CLASSIFICATION AND DETECTION OF *Plasmodium Vivax* INFECTED CELLS IN BLOOD SMEARS IMAGES

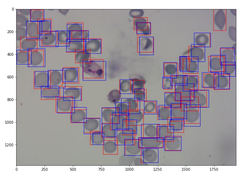
Erick W. Díaz & Jorge A. Balsells
Universidad San Carlos de Guatemala
ewdiazsa@gmail.com jorgebalsells@gmail.com

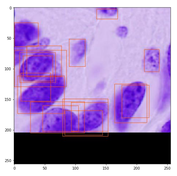
Allan Orozco
Industrial Engineering University of Costa Rica, School of Health Technologies,
allan.orozcosolano@ucr.ac.cr

INTRODUCTION

In Guatemala, this type of diagnosis (P Vivax malaria) has several dis-advantages, trained and experienced microscopist personnel are needed to identify and quantify the parasite in the thick drop (gold standard) and blood smears. There is also a high workload for the staff on the area in which the disease is concentrated; high estimated time in observation of blood films, on average a trained person observes a thick blood film for 10 to 20 minutes, in endemic areas a health center can receive between 200 and 300 weekly blood films, producing ocular fatigue, which is usually is the most influential factor in an incorrect diagnosis. Automation test through CNN networks it could accelerate diagnoses, increase the efficiency and performance of specialists.

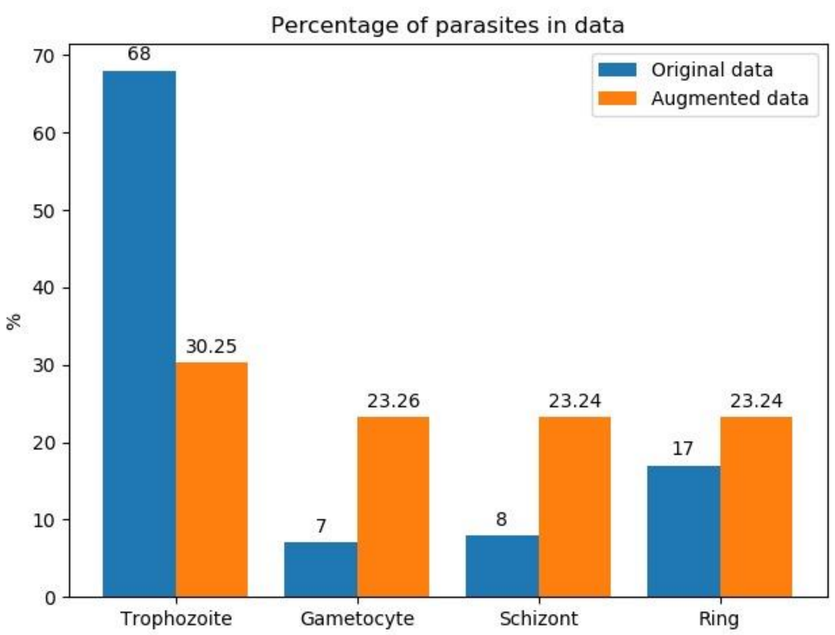
We used image set *P. Vivax(malaria)* infected human blood smears, accession number BBBC041version 1, available from the Broad Bioimage Benchmark Collection (Ljosa et al., 2012). The images in the dataset contains 1,364 images, with a total of 80,113 labeled cells, different researchers contributed labeling each cell in the dataset. These images were contributed by Jane Hung of MIT and the Broad Institute in Cambridge, MA. Each cell in the biological images has a class label an the bounding box coordinates, for the infected cell he have four classes gametocytes, rings, trophozoites and schizonts





METHODOLOGY

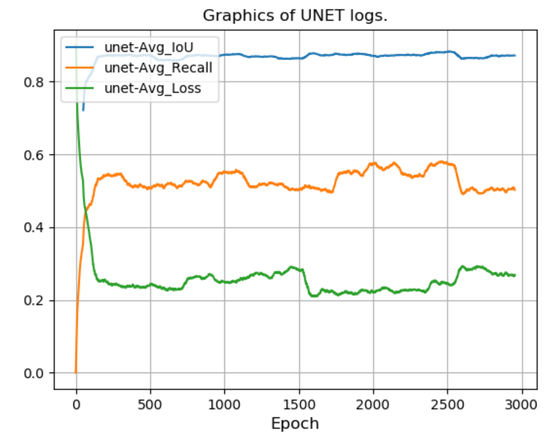
To increase the training data and remove the imbalance of the dataset for one of the classes and improve the generalization ability of the models, we fist cropped all parasites form images using the bounding box location and augmented these images with rotations, then we generated a method that randomly adds parasites into the training images that had less than four parasites, the method always add the parasite class that has fewer occurrences this runs until the different classes are close to be balanced.

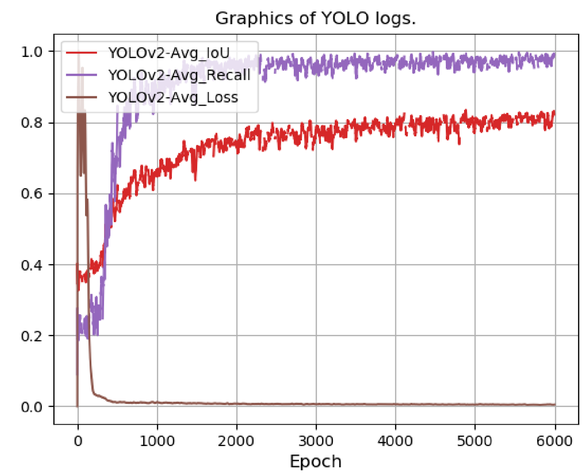


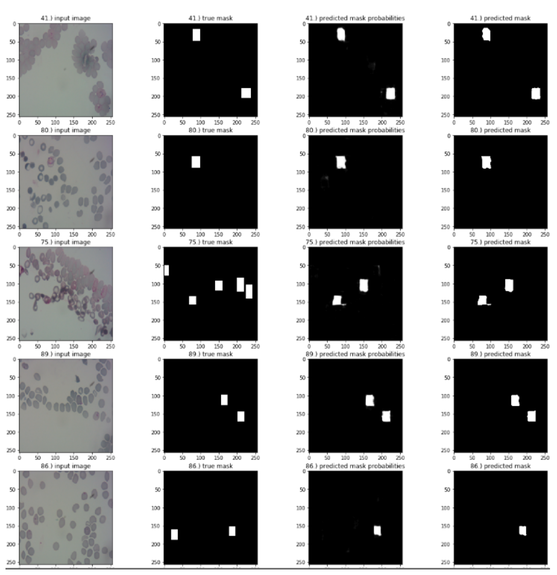
RESULTS

Add your information, graphs and images to this section.

Model	IoU	Precision	Recall	F1 Score
U-Net	71.1%	27.81%	44.29%	34.16%
YOLO	77.17%	45.41%	3.77%	6.96%
YOLO 1st augmented dataset	77.19%	41.56%	4.08%	7.44%
YOLO 2nd augmented dataset	84.85%	57.26%	5.13%	9.42%
YOLO 3rd augmented dataset	83.66%	46.71%	7.16%	12.42%

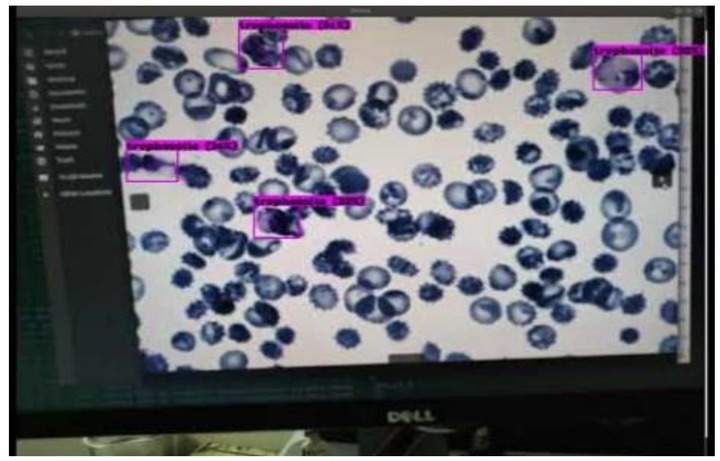


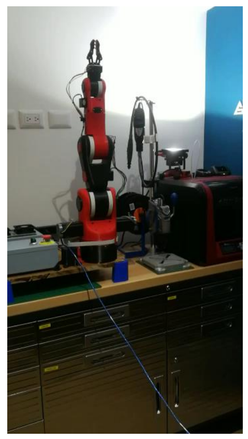




CONCLUSION

We were able to detect malaria parasites in the images, training the models with augmented datasets helped to improve the performance. In the future, we plan to extend this work, gathering our dataset, will be an important step, also creating an ensemble prediction using both models; we will also experiment with different novel models. Once we reach better performance measures we plan to automate the whole process using a robotic arm for manipulating the blood samples, the goal is to automate the complete process.



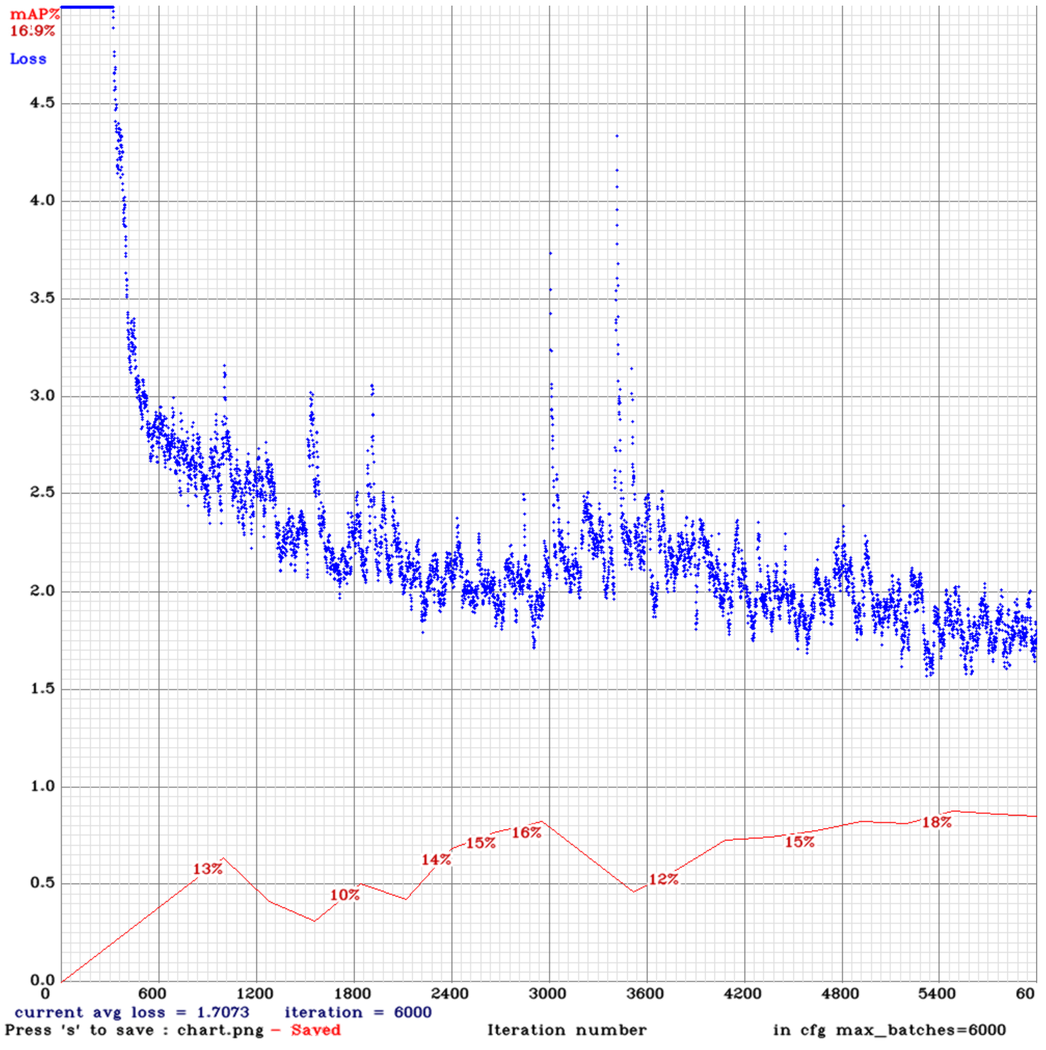


ACKNOWLEDGEMENTS

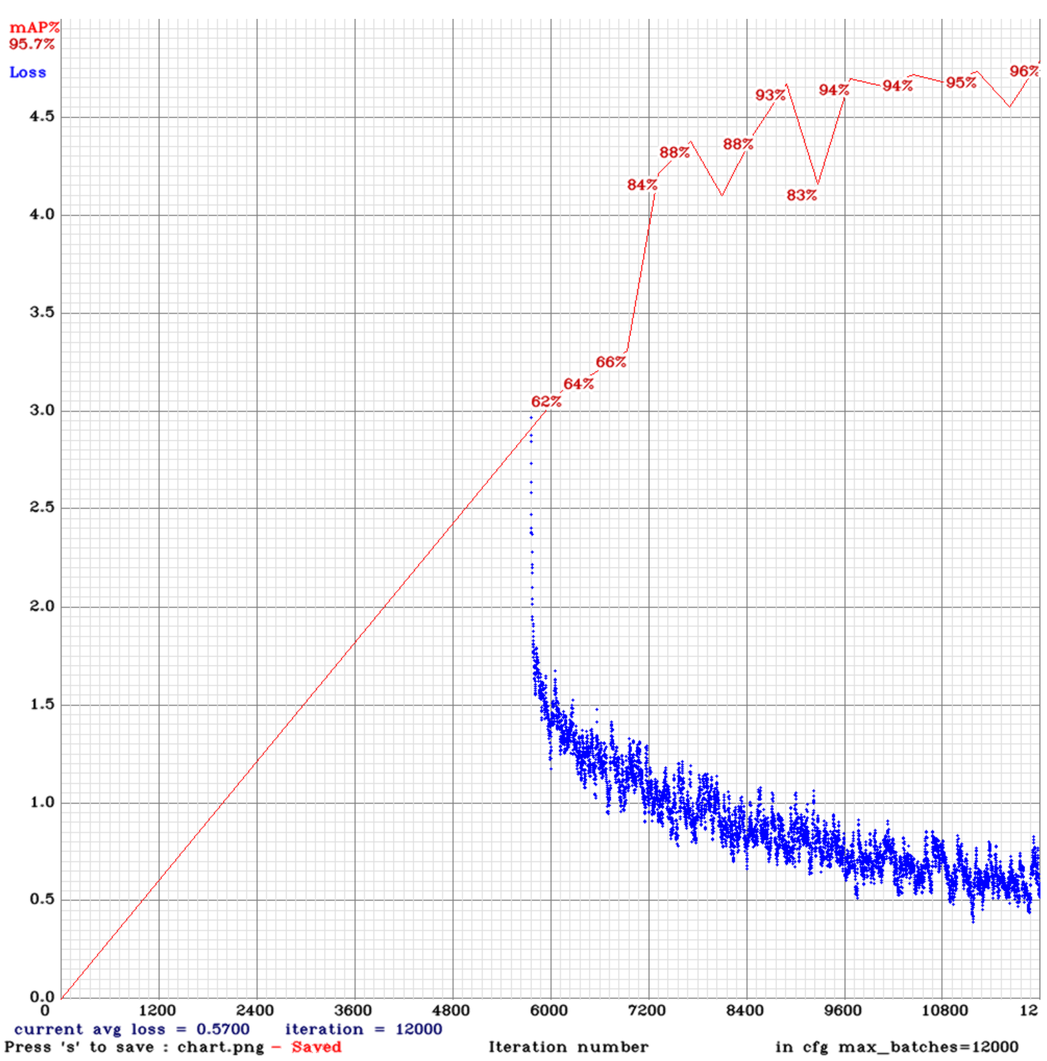
Original data available from the Broad Institute Repository at https://data.broadinstitute.org/bbbc/BBBC041/

These images were contributed by Jane Hung of MIT and the Broad Institute in Cambridge, MA.
We thank Jane Hung of MIT and the Broad Institute in Cambridge, for publishing and making available this dataset, and all the researchers that prepared the images.

Original Data Set



Augmented Data Set 1



Augmented Data Set 2