Blood Cells Classification with CNN

Jiaqi Hu, Fan Liang, Tuo Wu

Georgetown University

Abstract. The convolutional neural network is commonly applied to analyzing visual imagery. In CNN, the input is a tensor with shape (number of images) × (image width) × (image height) × (image depth), and convolutional kernels whose width and height are hyper-parameters, and whose depth must be equal to that of the image. In this project, we use CNN to research blood cell image classification. We will compare the influences of different regularization methods on CNN models, and extract the specific color from original images as input and compare the result with original images.

Keywords: Convolutional neural network · Parameter tampering · Exploit discovery.

1 Introduction

The diagnosis of blood-based diseases often involves identifying and characterizing patient blood samples. Automated methods to detect and classify blood cell subtypes have important medical applications.

In this project, we use CNN to recognize the colored cell, which is the white cell in each image and classify it in 4 types of blood cells: Eosinophil, Lymphocyte, Monocyte, and Neutrophil. In each image, the White Blood Cell is in red color, and the surrounding Red Blood Cells and Platelets are in light pink. The goal of the research is to classify the colored White Blood Cell. In the computer version, the model should recognize the colored cell first and then classify it into 4 types.

In the result part, we will compare the losses and accuracy rates for different models and input. The first input set is the original image, and the second input set is the color extraction image. Also, we use different models with different dropout rates and regularizers to compare their influences for the accuracy rates and losses.

2 Related Work

2.1 Yale University

From Yales Peripheral Blood Lab, the identification of blood cells is based primarily on observations of the presence or absence of a nucleus and cytoplasmic granules (Systems Cell Biology).

2.2 Kaggle

And previous kaggle researchers also used CNN to classify the blood cells. The previous models have over 400 thousand parameters and they used original images as input. While original images contain red blood cells that are unrelated to the goal, and these blood cells may be some noise. Our project focuses more on the effects of different dropout rates and tries to simplify the CNN model to have similar accuracy rates.

3 Datasets

We use a dataset called Blood Cell Images from Kaggle. This dataset has 12,500 augmented images of blood cells, and these images are labeled by cell type in the CSV file (Mooney, 2018). There are four different types of cells: Eosinophil, Lymphocyte, Monocyte, and Neutrophil. There are approximately 3000 images for each of 4 cell types split into training and test datasets.

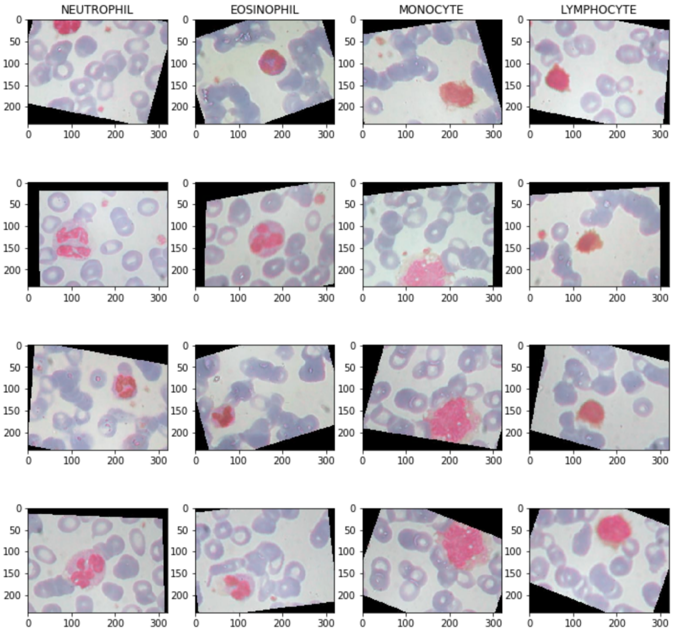

Fig. 1. Sample images of 4 types of blood cells

4 Method

In this project, we built several CNN models with different structures and techniques. First of all, we performed color segmentation to extract the useful cells which are red and purple in the image, because there are many unrelated cells in the image, we dont want these cells to affect our prediction. Also, we performed blur to our result, since we want to smooth our image noise and reducing details. In addition, we tried to build CNN models with different dropout rates and regularization to reduce overfitting. The basic CNN model contains 3 Conv2D layers, 3 MaxPooling2D layers, 1 ReLu Dense layer, and 1 softmax Dense layer. We also tried VGG16, but the results are not as good as other models.

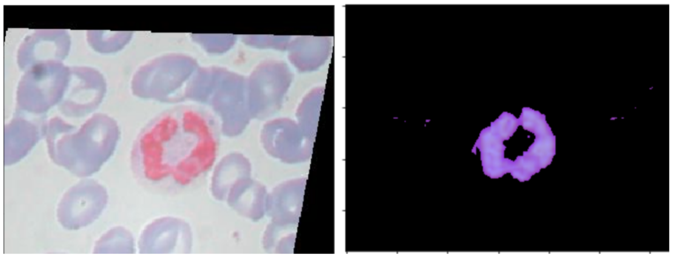

Fig. 2. Example of color extraction

5 Results

We run the experiments of 7 models with two input sets of images. The 7 models have different dropout rates or different regularizers. Below are the results.

	Original Image		Color Extraction Image	
Model	val_loss	val_acc	val_loss	val_acc
Basic	2.187	0.791	1.578	0.742
Dropout(0.1)	1.639	0.836	1.028	0.778
Dropout(0.3)	0.818	0.842	0.696	0.813
3 Dropout(0.1)	1.190	0.794	0.797	0.791
Lasso	0.978	0.776	0.806	0.786
Ridge	1.190	0.794	1.489	0.723
Ridge+Dropout(.1)	0.871	0.829	1.254	0.759
Vgg16	1.922	0.383	2.703	0.509

Table 1. Loss Rueslt

6 Discussion of Results

First of all, almost all other models that have dropout nodes or regularizers improve the result than the basic model.

Secondly, to compare models with different dropout rates and regularizers, the model with one dropout rate 0.3 has the lowest validation loss and highest validation accuracy for both original images and color extraction images.

The next part is we applied color extraction, which is the innovation of blood cell classifications. When the models only add dropout rate, the validation loss and validation accuracy both decrease. And in the lasso model, the color extraction has a better prediction. After the color extraction, the influence of surrounding non-targeting cells is reduced, while some detail information of target cells may also lose. Therefore, when we used color extraction images, the predictions do not have large improvements.

Further more, our models are much smaller than the previous kaggle works, while the validation accuracy rates are close. The previous works have more than 400 thousand parameters with a validation accuracy rate of 0.84 (Sharma, 2018). And our models only have 27 thousand parameters, while the highest accuracy rate is also 0.84.

7 Conclusions

In summary, to reduce the influence of non-target blood cells, we did color extractions on original images. Color extractions can recognize the target White Blood Cell with the color red and purple, while it also causes some details on the target cells vague. Therefore, the predictions do not have obvious improvement. Also, we can see that dropout nodes in layers or adding regularizers will improve the model performances.

For future works, we can use color extraction to make sure the position of target cells, and then do image segmentation to extract the target cells in the original images, and use the extraction images as input to train the CNN models. These processes will eliminate the influences of non-target cells and not lose any information of target cells.

8 Reference

Mooney, P. (2018, April 21). Blood Cell Images. Retrieved September 20, 2019, from https://www.kaggle.com/paultimothymooney/blood-cells.

Sharma, Kartik. (n.d.) Retrieved December 1, 2019
from https://www.kaggle.com/kartiksharma522/blood-cell-keras-inception.

Systems Cell Biology@Yale. (n.d.). Retrieved December 1, 2019,
from http://medcell.med.yale.edu/systems_cell_biology/blood_lab.php.