

Enhancing the Blood Cells Image Detection using YOLO, CNN, IOU and improving accuracy results using KNN

Melisa Krisnawati
Department of Computer Engineering
Dongseo University
Busan, South Korea
krisnawatimelisa@gmail.com

Lim HyoTaek
Department of Computer Engineering
Dongseo University
Busan, South Korea
htlim@dongseo.ac.kr

Lee Byung Gook
Department of Computer Engineering
Dongseo University
Busan, South Korea
lbg@dongseo.ac.kr

Kim JooHyung
Department of Computer Engineering
Dongseo University
Busan, South Korea
kjh3100210@gmail.com

Kim KyungHyun
Department of Computer Engineering
Dongseo University
Busan, South Korea
kkl5563@naver.com

Abstract—Blood cells are one of the most important parts of our body to transport the nutrients that have been processed to the whole organs for our health. It is very important to do a precise identification of blood cells to keep an eye on patient conditions. In this research, I will identify the types of blood cells accurately using the combination of machine learning CNN (Convolutional Neural Network) object detection YOLO (You Only Look Once) from a dataset of blood cell images that called as BCCD (Blood Cell Dataset Count) and doing the classification method. There will be 3 kinds of results: red blood, white blood, also thrombocyte. This technique can help the doctor faster and easily identify blood cells for giving the needs of data diagnosis from the results of the bloods cells classification to detect various diseases.

Keywords—blood cells, convolutional neural network, YOLO, k-nearest neighbor, machine learning, health

I. INTRODUCTION

Actually there are so many components that build our body, one of them is the blood cell. There are three types of blood cells which are with different looks, function and colours that will complete each other function and roles in the body. There are white blood cells to keep the body countering contagious disease, and red blood cells that are for delivering oxygen all over the body, also platelets to make sure the body will stop bleeding if it gets hurt and make the blood coagulate.

There is a conventional method that we usually know for testing the blood cells by observing using the microscope manually. But nowadays we can use machine learning to help identify the blood cells. To determine an object the methods will grabbing a classifier for objects and calculate it at many locations and scales in a test image. [3] YOLO is one of the object detection algorithms that can detecting objects model for images precisely by one process which is efficient for long term applications. YOLO proposes a neural network to make predictions of bounding boxes and class probabilities at once directly from the full image into one evaluation. If RCNN (Regional Convolution Neural Network) using region proposal methods (RPM) for generating potential bounding boxes on image then operating a classifier for the classification of the objects, then refine it on the post processing step removing the identical detections and rescore the boxes based on other objects, these are too slow for

optimisation because each part must be trained individually, YOLO sees entire images during training and testing so encoding contextual part about classes of image. [4]

II. METHODS

In this section, we will present a concise introduction to the utilized dataset, the techniques method for data pre processing, the machine learning algorithms utilized for detecting the blood cells, and a summary of the proposed methods for achieving the data with more accuracy. Figure 1 will show the workflow of this research.

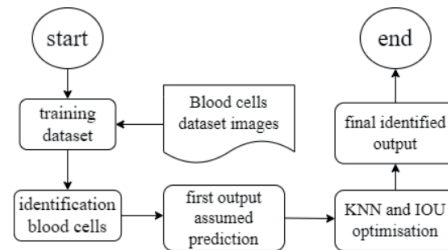


Fig. 1. Diagram workflow for the research methodology

First the dataset will be train and doing the identification of the blood cells. Then the first detection output will be assumed for the predicted blood cells. Afterward, we divided the dataset into 80% for training and 20% for testing. Then many machine learning techniques are used on the training data to train the model. Following that, running the machine learning models through the test dataset. KNN and IOU are also used to evaluate the effectiveness of the model detection from the image. Finally we compare the results to determine the best model as an evaluation. Once the model is validated, it can be used to maximize the blood cells detection results.

A. About dataset

As a researcher, this blood cells detection research is using the standard public dataset that called Blood Cell Count Dataset (BCCD) which can be access at mxnet website in .rec format (a text database file format that human editable) using bases mxnet.imageiter and class ImageDetIter that are image iterator with a large number of augmentation choices for detection. Then the dataset is reconstructed into VOC (Visual Object Classes) format

which can be found in the github link from cosmicad/dataset. This dataset can be directly used because it is already in jpg and xml format. These dataset include 364 total images of the human blood cells containing white blood cells, red blood cells, and platelets. And there are 4 images shown that have a very low quantity of red blood cells then I remove these 4 files. After that I split into 80 percent of training sets, that is 291.2 and I decide to round up into 300 training data and the rest is 20 percent validation set which is 60 for the testing data. The resolution of each image is 448x448. Processor that I used is Intel® Core™ i7-8700 CPU @3.20Ghz, with memory of 16GB, Nvidia Geforce GTX 1080 for graphic card, using windows 10 as operating system, 64 bit. The process is using the pytorch 1.12 framework and python 3.7 as deep learning framework. GPU that runs for the acceleration software is using CUDA version 11.3, tensorflow version 2.16.2 The example of the dataset can be seen in figure 2.

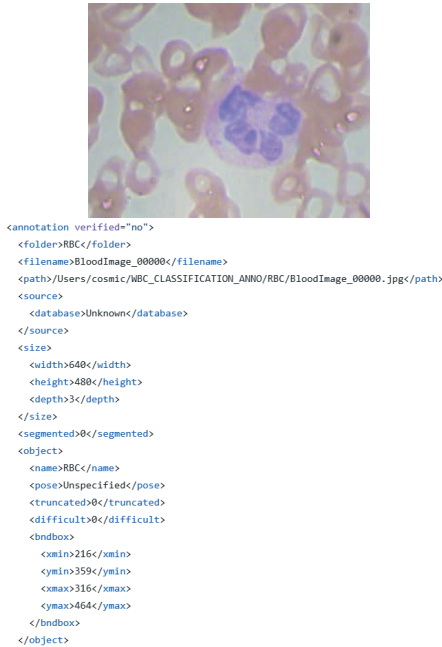


Fig. 2. Dataset examples in jpg and xml

B. Dataset Training

To make prediction faster for the image class and location, I split the image for a 7x7 grid square and each of it concludes into bounding boxes and every cell will forecast about two bounding boxes and the confidence score for every boxes. First employment of the YOLO model as a CNN calculated on a PASCAL VOC dataset that contains a diverse set of images with various object categories and complex scenes. This network that was possessed by the GoogleNet model has 24 convolutional layers that come after 2 fully associated layers for the image classification. [2] Rather than the inception method from GoogleNet, we can alternate to decrease the feature space followed by 3x3 convolutional layers using the 1x1 convolutional layers. The architecture details is the input image has a size of 448x448 pixels with 3 colour channel (RGB). First convolutional layer has the number of filter 64, size 7x7 and the convolutional filters will move across the input image/feature map (stride) by 2, and the max pooling is 2x2 (to reduce the spatial

dimension of the feature maps for the computational load) then the output will be 112x112x192. Second convolutional layer has the filter size 3x3, number of filters 192 and the max pooling and stride will be the same 2x2 then the output will be 56x56x256 and so on like described on the figure 3.

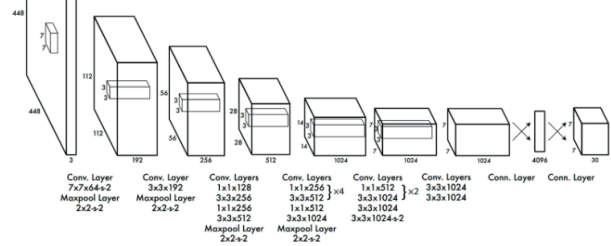


Fig. 3. The architecture of YOLO

For every anchor boxes there are 5 value predictions to implement for this blood cells identification using YOLO. The coordinates centre box of the objects to the grid cell X and Y, confidence level the intersection of union (IOU) between predicted box and the ground truth box, height and weight of the entire image. This VOC dataset was trained by 20 classes but for this case only need 3 white blood, red blood, and thrombocytes, so we also change the final convolutional layer at the architecture for the number of filters. In this research the anchor boxes we used are 5 because they provide good balance among the computational efficiency, detection accuracy, simplicity of the applied method based on models like YOLO. [4] More anchor boxes might improve accuracy slightly but for the cost the computational complexity will increase and become slower.

To predict bounding boxes, confidence scores and class probability for each anchor box, the final convolutional layer's number of filters is calculated based on the number of anchors and number of classes plus additional information needed for each detection. So the formula to calculate the final convolutional layer's number of filter as in:

$$F = A \times (C + 5) \quad (1)$$

Where F is number of filters, A is number of anchor boxes, C for number of classes and 5 means the bounding boxes coordinates, there are 4 items (x, y, w, h) and the confidence level IOU (intersection of union) between predicted box and the ground truth box.

For the image processing the number of anchor box (A) to predict the final bounding boxes around the objects are 5. The number of class (C) are 3, because there are red blood, white blood and thrombocytes. And the result, we have is 40 for the number of filters (F) on the final convolutional layer. Each grid cell in the output feature map will have 40 values, that having 5 bounding boxes each with 3 class probabilities, 1 confidence level and 4 coordinates of the bounding boxes.

Key metrics to compare and evaluate the performance of object detection for making a better detector are using a precision rate (P), recall (R), average precision (AP) and mean average precision (MAP). For each class the calculation of precision and recall can be seen below:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

So precision is the proportion of detected bounding boxes that are correct. Recall is the actual objects that are correctly detected (true positives out of all actual ground truth boxes). TP represents True Positive, the predicted bounding boxes that match ground truth boxes (using IOU), which is FP (False Positives) are the opposite, that does not match any ground truth box. And False Negatives (FN) are the ground truth boxes that are not detected by the model. High precision with low recall means the model is very accurate when it does predict, but misses many objects. At the other side if high recall with low precision means the model detects most objects, but with many false positives

Key metric for evaluating the object detection models are used in this research. It summarizes the relationship between precision and recall for a single object class at a particular intersection over union (IoU) threshold using Average Precision and Mean Average Precision.

$$AP = \int_0^1 P(R) dR \quad (4)$$

Average Precision (AP) is the area under the precision recall and used to assess the overall performance of each category on the test model. This curve plots precision against recall at various threshold levels for detection confidence scores. The resulting average precision provides a single score that captures how well the detector balances finding all objects (recall) with minimizing false positives (precision). On equation 4, where $P(r)$ is the precision measured as a function of recall r evaluated for recall values from 0 to 1. This $P(r)$ curve is obtained by varying the confidence threshold for your detector and plotting the resulting precision and recall at each point. [7] The AP formula with an integral measures the total area under the precision-recall curve for a class, representing the model's detection quality for all possible thresholds. A larger area (higher AP) means the model achieves both high precision and high recall at multiple thresholds, a lower area means poor trade off. In computation, this area (integral) is estimated using discrete thresholds as sum of rectangle or trapezoid areas. [5]

$$AP = \sum_n (R_n - R_{n-1}) P_n \quad (5)$$

Because precision and recall are computed at discrete thresholds, the integral is approximated as sum that we can see on equation 6. Where P_n is the precision at recall R_n and R_{n-1} is the previous recall value. This is directly computed by sorting detections by decreasing confidence, calculating precision and recall at each prediction, summing the products of the increase in recall and the corresponding precision. Numerically, the integral is typically approximated by summing regions between consecutive points on the PR curve (using, for example, the trapezoidal rule or step functions). It provides a comprehensive measure of model performance at every possible threshold, not just one chosen operating point. That is why AP (area under the PR curve or its integral) is more informative than using either precision or recall alone for evaluation in object detection.

Since the blood cells images data are discrete in the context of image analysis, the integral is typically approximated as a sum over all prediction thresholds. [6] On equation 6, where $P(k)$ is the precision at step k and the $\Delta r(k)$ is the increase in recall at that step. This approximation means the integral formula essentially "adds up" all the small rectangles (or trapezoids) under the PR curve between successive recall values, providing the average precision value. Very useful because summarize the overall detection quality as a single number by considering all operating points, rather than at a single arbitrary threshold and reflect the balance between missing true objects (recall) and avoiding false detections (precision), which is crucial in real-world detection scenarios.

$$AP = \sum_{k=1}^n P(k) \Delta r(k) \quad (6)$$

Mean Average Precision extends Average Precision to multi class object detection tasks, rather than at a single fixed threshold. MAP counting the average of all AP in every category to determine the overall detection capability of a detector. Compute the AP at a specified IoU threshold. It will providing a single, comprehensive metric reflecting the overall performance across all object categories. mAP is evaluating the model on both detection accuracy (correctness of identification) and localization (accuracy of bounding box placement) across all classes and thresholds and considering both false positives and false negatives, which giving a more robust measure of real world detection performance that makes it as the standard benchmark for evaluating and comparing object detection models. It reflects the model's consistency and effectiveness across different conditions and object types. Equation 7 is the calculation formula for Mean Average Precision. N is the validation dataset size in this research I use 60 because it is being performed on custom dataset (blood cell image). And counting the Σ amount of the average precision which is the precision and recall (the ground truth and the estimated bounding box).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (7)$$

Adding KNN can be used within the detection as a component for feature based classification. So after using the CNN it can classify proposed object regions by comparing their features to stored examples. When there's limited training data per class, KNN can help generalize from just a few labeled examples, boosting performance especially in rare classes or new object categories. Because it relies on similarity in feature space, KNN is robust to changes in lighting, orientation, or object pose in this case the it can handle the blood cells appearance on the dataset for the variations of pictures. In ensemble systems, KNN can be combined with other classifiers to improve final predictions, especially when dataset size or balance is a concern, so it will complementing deep models. [8] KNN also requires no retraining when new samples are available. We can simply add new labeled data so it will very useful in dynamic environments or for incremental updates, in this term medical field that will always have new data and changes. [7]

III. RESULTS

For each class it will compute the precision recall curve based on the model's prediction and calculating the AP as the area under curve. mAP will sum all the AP then divided

by N. IoU threshold defines whether predicted bounding boxes matches a ground truth box, if it is more than 0.5 then usually counts as the correct prediction. AP provides a reliable score for a model's detection quality on a per-class basis, balancing precision and recall. mAP gives an overall evaluation of a detector's performance across all object types, commonly used as the main metric for comparing models in research and benchmarking. A high mAP score indicates that the object detector can both correctly identify objects and avoid false positives. A crucial requirement for application especially for medical imaging.

mAP is less sensitive to the number of objects in an image, making it suitable for comparing models across datasets. Mean Average Precision threshold of each blood cell is also different for each variant of blood type. The less error rate we found for red blood cell is 55 percent, white blood is 35 percent, thrombocytes (platelets) is 25 percent. The result can be seen on table 1.

TABLE I. MAP THRESHOLD OF EACH BLOOD CELLS

MAP Threshold (in percent %)	Red Blood	White Blood	Platelets
20	5.450	0.086	0.238
25	4.438	0.048	0.082
30	3.450	0.068	0.077
35	2.867	0.016	0.081
40	2.652	0.053	0.079
45	2.284	0.096	0.084
50	2.367	0.124	0.105
55	2.086	0.217	0.118
60	2.143	0.365	0.125

And then we counted the accuracy with their confidence level from different variations of cells. The accuracy with confidence level of 55 percent for red blood cell is 96.23, white blood cell is 87.83 with confidence level 35 percent, platelets is 96.53 with confidence level 30 percent. After that KNN is used to refine the candidate detections from standard object detectors, check for consistency, or improve detection confidence. Implementing KNN algorithm in each cell and using the IOU among two detected closest cells and count the extent of overlap by 10 percent to evade the double detection on the blood cells. We ignore the cell that detected as double if the intersection is bigger than 10 percent. This figure 4 can explain the example of YOLO algorithm detected the cells double, and by using the KNN and IOU methods this detection can be deleted.

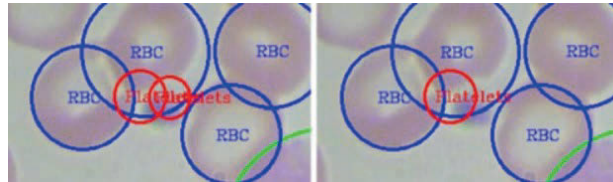


Fig. 4. Differences of blood cells detection, before detection (left), after detection (right)

IV. CONCLUSION

In this process we can directly identify 3 types of blood cells, red blood, white blood and thrombocyte according with YOLO methodology. As a core part of YOLO to compare and evaluate the performance of object detection using the Average Precision (AP) and Mean Average Precision (mAP). We also using a circular bounding boxes because it better fits for objects that are naturally circular, which is in this research the objects are blood cells. Also it can be more efficient since it can reduce the amount of the backgrounds for the objects. For making better precision at the prediction we are using the KNN (k-nearest neighbour) and IOU (intersection over union) for eliminating multiple detectors of the identical entity. KNN is a complementary component in the detection process that will helping the situation where standard detectors are limited in robust classification to improving/adjusting blood cells detection type and classification method. With the correctness and performance of this technique, this can be told that this system has the possibility to make it easier and faster for the traditional style blood cell recognition process.

ACKNOWLEDGMENT

Following are results of a study on the "Busan Regional Innovation System & Education (RISE)" Project, supported by the Ministry of Education and Busan Metropolitan City.

REFERENCES

- [1] Ren, S., He, K., Girshick, R., et al.: 'Faster R-CNN: towards real-time object detection with region proposal networks', IEEE Trans. Pattern Anal. Mach. Intell., 2017
- [2] Redmon, J., Divvala, S., Girshick, R., et al.: 'You only look once: unified, real-time object detection'. IEEE Conf. on Computer Vision and Pattern Recognition, December 2016
- [3] Acharjee, S., Chakrabarty, S., Alam, M.I., et al.: 'A semiautomated approach using GUI for detection of red blood cells'. Proc. Int. Conf. on Electrical, Electronics and Optimization Techniques, 2016, pp. 525-529
- [4] Redmon, J., Ali, F.: 'YOLO9000: better, faster, stronger', arXiv preprint arXiv:1612.08242, 2016
- [5] Bodzas, A., Kodytek, P. & Zidek, J. A high-resolution large-scale dataset of pathological and normal white blood cells. *Sci Data* 10, 466 (2023). <https://doi.org/10.1038/s41597-023-02378-7>
- [6] Shahzad M, Ali F, Shirazi SH, Rasheed A, Ahmad A, Shah B, Kwak D. Blood cell image segmentation and classification: a systematic review. *PeerJ Comput Sci.* 2024 Feb 2;10:e1813. doi: 10.7717/peerj-cs.1813. PMID: 38435563; PMCID: PMC10909159.M.
- [7] Boyd, K., Eng, K.H., Page, C.D. (2013). Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals. In: Blockeel, H., Kersting, K., Nijssen, S., Železný, F. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2013. Lecture Notes in Computer Science(), vol 8190. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40994-3_29D.
- [8] Kurniati, F. T., Manongga, D. H., Sembiring, I., Wijono, S., & Huizen, R. R. (2024). The object detection model uses combined extraction with KNN and RF classification. *arXiv preprint arXiv:2405.05551*.