Multiclass Deep Active Learning for Detecting Red Blood Cell Subtypes in Brightfield Microscopy

Ario Sadafi[1,2,3], Niklas Koehler[1], Asya Makhro[4], Anna Bogdanova[4], Nassir Navab[2], Carsten Marr*[1], and Tingying Peng*[1,2]

[1] Institute of Computational Biology, Helmholtz Center Munich, Neuherberg 85764, Germany
[2] Department of Mathematics, Technical University of Munich, Munich 85748, Germany
[3] Arivis AG, Munich 80636, Germany
[4] Red Blood Cell Research Group, Institute of Veterinary Physiology, Vetsuisse Faculty and the Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland

Abstract. The recent success of deep learning approaches relies partly on large amounts of well annotated training data. For natural images object annotation is easy and cheap. For biomedical images however, annotation crucially depends on the availability of a trained expert whose time is typically expensive and scarce. To ensure efficient annotation, only the most relevant objects should be presented to the expert. Currently, no approach exists that allows to select those for a multiclass detection problem. Here, we present an active learning framework that identifies the most relevant samples from a large set of not annotated data for further expert annotation. Applied to brightfield images of red blood cells with seven subtypes, we train a faster R-CNN for single cell identification and classification, calculate a novel confidence score using dropout variational inference and select relevant images for annotation based on the confidence of the single cell detection and the rareness of the classes contained in the image. We show that our approach leads to a drastic increase of prediction accuracy with already few annotated images. Our original approach speeds up annotation and improves classification of red blood cell subtypes. This important step in diagnosing blood diseases will profit from our framework as well as many other clinical challenges that suffer from the lack of annotated training data.

Keywords: Active learning · Multiclass annotation · Single cell microscopy.

1 Introduction

A typical human red blood cell can be morphologically described as a biconcave discoid, called a discocyte [11]. Changes in the volume of the cell change

its appearance: as the volume decreases, it shrivels into a star-like shape called echinocyte with distinguishable convex rounded protrusions. As volume increases, the cell expands into a shape with single- or multi-concave invaginations, called a stomatocyte. In physiological conditions, seven different morphological sub-types can be distinguished by cytologists (see Fig. 1) and appear in a particular frequencies, which change upon environmental challenges or in a course of a number of diseases [8].

Detection and classification of the red blood cell subytpes is a crucial step for blood sample analysis and the diagnosis of blood diseases [11]. However, images are produced with different modalities, illumination conditions and zoom levels. Thus, classification of red bloods cells nowadays still relies on manual annotation by an expert.

Deep learning approaches are known to be versatile and adaptive to new environments and excel on a couple of recent biomedical challenges, like the classification of skin cancer (Esteva et al., 2017 [4]) or the prediction of mutations from histopathological slides (Coudray et al., 2018 [2]). A first approach to the classification of red bloods cells has been recently also proposed (Xu et al., 2017 [14]). However in general, the application of powerful deep learning algorithms in clinical applications is heavily limited by the need of large amounts of well annotated data, since expert time is typically scarce and expensive. We thus want to significantly reduce redundancy in manual annotation by developing uncertainty based scores that allow us to involve expensive expert knowledge only where necessary.

One promising approach to break the bottleneck of data annotation is active learning, which uses a learning algorithm that carefully selects unlabeled data points to interactively query experts for new annotations. This expert-in-the-loop process has been demonstrated to achieve similar or even greater performance as compared to a fully labelled data-set, with a fraction of the cost and time that it takes to label all the data [9]. Here, we combine active learning with object detection and develop a novel active learning annotation tool to guide expert annotation. Although different active learning methods have been proposed to accelerate the annotation process for classification problems, e.g. [5], few approach exist that allows to select those for object detection, and none for a multiclass detection problem with clinical relevance.

Our active learning annotation tool interactively select candidate annotation set by measuring the uncertainty of classification and detection of single cells, and by considering rare classes in our data set. Our approach is the first to calculate relevance for the goal of active learning in multiclass object detection and the first to come up with intelligent data selection for expert annotation for biomedical images.

2   Method

Our proposed active learning annotation tool starts from an Faster R-CNN model trained with an annotated training set (see Fig. 2). We apply the trained
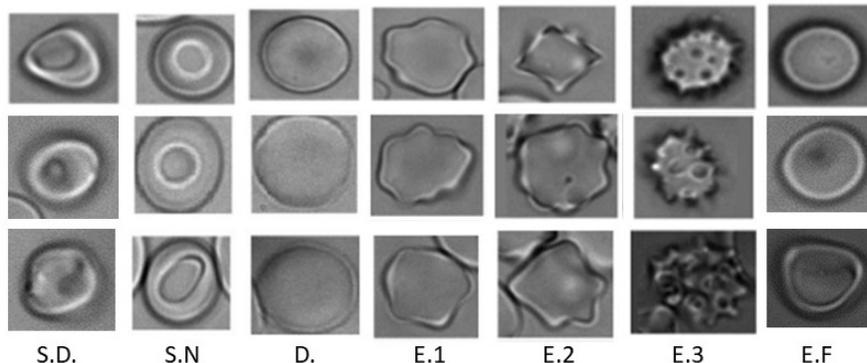
Fig. 1. Red bloods cells change their morphology due to environmental changes or in the course of a disease. They can be classified into seven subtypes, from left to right: dehydrated stomatocyte (S.D.), normal stomatocyte (S.N), discocyte (D), primary, secondary, tertiary and final echinocyte (E.1, E.2, E.3, E.F).

model on not annotated images and select the most relevant cells in these images based on a novel uncertainty analysis in order to ask for expert annotations. With these additional annotations, we update our model and select new cells for more annotations based on the updated algorithm. We keep iterating this annotation process until all cells above a particular uncertainty are annotated or a desired classification performance is achieved.

2.1 Object detection with Faster R-CNN

Faster R-CNN is an advanced version of Fast R-CNN [6] and R-CNN [7] and was first proposed in [12]. In this approach a Fast R-CNN is coupled with a Region Proposal Network (RPN) and both networks are trained together: convolutional layers extract features from the input image, the RPN generates object proposals based on the feature map, and each proposal is classified into one of the defined classes. We used a VGG-16 network [13] pretrained on ImageNet [3] as the backbone. More formally considering F_θ a Faster R-CNN model with weights θ and I an input image we have:

$$p, t^k = F_\theta(I) \tag{1}$$

where p is discrete probability distribution over all classes (as is normally computed with soft-max over the last fully connected layer) and t^k is the bounding box regression for every class. The multi-task loss of the Fast R-CNN can be defined as follows:

$$L = L_{cls}(p, u) + \lambda[u > 0]L_{loc}(t^u, v) \tag{2}$$
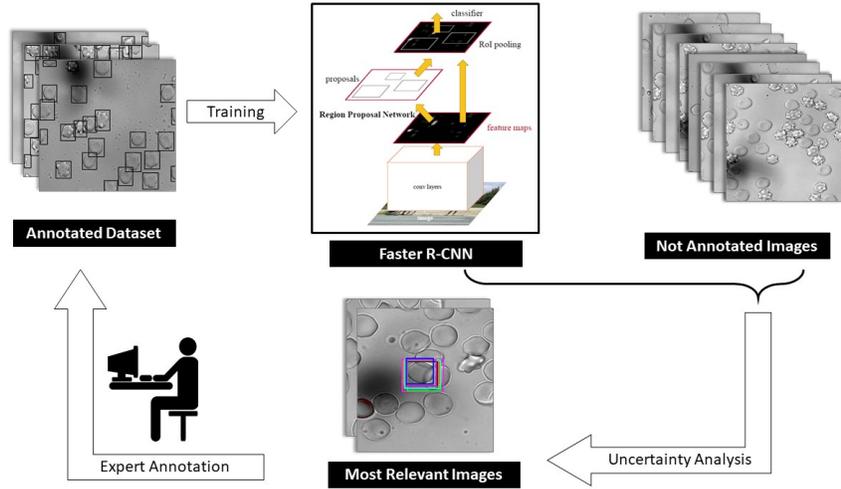
Fig. 2. Overview of the proposed active learning annotation tool. First, a Faster R-CNN model is trained on an annotated dataset. Not annotated images are then analyzed with the trained model, the uncertainty of detection and classification is determined, and the most relevant cells are passed to the expert for annotation. With the new annotations, a new cycle starts.

where u,v are annotations of bounding boxes and their classes from the dataset respectively and t^u is the bounding box regression corresponding to the ground truth u. Brackets are Iverson brackets which yield 0 for the background class ($u = 0$) and 1 for the rest. λ is a balancing parameter between classification loss L_{cls} and localization loss L_{loc} [6]. The localization loss is defined as

$$L_{loc}(t_i, v) = \sum_{m \in \{x,y,w,h\}} \text{smooth}_{L1}(t_i^m - v^m). \quad (3)$$

The classification loss L_{cls} is calculated with softmax cross entropy:

$$L_{cls} = -\log p_u. \quad (4)$$

For training we used an approximate joint training method [12].

2.2   Uncertainty score per cell

For each cell, we measure the uncertainty of our model prediction with three scores: (i) the detection uncertainty, (ii) the classification uncertainty and (iii) a binary score of the possibility of the object belonging to a rare class. We explain each score in the following subsections.

Detection uncertainty For each non-annotated image, we performs N inferences with our model and quantify the model uncertainty using dropout variational inference [10]. To evaluate the certainty of the model in detecting cells, we compare bounding boxes of each cell across N inferences. For every inference, we only keep the bounding box of the class that has the highest probability in p. We call this bounding box d. Our uncertainty score of detection U^d is thus defined as:

$$U^d = \frac{1}{N-1} \sum_{i=2}^{i=N} \text{IoU}(d_1, d_i) \tag{5}$$

where d_i is bounding box d in the ith inference and IoU measures intersection of union between two given bounding boxes. It is clear that $U^d \in [0, 1]$.

Classification uncertainty For each detected cell, the most probable class c from p is picked. Having N inferences we have the set $c = \{c_1, c_2, ..., c_N\}$. Hence ,we measure the uncertainty of classification using:

$$U^c = \frac{1}{N} \sum_{i=1}^{N} [c_i = c_m] \tag{6}$$

where c_m is the mode, i.e. the item with the most frequency, in set c and U^c is its frequency. Similar to U^d, U^c is also bound in $[0, 1]$.

Rare class prediction The red blood cell dataset has a strong class unbalance: cells belonging to the discocyte (D.) or primary echinocyte (E.1, see Fig. 1) class are much more frequent compared to dehydarted stomatocytes (S.D) or final echinocytes (E.F), which is an irreversible state. Blood cells of rare classes are clinically interesting yet detecting them is extremely challenging due to the small number of samples and large variations in appearance. In order to boost the precision of detection in rare classes, we introduce another metric to prioritize annotation of those cells that are likely to belong to a rare class.

$$\forall j \in R : U^r = \begin{cases} 0 & p^j \leq 0.2 \\ 1 & p^j < 0.2 \end{cases} \tag{7}$$

where R is the set of rare classes which is previously known to us and p^j is the probability of class j in the Faster R-CNN discrete class probability.

2.3 Relevance score per image

To save expert's time and annotate images more efficiently, we rank every non-annotated image with a relevance score defined as:

$$R_{\text{img}} = \sum_{j=1}^{M}(U_j^c <= \alpha) + \sum_{j=1}^{M}(U_j^d <= \beta) + \gamma \times \sum_{j=1}^{M} U_j^r \tag{8}$$

where M is the number of detected cells in the image, α and β are thresholds defined for detection and classification uncertainties respectively. Cells having a higher uncertainty are selected for the analysis of the expert. γ weights the contribution of cells that are suspected to be rare classes. In our experiments, we chose $\alpha = 0.80$, $\beta = 0.90$ and $\gamma = 10$.

3   Experiment & Results

Our approach is able to determine the uncertainty of each single cell, calculate the relevance of each image and rank images for annotation accordingly. Fig. 3 shows three exemplary images with different types of informative cells selected by our strategy: cells associated with high detection uncertainty (Fig. 3a), cells associated with high classification uncertainty (Fig. 3b), and cells that are predicted to belong to a rare class (Fig. 3c). We highlight these cells in red boxes. In contrary, cells in green boxes are considered to be less informative for the model and do not require expert review. Images containing many or highly uncertain cells are ranked as highly relevant and presented to the expert for annotation.
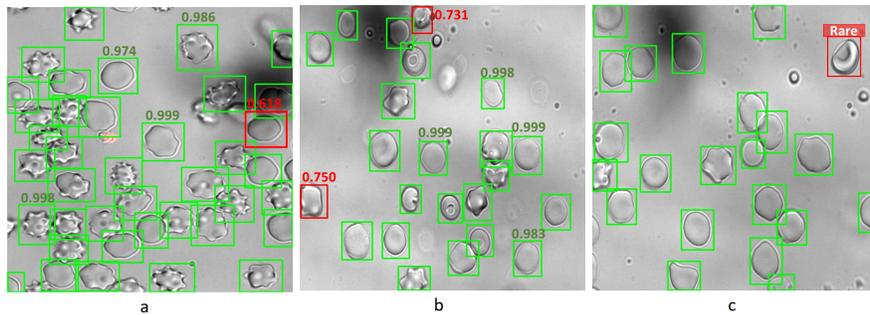
Fig. 3. Exemplary cells (marked with a red box) that are considered to need expert annotation by our uncertainty assessment due to uncertain classification (a), uncertain detection (b) and association to a rare class (c).

We evaluate our active learning annotation tool systematically by comparing its performance with a baseline method where the expert is asked to annotate randomly selected images. In Fig. 4a we show the the object detection precision for all seven classes, weighted by the the number of cells in each class:

$$\text{Precision}_{\text{all}} = \frac{\sum_{k=1}^{K} N_k \times AP_k}{\sum_{k=1}^{K} N_k} \tag{9}$$

where N_k is number of detected cells in class k. This value increases by 5% as we add 1000 newly annotated cells using our active learning annotation tool. In contrast, the performance boost with the same number of randomly annotated cells is slower and around only 2% for 1000 additionally annotated cells. The difference between the two methods is even more pronounced in the detection precision of blood cells of a rare class. The peculiar morphology of dehydrated stomatocytes and a potential over-representation has been linked to disease mutations [1]. Hence an accurate detection of this subtype is clinically highly important by impeded by the rareness of the cells. While dehydrated stomatocytes can be hardly captured by random annotation, our active learning annotation tool highlights cells that are predicted to belong to this rare class and prioritizes them for expert annotations. This leads to a fast increase of the average detection precision of this rare class from around 15% to around 50% for 1000 newly annotated cells (see Fig. 4b), while the average precision is unchanged in the random approach, where few if any dehydrated stomatocytes are annotated among the 1000 randomly selected new annotations.
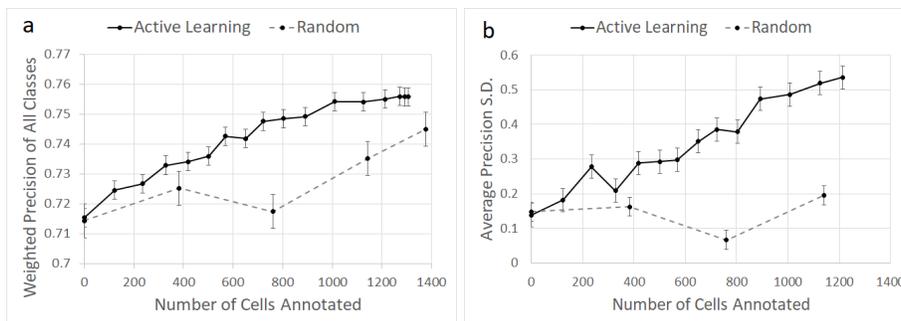
Fig. 4. Our active learning based annotation boosts the precision in cell detection and classification. (a) Weighted detection precision of all classes increases more rapidly when our active learning annotation tool is used (solid line) as compared to a random selection of cells (dashed). (b) Average precision for the rare class of dehydrated stomatocytes, a rare subtype with high clinical relevance, increases sharply when active learning annotation is used. We show the mean and standard deviation from 10 experiments, where we order the images to be newly annotated either randomly, or by sorting 50 randomly selected images according to their relevance score.

4   Conclusion

Our original active learning annotation approach is able to speed up annotation and improve classification of red blood cell sub-types. This is an important task in diagnosis and prognosis of many blood diseases. However, efficient annotation

is urgently required for other biomedical data sets, and in particular for digital pathology applications. An extension of our framework into a software prototype will boost annotated data sets and open new avenues for computational pathology solutions.

References

1. Andolfo, I., Alper, S.L., De Franceschi, L., Auriemma, C., Russo, R., De Falco, L., Vallefuoco, F., Esposito, M.R., Vandorpe, D.H., Shmukler, B.E., et al.: Multiple clinical forms of dehydrated hereditary stomatocytosis arise from mutations in piezo1. Blood **121**(19), 3925–3935 (2013)
2. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A.: Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine **24**(10), 1559 (2018)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09 (2009)
4. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115 (2017)
5. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 1183–1192. JMLR. org (2017)
6. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587 (2014)
8. HW, G.L., Wortis, M., Mukhopadhyay, R.: Stomatocyte–discocyte–echinocyte sequence of the human red blood cell: Evidence for the bilayer–couple hypothesis from membrane mechanics. Proceedings of the National Academy of Sciences **99**(26), 16766–16769 (2002)
9. Ilhan, H.O., Amasyali, M.F.: Active learning as a way of increasing accuracy. International Journal of Computer Theory and Engineering **6**(6), 460 (2014)
10. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in neural information processing systems. pp. 5574–5584 (2017)
11. Minetti, G., Egée, S., Mörsdorf, D., Steffen, P., Makhro, A., Achilli, C., Ciana, A., Wang, J., Bouyer, G., Bernhardt, I., et al.: Red cell investigations: art and artefacts. Blood reviews **27**(2), 91–101 (2013)
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
14. Xu, M., Papageorgiou, D.P., Abidi, S.Z., Dao, M., Zhao, H., Karniadakis, G.E.: A deep convolutional neural network for classification of red blood cells in sickle cell anemia. PLoS Computational Biology **13**(10), e1005746 (2017)