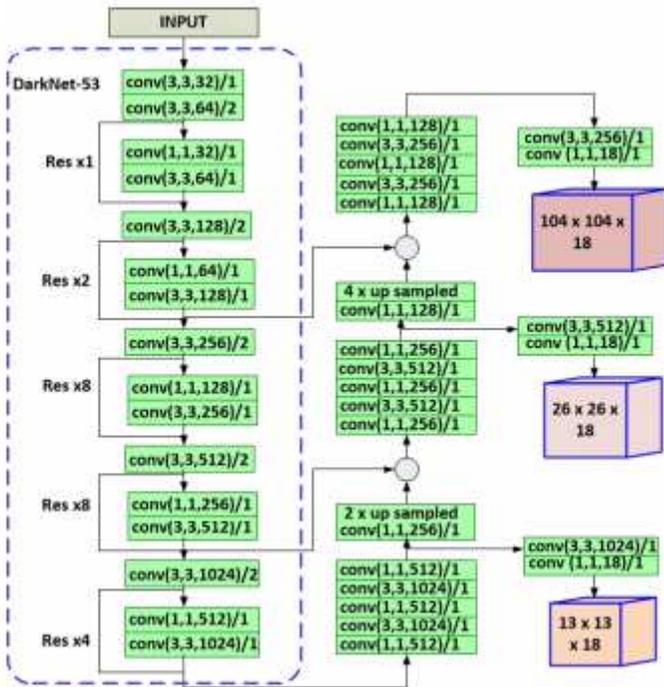


Figure 13

Architecture of YOLOV3-MOD1

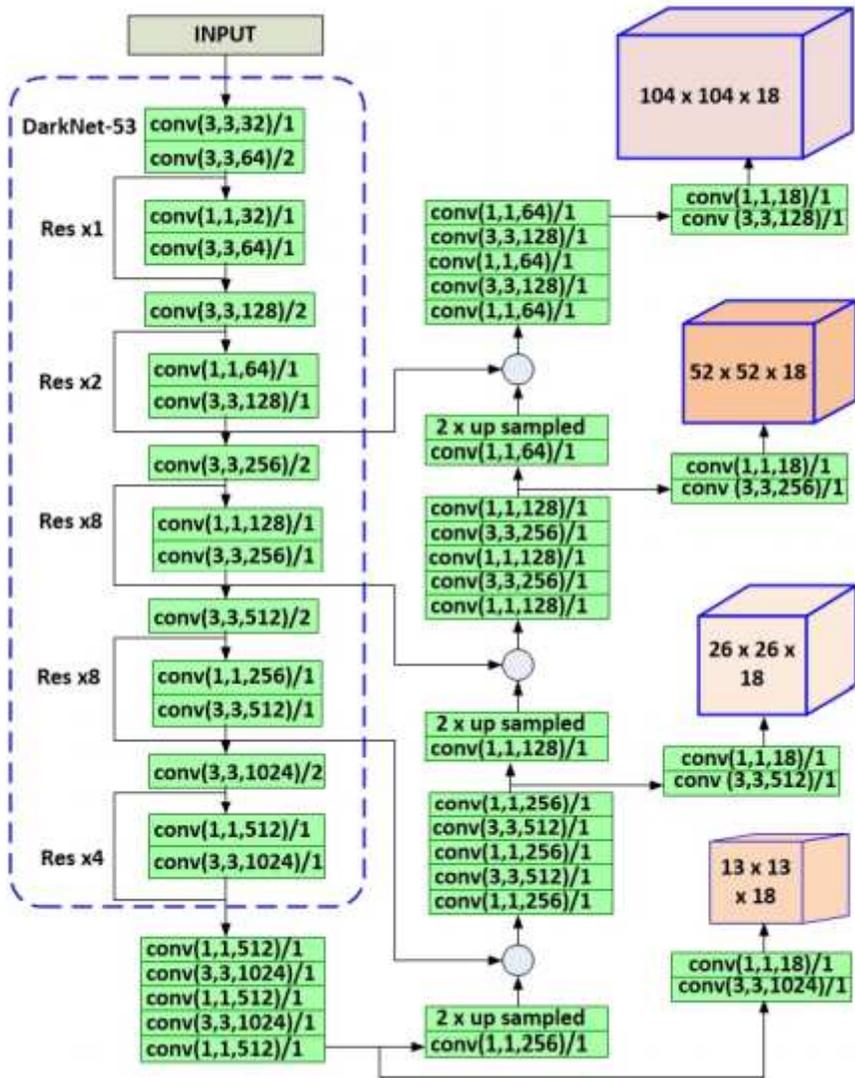


Figure 14

Architecture of YOLOV3-MOD2