Development of a service for the search and classification of leukocytes on bone marrow blood smears

together with THIRD OPINION

Student(s)
Evgeniy Garsiya, Evgeniy.Garsiya@skoltech.ru, tel: +7 905 736 11 08

Program
Data Science

Skoltech supervisor
Ivan Oseledets

Company supervisor
Evgeniy Sidorov

Context

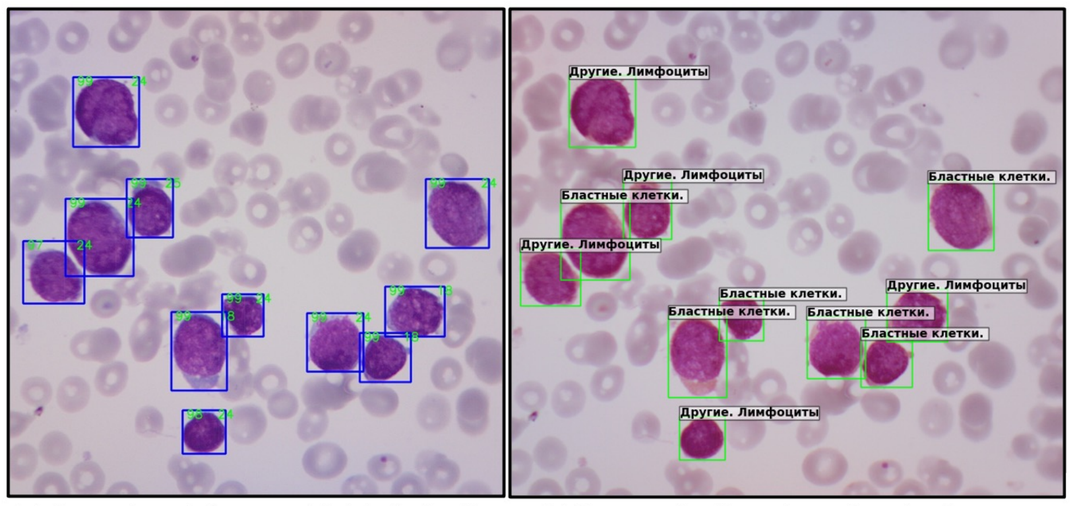
- One of the most terrible blood disease which needs to be detected as fast as possible is leukemia. Presence of leukemia can be defined based on the amount of blast cells (which is a type of leukocyte) comparing to the amount of all other cells.
- To automize, speedup and reduce the cost of blood analysis a computer vision techniques can be used. In particular, it is possible to develop a detection and classification networks which can process the patient images and predict the presence of leukemia.

Objectives

- The first aim was to develop an accurate detector which can catch a blood cell on the image.
- Secondly, the classifier network must be trained to classify every blood cell, caught by detector. The main aim here is to avoid false positive predictions when the model fails to correctly classify the blast cell.
- Thirdly, it was required to create a full pipeline as a combination of both networks with patient images as an input and blast cells percentage as an output. After that it was covered with Docker.

Process

- The first step was to clear and prepare the provided data and make a correct data splitting, which is quite important in our case.
- Then, a YOLOX Nano network was used for detection with 0.98 mAP-50 metric on validation set. It is a SOTA approach meaning “You Only Look Once” which highlights that the network observes the whole image info at once (no sliding windows comparing to other approaches).
- Secondly, I used EfficientNet B3 as a classifier and added a hierarchy to provide a two step classification (blast cell or not - on the 1st step and multilabel classification for the case if the cell is not blast). This improved results comparing to classical multilabel classification. The resulting validation binary and multilabel accuracy was 96% and 83%.
- After networks' testing and tuning a research pipeline was developed and possibility to run through a Docker container was added.



(a) Ground truth image with labeled cells (b) Detected cells and predicted cell types

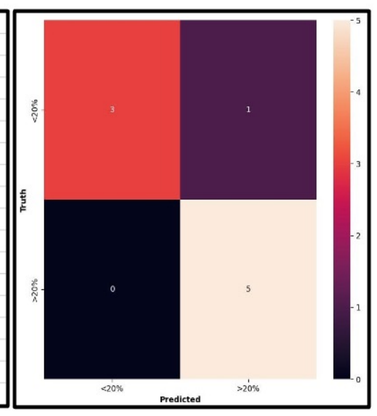
Fig. 1

Results

- Detector and classifier models were developed. On Fig. 1 you can take a look at output visualisation.
- The patients’ decision making functionality was added to define the leukemia presence based on the models output (Fig. 1). The resulting file with patients statistics is shown on Fig. 2 (leukemia presence defined as >20% of blast cells).
- The outcome of the project is a ready system which can be launched on server through Docker with build-in networks, which can be used to analyse a patient blood images and make a decision on patients' leukemia status.

Общее количество обнаруженных клеток	649	717	686	860	683
Бластные клетки	258 (39.8%)	375 (52.3%)	119 (17.3%)	355 (41.3%)	44 (6.4%)
Палочкоядерные нейтрофилы	10 (1.5%)	11 (1.5%)	43 (6.3%)	14 (1.6%)	77 (11.3%)
Сегментоядерные нейтрофилы	15 (2.3%)	1 (0.1%)	52 (7.6%)	4 (0.5%)	77 (11.3%)
Лимфоциты	237 (36.5%)	254 (35.4%)	162 (23.6%)	386 (44.9%)	146 (21.4%)
Моноциты	36 (5.5%)	6 (0.8%)	11 (1.6%)	18 (2.1%)	15 (2.2%)
Не учитываются в подсчете	42	38	49	100	49
Полихроматофильные нормобласты	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.1%)	1 (0.1%)
Все эритрокариоциты	54 (8.3%)	24 (3.3%)	148 (21.6%)	36 (4.2%)	114 (16.7%)
Метамиелоциты нейтрофильные	4 (0.6%)	1 (0.1%)	20 (2.9%)	4 (0.5%)	35 (5.1%)
Базофильные нормобласты	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Митозы. Тоже лучше не учитывать в подсчете	13	17	68	29	48
Миелоциты нейтрофильные	14 (2.2%)	5 (0.7%)	21 (3.1%)	4 (0.5%)	84 (12.3%)
Плазматические клетки	1 (0.2%)	0 (0.0%)	37 (5.4%)	0 (0.0%)	3 (0.4%)
Оксифильные нормобласты	4 (0.6%)	7 (1.0%)	16 (2.3%)	7 (0.8%)	35 (5.1%)
Промиелоциты	13 (2.0%)	31 (4.3%)	54 (7.9%)	30 (3.5%)	51 (7.5%)
Эозинофилы	2 (0.3%)	1 (0.1%)	2 (0.3%)	0 (0.0%)	0 (0.0%)
Базофилы	1 (0.2%)	1 (0.1%)	1 (0.1%)	1 (0.1%)	1 (0.1%)

(a)



(b)

Fig. 2 The models’ output file

(a) Patients’ blood cells statistics. (b) Confusion matrix for 9 real test patients

Conclusions

- During the project I faced a lot of challenges. First of all it was difficult to get know and work with such a complex repository as YOLOX.
- Secondly, a work with Linux and Docker was quite a new and unknown thing for me, but it is a useful experience.
- Thirdly, I implemented and tried a big amount of experiments during a classifier development. Some of them were easy, but others were quite challengeable (like hierarchy implementation). I studied a lot of useful tricks and approaches for network development.
- I would like to thank Anna Mesheryakova (CEO) and Ivan Oseledets for providing such a great opportunity for an internship. Great thank to my mentor - Evgeniy Sidorov for advising my work and helping with difficult tasks.