The threshold of resolution for the Classification of microscopic blood images with or without malaria

Zhouyang Min
Department of Biomedical Engineering
Duke University
Durham, NC 27708
zhouyang.min@duke.edu

Abstract

Medical image is a crucial portion in today's disease diagnosis such like CT, MRI and PET. Even though we have many newly developed imaging diagnosis methods today, some conventional microscopy diagnosis methods still cannot be substituted. The diagnosis of malaria through conventional microscopy is still necessary. Machine learning has been involved in biomedical filed for many years and developed rapidly, lots of research use this method to obtain a more efficiency way to detect the plasmodium parasites in the thick blood smear. The resolutions of images which get from the conventional microscopy are usually not super high. This project will modify the resolution of image dataset obtained through conventional microscopy, then interpret and summary the result of classification of images with or without plasmodium parasites. Finally made a conclusion on the effect of image resolutions will bring to the convolutional neural network image classification.

Introduction

Malaria is a serious and sometimes life-threatening disease that causes more than 400000 deaths worldwide each year [1]. Malaria is caused by Plasmodium parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. Early parasite-based diagnostic testing of malaria using microscopy helps to reduce disease spread and prevent deaths especially in developing countries [2,3]. Machine learning approaches are got succeed in many image-based diagnosis, disease prognosis and risk assessment [4]. We propose to construct a convolutional neural network (CNN) that can efficiently classify microscopic images of blood samples and tell if a sample has or does not have any parasites. We want to optimize some physical parameters to evaluate the factors that can affect our CNN classification result.

Related work

Recently, Kannojia et al points out that variation in images resolution will change the visual information of images which can influence the performance of CNN classification. In their study, they separate the resolution modified and not modified data and did classification under two conditions, one has trained on original resolution dataset and tested on caring resolution dataset (TOTV), another is trained and tested on each varying resolution dataset separately (TVTV) [4].

Their result shows that TVTV methods will return an overall higher classification accuracy but method will lead to the classification decrease as the dataset resolution decrease.

In our project, we want to do more analysis on how the resolution of images can affect the CNN imaging classification. We will choose the TVTV method only and keep decreasing the resolution of images to see if we will meet a threshold that the CNN classification gets failed. And we will add noise to the image to see if the more physical parameters added can affect the performance of CNN imaging model.

Methods

- **Dataset pre-processing:**

In our project, we used the annotated malaria image dataset released by artificial intelligence and data science group from Makerere University [6]. They captured those image data by using the smartphone adapter setup they designed. This dataset includes 1182 thick blood smear images, each image includes several plasmodium parasites in it. Each image in this dataset has a .xml file relate to it. The .xml file includes all the location information of the plasmodium parasites in that image, it saved the axis values of four corner of the 40X40 squares, and each square has a parasite in it. We first extracted all the axis information of the .xml file, grouped them and saved them to a .json file with correct image number label.

With the .json file, we cropped the square image with resolution 40X40 from each raw thick blood smear images, labeled them as plasmodium parasites images and then removed that square part from the raw image. Each image with parasites removed will be saved. As each raw image usually have an average of 10 parasites on it, then we used the saved raw images which has parasites removed to randomly crop 10 square images with 40X40 resolution and label as non-plasmodium parasites image. All the cropped 40X40 images form the datasets will be used to do the classification. We cropped in total 7090 images with parasites and 6387 images without parasites. This process was down with colab, we saved all the raw image data and .json file to a google drive folder and then mounted the drive in google colab.

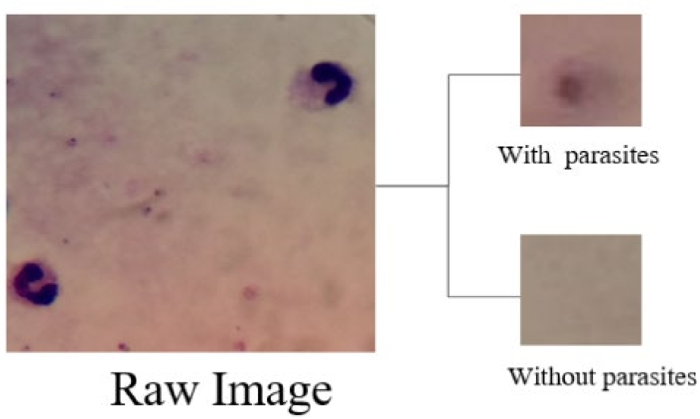

Figure 1: Image with and without parasites cropped from raw images.

- **Optimization of physical parameters**

With the classification data prepared, we read and splited our cropped data image into train and validation data. The data reading process took very long period, even though we set the running type to be GPU, it still needs over 20 minutes to finish. We defined an Gaussian noise by using tf.random.uniform function, and changed the standard deviation of it to increase the noise. Then we lowered the resolution of cropped image by using cv2.resize function.

- **CNN architecture and classification:**

As some popular CNN architecture like VGG usually requires images with higher resolution, in order to have a CNN model can train images with very small resolution, we built our own CNN architecture as Table 1 shows. After each modification of parameter, we ran our classification model (CNN architecture we created) and outputed our classification result.

CNN architecture		
Layer	Layer Parameter	Activatve function
Conv2D	filter = 8,size = (5,5),strides = 1, padding = 'same'	tanh
Conv2D	filter = 8,size = (5,5),strides = 1, padding = 'same'	relu
Conv2D	filter = 8,size = (5,5),strides = 1, padding = 'same'	tanh
Conv2D	filter = 8,size = (5,5),strides = 1, padding = 'same'	relu
Dense	units = 8	softmax
Flatten	default	default
Dense	Units = 8	softmax

Table 1: CNN architecture created

Results:

- **Classification under different noise**

The standard deviations of the gaussian noise we used include 0.0002, 0.002, 0.02, 0.2, 2, and it returned the corresponding train accuracy 0.9793, 0.9787,0.9645,0.9449,0.8987. It shows a small decrease in accuracy if we only compared the accuracy value. The accuracy can still reach 90% when the standard deviation increased to 2, at that time, we cannot identify what output image is as it has too much noises.

- **Training accuracy comparation**

When directly compare the training accuracy and validation accuracy value, we will find the accuracy will reach the highest value when the resolution of images decreased to 16X16. And when the resolution decreased to 3X3, the CNN image classification suddenly has a huge decrease. With different noise added, the changes of the classification performance are similar.

Training accuracy with different noise added	Resolution of Training and Validation Data Image					
	40X40	32X32	16X16	8X8	4X4	3X3
Gaussian noise with stddev = 0.0002	0.9618	0.9635	**0.9762**	0.9590	0.9471	**0.6087**
Gaussian noise with stddev = 0.002	0.9646	0.9679	**0.9771**	0.9585	0.9328	**0.6087**

Table2: The training accuracy of data images with different resolution

Validation accuracy with different noise added	Resolution of Training and Validation Data Image					
	40X40	32X32	16X16	8X8	4X4	3X3
Gaussian noise with stddev = 0.0002	0.9521	**0.9620**	**0.9606**	0.9465	0.9380	**0.6085**
Gaussian noise with stddev = 0.002	0.9676	0.9718	**0.9817**	0.9507	0.9183	**0.6085**

Table2: The validation accuracy of data images with different resolution

- **Training Accuracy curve comparison**

From the training and validation accuracy data, we already learned that before the resolution of image decreased to 3X3, it will always have the accuracy over 90%, The accuracy curve of training accuracy shows that when decrease the resolution, the epoch needed to finish the classification will increase.

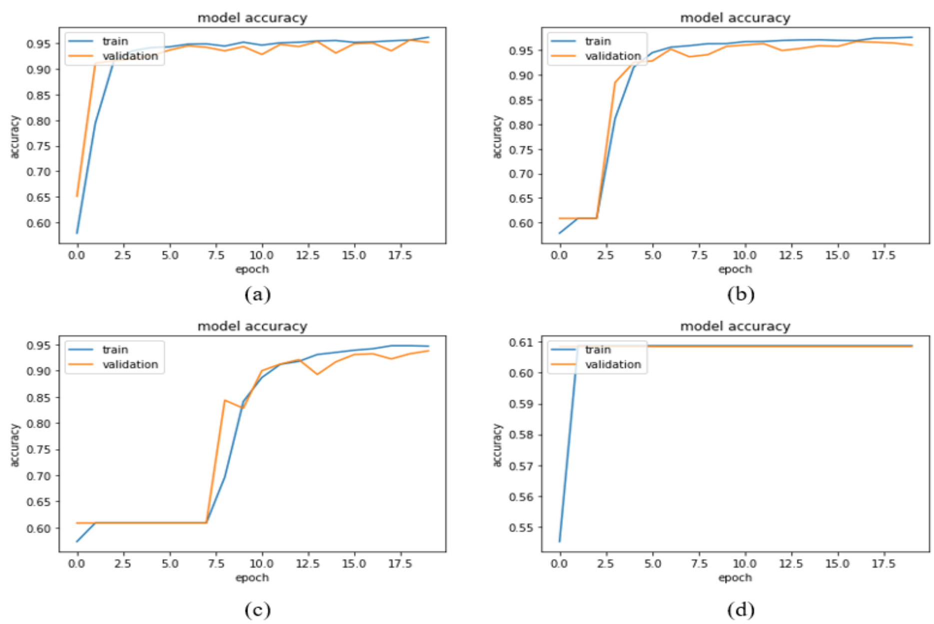

Figure 2: Training and validation accuracy curve with different data image resolution. (a)Data image with resolution 40X40, (b) Data image with resolution 16X16, (c) Data image with resolution 4X4, (d) Data image with resolution 3X3.

- **ROC curve comparation**

The ROC curve is relate to the sensitivity and specificity of the classification, the ROC curve of our result did not have a huge difference until the resolution of image changes from 4X4 to 3X3 when we have a large decrease in classification accuracy.

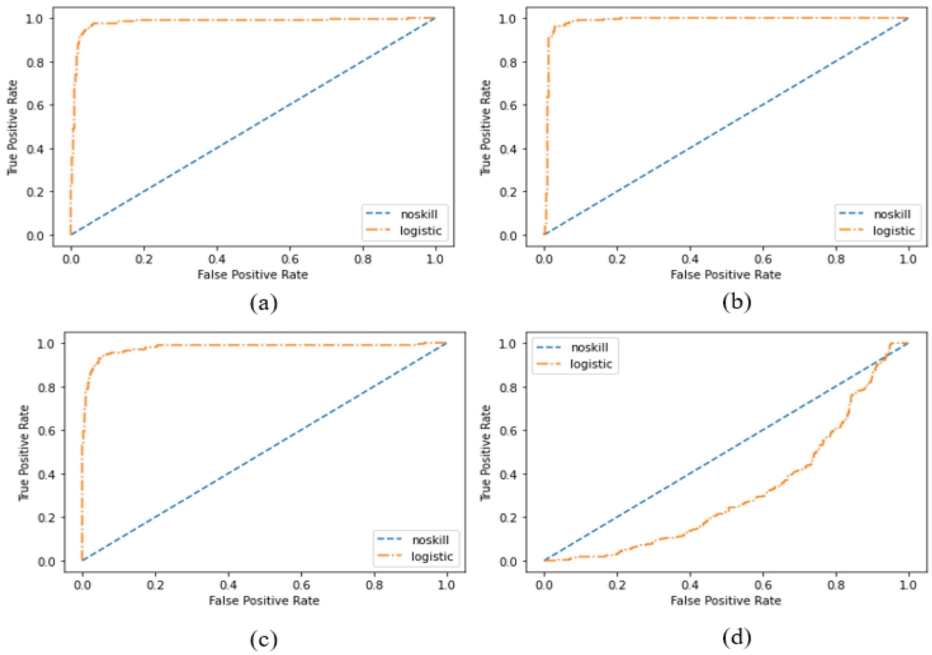

Figure 3: ROC curve with different data image resolution. (a)Data image with resolution 40X40, (b) Data image with resolution 16X16, (c) Data image with resolution 4X4, (d) Data image with resolution 3X3.

Discussion:

In this work, we use our own created CNN architecture to classify the cropped images. We optimize our classification by changing the images' resolution and compare the result in many ways. From our result, we found when the resolution of the image keep decreasing, the CNN imaging classification will finally failed, the threshold of our image dataset is 4X4 because when the resolution of images decrease from 4X4 to 3X3, the accuracy suddenly decreased from 90% to 60%. However, 4X4 is actually a very small resolution value. In actual life, it seems impossible that we need to do machine learning with such a small resolution image. Even though we found the threshold value, the resolution of images should not affect the result of classification too much as we will not have to use such low-resolution image dataset.

Besides from finding 4X4 is a threshold image resolution, we also found that the classification performance did not keep decreasing when we lowered the resolution of images. It reaches the highest value when the resolution changed from 40X40 to16X16, this result is different from the former related work, we think that may because we use different dataset and different CNN architecture. To verify our idea, some future works are needed.

- **Future work**

To be more confident about our result, we should analyze more datasets. We need to include images with relatively high original resolution and complicated pattens, to see if we will still be able to get the same result. To make sure the changes of classification performance is totally due to the changes of resolution of image, we will have add more types of physical parameters and train dataset with different CNN architecture to see if each time it will return the same results trends.

Reference:

[1] Mtalimanja, Michael, et al. "Malaria Funding Profile Against Disease Burden: Regression Analysis Case of Zambia, 2009 to 2018." (2020).

[2] World Health Organization. "Malaria surveillance, monitoring and evaluation: a reference manual." (2018).

[3] Jan, Zahoor, et al. "A review on automated diagnosis of malaria parasite in microscopic blood smears images." *Multimedia Tools and Applications* 77.8 (2018): 9801-9826.

[4] Komura, Daisuke, and Shumpei Ishikawa. "Machine learning methods for histopathological image analysis." *Computational and structural biotechnology journal* 16 (2018): 34-42.

[5] Kannojia, Suresh Prasad, and Gaurav Jaiswal. "Effects of varying resolution on performance of CNN based image classification: An experimental study." *Int. J. Comput. Sci. Eng* 6 (2018): 451-456.

[6] Quinn, John A., et al. "Deep convolutional neural networks for microscopy-based point of care diagnostics." *Machine Learning for Healthcare Conference*. 2016.