Research Article

State of the Art Cell Detection in Bone Marrow Whole Slide Images

Philipp Gräbel[1], Özcan Özkan[1], Martina Crysandt[2], Reinhild Herwartz[2], Melanie Baumann[2], Barbara Mara Klinkhammer[3], Peter Boor[3], Tim Hendrik Brümmendorf[2], Dorit Merhof[1]

[1]Institute of Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany, [2]Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation, University Hospital RWTH Aachen University, Aachen, Germany, [3]Institute of Pathology, University Hospital RWTH Aachen University, Aachen, Germany

Submitted: 03-Sep-2020 Revised: 29-Apr-2021 Accepted: 23-Aug-2021 Published: 17-Sep-2021

Abstract

Context: Diseases of the hematopoietic system such as leukemia is diagnosed using bone marrow samples. The cell type distribution plays a major role but requires manual analysis of different cell types in microscopy images. **Aims:** Automated analysis of bone marrow samples requires detection and classification of different cell types. In this work, we propose and compare algorithms for cell localization, which is a key component in automated bone marrow analysis. **Settings and Design:** We research fully supervised detection architectures but also propose and evaluate several techniques utilizing weak annotations in a segmentation network. We further incorporate typical cell-like artifacts into our analysis. Whole slide microscopy images are acquired from the human bone marrow samples and annotated by expert hematologists. **Subjects and Methods:** We adapt and evaluate state-of-the-art detection networks. We further propose to utilize the popular U-Net for cell detection by applying suitable preprocessing steps to the annotations. **Statistical Analysis Used:** Evaluations are performed on a held-out dataset using multiple metrics based on the two different matching algorithms. **Results:** The results show that the detection of cells in hematopoietic images using state-of-the-art detection networks yields very accurate results. U-Net-based methods are able to slightly improve detection results using adequate preprocessing – despite artifacts and weak annotations. **Conclusions:** In this work, we propose, U-Net-based cell detection methods and compare with state-of-the-art detection methods for the localization of hematopoietic cells in high-resolution bone marrow images. We show that even with weak annotations and cell-like artifacts, cells can be localized with high precision.

Keywords: Bone marrow, detection, hematopoietic cells

Introduction

Diseases such as leukemia affect the blood-forming process (hematopoiesis), which occurs in the bone marrow. The result is the overproduction of cells which are not able to perform vital functions. Although these negative effects of leukemia can be observed in peripheral blood, a detailed diagnosis requires the analysis of bone marrow samples, which also shows different kinds of immature cells.

Classifying and counting a large enough number of hematopoietic cells from sufficiently many bone marrow samples is necessary to accurately determine the distribution of cell types. In clinical routine, this task is performed manually by trained hematologists. Using modern machine learning methods, automating this task becomes a realistic and appealing option, which would not only reduce the manual burden for diagnosis but also yield reproducible and objective results. A clinician's role would change from time-consuming and error-prone cell counting to validation and supervision of the results, thus enabling faster treatment and more objective cell counts for diagnosis and research purposes.

After the acquisition of Whole Slide Images (WSI), the next necessary step is to localize individual cells (cell detection). This task can already be performed with sufficient precision in

Access this article online

Quick Response Code:

Website: www.jpathinformatics.org

DOI: 10.4103/jpi.jpi_71_20

Address for correspondence: Philipp Gräbel, Lehrstuhl für Bildverarbeitung, Kopernikusstr. 16, 52074 Aachen, Germany. E-mail: graebel@lfb.rwth-aachen.de

How to cite this article: Gräbel P, Özkan Ö, Crysandt M, Herwartz R, Baumann M, Klinkhammer BM, *et al*. State of the art cell detection in bone marrow whole slide images. J Pathol Inform 2021;12:36.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2021/12/1/36/326215

peripheral blood smear microscopy images and (commercial) software for the detection and even classification of leukocytes in peripheral blood is readily available. However, bone marrow microscopy images are much more challenging and cannot be sufficiently processed by such tools. They feature a varying density of cells including very dense areas with cell clusters, which can be difficult to separate into individual cells.[1] Furthermore, a higher number of cell types and several different stages of immature cell types lead to a greater variability in the visual characteristics of cells. Furthermore, the bone marrow images often show many artifacts, which should not be detected despite a cell-like appearance. This includes cells that are damaged during the preparation of the slide.

Similar challenges as in other medical image analysis tasks arise - first and foremost the limited availability of annotated training data. In 2017 and 2018, Song *et al*.[2,3] performed research related to the detection of two groups of cell types in bone marrow images. While they show promising results, the data only has ×40 magnification and only allows distinguishing between two groups of cell types. Chandradevan *et al*.[4] also, propose detection and classification methods on a similar dataset with ×40 magnification. Their detection part consists of only Faster R-CNN, no other architecture is considered. In 2018, the feasibility of using deep convolutional networks for the classification of a larger number of hematopoietic cells were shown.[5] Furthermore, a successful implementation of a segmentation task[6] as well as an improved detection through RetinaNet was presented.[7] Apart from the improved RetinaNet,[7] we are not aware of research on detection of hematopoietic cells incomparable, high-resolution images.

In this work, we investigate the feasibility and performance of several common object detection networks, in particular Mask R-CNN,[8] Yolo v3,[9] and RetinaNet[10] for the detection of hematopoietic cells in bone marrow WSIs. We furthermore propose and analyze several approaches for U-Net[11] based detection of blood cells, which rely on an intermediate segmentation task. These utilize various preprocessing methods and representations for ground truth contours such as weighting based on a distance transform, shrinking of contours, "*Don't Care*" labels, and separate edge labels that enable training a U-Net for instance segmentation with weak annotations. We additionally research the handling of cell-like artifacts and the impact of using weak annotations instead of precise contours.

Our main contributions

We show the feasibility of several one and two-stage detectors for the detection of hematopoietic cells in human bone marrow whole slide microscopy images. Further, we propose and research various ways of preprocessing and representing ground truth data, which makes it possible to utilize the U-Net architecture with weak annotations. We find that including cell-like artifacts in training data or adequately applied weak annotations can improve results.

Compared to the previous state-of-the-art in hematopoietic cell detection, we base our work on more challenging, high-resolution microscopy data, which enables a more detailed analysis. This is necessary for most clinical applications. The U-Net methods introduce novel ways of handling weak annotations for detection tasks.

Subjects and Methods

Image data and preprocessing

Our work is a purely retrospective analysis of bone marrow samples under the Helsinki Declaration of 1975/2000 with written informed consent of all patients. The image data includes one patient with healthy bone marrow and one chronic myelogenous leukemia as well as one acute myeloid leukemia patient each. All samples are pseudonymized with only information about the diagnosed disease available.

Images are acquired from Pappenheim stained bone marrow samples using a whole slide imager with a maximum of ×63 magnification and immersion oil, resulting in extremely large image files. To reduce the file size and acquisition time, a selection of individual representative regions to be scanned in the highest resolution is required. While this could in theory be done automatically by neural networks, it is performed manually in the same way suitable regions are selected in manual diagnosis based on an overview scan. To counteract a possible selection bias, multiple regions are extracted from different areas of the WSI. The annotation is performed in collaboration with two medical experts with many years of training and clinical experience.

For this work, we utilize annotations that not only provide the position and class of each cell type but also the contour. In addition to the various cell types of the hematopoiesis, cell-like artifacts constitute about 40% of the annotated cells. Furthermore, we observe a significant class imbalance with neutrophilic granulocytes making up the largest part of the dataset. Apart from the distinction between artifacts and hematopoietic cells, the classification task is not further investigated in this work.

Weak annotations

Weak annotations are artificially created using the minimal enclosing circle of ground truth contour annotations. To account for the possible differences between the minimal enclosing circle and manually placed circular annotations, we slightly distort each minimal enclosing circle. Firstly, each center coordinate is individually translated by a random number from a Gaussian distribution with 0 mean and a standard deviation of three pixels. Afterward, the radius is dilated by the maximal displacement of the center plus another random number from a Gaussian distribution with 0 mean and a standard deviation of one pixel. This produces weak annotations that mimic human annotations which usually cover the entire cell but do not necessarily have the minimal radius.

The acquisition of weak annotations is significantly faster and much more convenient, resulting in a larger number of annotated cells in a manageable amount of time. This could

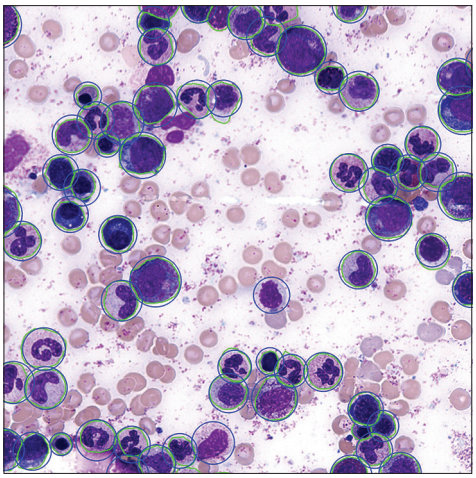

Figure 1: An excerpt from the dataset. The blue circles are the result of the U-Net automated detection method presented below; the green contours show precise ground truth annotations

potentially lead to large improvements with respect to related questions such as the classification of cell types in this dataset.

In total, the dataset used in this work has approximately 7400 annotated cells in 101 image patches of size 2048 × 2048 [Figure 1]. These patches are split into four subsets for training (56%), validation (11%), hyperparameter optimization (11%), and testing (22%). Each set is arranged such that its cell type distribution is similar to the overall distribution. Furthermore, we apply data augmentation to each data in the form of random crop, arbitrary rotation, and slight Gaussian blur.

Methods

Dedicated detection architectures

In a first step, we evaluate the common detection architectures Mask R-CNN, RetinaNet, and Yolo v3. Furthermore, we optimize these architectures to increase their performance for this task.

For each of the networks, we adapt the nonmaximum suppression (NMS) step and utilize the area-based NMS (ANMS), which was presented in 2020.[7] ANMS additionally excludes smaller predictions that are largely within another prediction. This results in a significantly lower number of false-positive detections of, for example, a detected nucleus within another cell. We furthermore use test-time augmentation (two-axis mirroring) and ensembling (with five models), resulting in 20 predictions which are combined using weighted box clustering (WBC). For RetinaNet, we also evaluate a version with circular instead of axis-parallel, rectangular anchors.[7] In the following, this is denoted as Circular RetinaNet. For implementation details, we refer to the Medical Detection Toolkit.[12] All parameters are chosen in an automated hyperparameter search as described in the experimental setup.

U-Net based detection

While the U-Net is a powerful architecture for many medical image segmentation tasks it is not capable to perform instance segmentation (and thus, detection) out-of-the-box. A postprocessing step on the predicted segmentation map is required. This corresponds to the step (b) marked in Figure 2. While Schmidt *et al*.[13] suggested connected component labeling (CCL), we found that the watershed algorithm is better suited to separate cells in dense areas. As we apply marker-based Watershed,[14] background as well as individual cells need to be marked. This is achieved by first Otsu-thresholding[15] the image and then denoising it using morphological opening. To obtain the marker for the background, the cells are dilated and the remaining background is used as marker. For the markers of the cells, a distance transform is applied to the denoised image, followed by another threshold operation. All cells should be separable at this point, such that approximate cell centers can be determined using CCL and precise positions as well as a segmentation using the extracted markers for Watershed.

In another postprocessing step, all predictions smaller than 50 pixels are discarded. This threshold denotes the smallest cell size in the dataset.

Based on different predictions from model ensembling and test time augmentation, we can determine a confidence score similar to dedicated detection architectures. To this end, we use the mean value of all predicted values inside the predicted contour, as these tend to correlate with the confidence of the network in its prediction.

Hyperparameters, including the depth of the U-Net are determined in a hyperparameter optimization step as described in the experimental setup section.

Incorporating weak annotations

To incorporate weak annotations, namely, circles instead of precise ground truth contours, into the training process for each of the presented architectures, we propose several options for preprocessing the annotations for the U-Net-based methods. This becomes necessary, as weak annotations overlap fairly often in contrast to ground truth contours, which have no overlap at all. Some of the following methods are also beneficial for training with precise contours, as it can improve robustness in areas with high cell density. This section corresponds to step (a) in Figure 2.

The first and simplest option is shrinking annotations through erosion. Eroded circles, of course, do not capture the cell perfectly: some parts will be outside the annotations, some parts inside. However, the benefit of having no overlapping annotations is more important.

Of course, this kind of preprocessing requires predicted cells to be dilated with an equal kernel size. It is further possible to use a third class for the cell contour or represent it through a "*Don't Care*" label that is ignored in the loss function. Shrinking the annotations is the default setup in the experiments if not mentioned otherwise.

The second option considers continuous value segmentation masks. Instead of obtaining a binary ground truth from

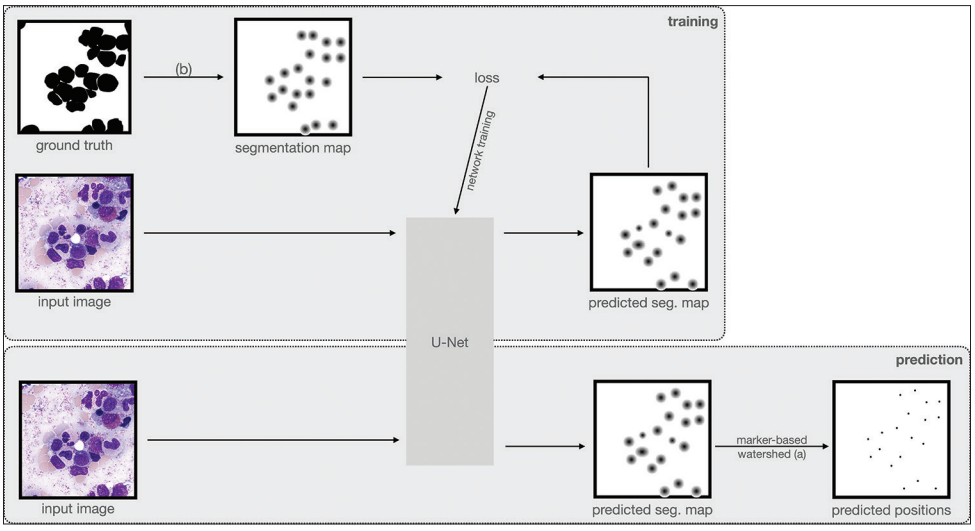

Figure 2: Pipeline for the U-Net based detection, including the steps (a) watershed and (b) segmentation map pre-processing to incorporate weak annotations

weak annotations, segmentation masks in this case can also have values between 0 and 1. The most straightforward implementation of this is using a normalized distance transform that assigns the label 1 (foreground) to the center of a weak annotation and 0 (background) to its outer contour. This already results in much better separable cell annotations. However, if the value is considered as a probability of the foreground label, this is highly unrealistic: close to the center, there is a much higher likelihood of a pixel being a foreground pixel. Consequently, we propose using a sigmoid function (shifted by 0.5 and compressed by 10) on the distance transform values, to get a more realistic curve.

Further analysis of the underlying data reveals a statistical evaluation of the probability of a pixel with a given distance transform value being part of the foreground class. This results in a function that is similar in shape to sigmoid function, but more shifted toward the cell contour and thus steeper. In addition to using this function directly as a continuous representation, we propose a version that goes toward −1 at the cell contour, to make the loss function punish possible false positives at the edge of weak annotations.

The resulting mappings are shown in Figure 3.

As there is no pixel-wise loss (apart from the optional segmentation branch in mask R-CNN, which we disable for weak annotations) in other network architectures, adapting those to weak annotations is straightforward. However, it is also necessary to slightly shrink bounding boxes (BBs), as the overlap of cells makes prediction slightly worse otherwise, particularly through the NMS.

Cell-like artefacts

Analyzing the results of detection architectures shows that a large number of false-positive predictions is due to smudge cells, granules, and other cell-like artifacts. In addition, there are cells that were damaged during sample preparation. These,

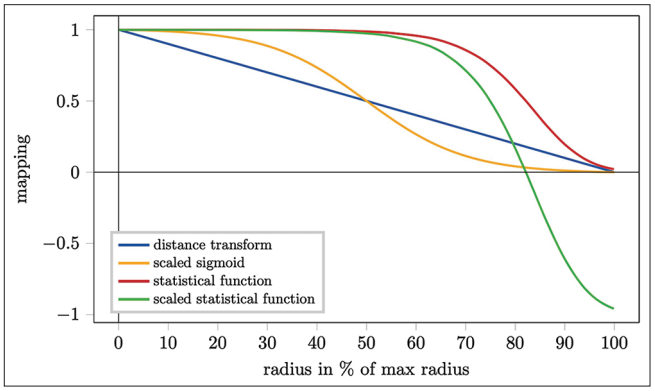

Figure 3: Different mappings for creating continuous segmentation mask

of course, look very similar to valid cells, often differing in no more than a small cut. Examples are shown in Figure 4. While they should not be used for the diagnosis, the amount of those cells is indicative for the quality of a sample. Consequently, they cannot be fully discarded. In general, it is more desirable to have a low false-negative rate than a false-positive rate, as erroneous detections can be filtered out in subsequent steps. In addition to discarding cell-like artifacts by assigning them to the background class, they can either be included in the foreground class and treated like actual cells or be split into a separate class. Both options aim at reducing the number of false-negative predictions for the actual cells.

Results

Setup

Metrics

We apply two metrics to evaluate and compare each method: average precision (AP)[16] and the F1-score (F1). AP is a standard object detection metric, used in popular challenges such as COCO,[17] Pascal VOC,[16] and ILSVRC.[18] This metric combines precision and recall depending on the predicted

confidence. The F1-score is usually applied to evaluate classification results as the harmonic mean of precision and recall. This requires a confidence threshold to determine exactly, which predictions to discard. The threshold is determined individually for each experiment during the hyperparameter optimization.

Furthermore, we consider two possible options to assign a prediction to a corresponding ground truth: matching through BB and matching through center point distance (CD). The former assigns a prediction to a ground truth if the predicted overlap of the BBs is larger than 0.5. The latter assigns a prediction to a ground truth if the center points are no more than 55 pixels apart. This value roughly corresponds to the BB overlap. The CD measure is more interesting if the position is more important than the size. Depending on the use-case (e.g., further classification) this might indeed be the case.

Evaluation pipeline

We perform our evaluation by utilizing four different subsets. The largest set is used for training the networks, a smaller subset is used for validation. This step is repeated for several combinations of hyperparameters, which are evaluated on the optimization subset. The best performing model is retrained including the optimization subset and used to obtain the final results on the held-out test set.

In the hyperparameter optimization, we considered the following parameters (where applicable): U-Net depth (U-Net only), optimizer (including learning-rate, weight decay, and momentum), loss function, patch size, batch size, contour class size, loss weight for contour class, confidence threshold, distance transform threshold for marker-based watershed, NMS threshold, batch versus instance normalization, segmentation branch usage (mask R-CNN only), ANMS threshold, and WBC threshold.

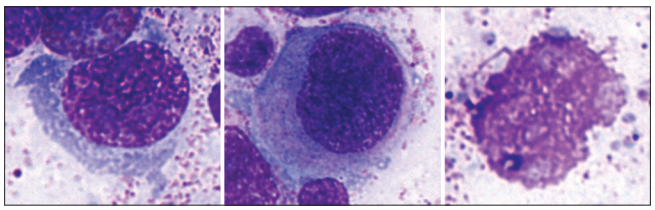

Figure 4: Two cells that were damaged while preparing the sample and one smudge cell

Experiments

This analysis compares the three detection networks Mask R-CNN, RetinaNet, and Yolo v3 as well as U-Net based detection and the RetinaNet with circular anchors (Circular RetinaNet). We first evaluate the detection algorithms with training on precise annotations. Afterward, we evaluate different methods to utilize weak annotations in the U-Net architecture, followed by an evaluation of all detection methods trained on weak annotations. Lastly, we use the U-Net architecture to compare three methods of handling cell-like artifacts.

Results

Precise annotations

Table 1 and Figure 5 show the results for training on precise contour annotations. With respect to most metrics, mask R-CNN yields the best performance, whereas Yolo performs worst in terms of F1-score. In general, matching through CD shows higher metrics compared to BB-based matching, suggesting that the position is estimated more precisely than the size. RetinaNet has slightly better results when using circular anchors. Overall, most scores reach quite high values.

U-Net based detection with weak annotations

Table 2 and Figure 6 show the results of U-Net-based detection using various preprocessing steps on weak annotations. One of the simplest methods, shrinking a weak annotation and adding a separate contour class, yields the overall best results, slightly outperforming shrinking without a separate contour class. Using a "*Don't Care*"-label instead of a contour class, yields worse results, particularly when using BB matching. When comparing the four continuous representations, the statistical function yields the worst results. Although other continuous representations perform better, they are still largely outperformed by noncontinuous segmentation maps.

Weak annotations

The results obtained in the case of weak annotations [Table 3] show one major difference compared to the results with precise annotations. In terms of the F1-score, the U-Net-based method

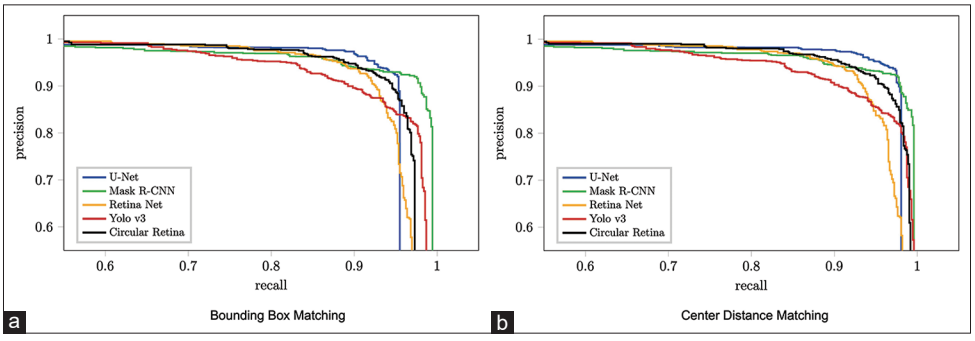

Figure 5: Precision-recall curves for detection usig strong annotations with BB matching [Figure 4a, left] and CD matching [Figure 4b, right]

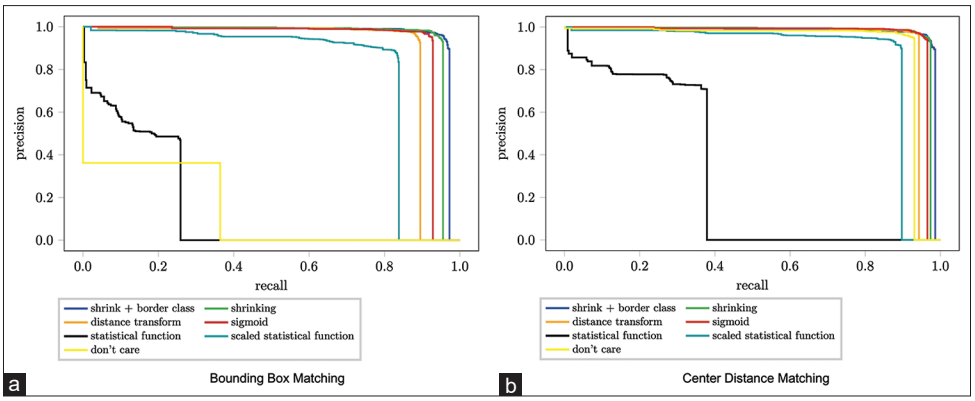

Figure 6: Precision-recall curves for U-net based methods detection using weak annotations with BB matching [Figure 5a, left] and CD matching [Figure 5b, right]

Table 1: Results with precise annotations

Score Matching	F1-score		Average precision	
	BB	CD	BB	CD
U-Net	93.2	95.2	94.2	96.7
Mask R-CNN	94.5	94.6	97.5	97.8
RetinaNet	91.9	92.5	95.2	96.6
Yolo v3	88.8	89.0	96.1	96.9
Circular RetinaNet	92.5	93.4	95.6	97.6

BB: Bounding box; CD: Center point distance

Table 2: Results for U-Net based detection using weak annotations

Score Matching	F1-score		Average precision	
	BB	CD	BB	CD
Shrink + contour class	95.2	95.8	96.3	97.7
Shrinking	94.9	95.7	94.9	96.7
Distance transform	90.7	95.4	88.6	93.6
Sigmoid function	93.1	95.5	92.0	95.9
Statistical function	33.3	49.3	14.9	29.9
Scaled statistical function	84.1	90.1	80.0	87.0
Don't care label	33.9	91.8	13.2	91.7

BB: Bounding box; CD: Center point distance

Table 3: Results with weak annotations

Score Matching	F1-score		Average precision	
	BB	CD	BB	CD
U-Net	95.2	95.8	96.3	97.7
Mask R-CNN	93.0	94.0	97.3	98.2
RetinaNet	93.4	94.2	96.3	97.8
Yolo v3	90.4	90.9	96.4	97.1
Circular RetinaNet	94.0	94.3	96.6	97.5

BB: Bounding box; CD: Center point distance

performs much better than mask R-CNN, even compared to training on precise annotations. In terms of AP, mask R-CNN is still the best performing model.

Again, circular anchors improve the results of RetinaNet slightly for most scores.

Figure 7 shows in detail that U-Net slightly outperforms the other methods, particularly when choosing an appropriate confidence threshold for the detection. Apart from the Yolo architecture, all remaining methods yield similar results.

Inclusion of cell-like artefacts

According to Figure 8, a large number of false-positive predictions belong to the class of cell-like artifacts. At the same time, some blood cells are not detected (false negatives). Including artifacts into the training data reduces, the number of false-positive as well as false-negative predictions for blood cells, whereas the false detection rates for artifacts in the data are increased. With artifacts as a dedicated additional class, the numbers of false positives for blood cells as well as artifacts lie between the other two methods.

Discussion

The results show that most of the detection networks are capable of performing the requested task: finding hematopoietic cells in whole slide microscopy images of human bone marrow. Subsequent (or parallel) classification tasks will benefit from the high resolution, enabling a more detailed analysis compared to the previous state-of-the-art methods[2,3] on human bone marrow images. While several methods exist for the analysis of peripheral blood images, these do not cover the entire hematopoiesis, which is important for the diagnosis of several hematopoietic diseases. Consequently, this paper with its state-of-the-art presentation of detection methods in human bone marrow microscopy images is an important step toward automated analysis tools for hematologists.

The U-Net-based methods are the most complex in terms of available configuration. The results show that a fairly simple standard method performs best: shrinking weak annotations and using a separate class for the border. Thereby, shrinking is a key step and is essential to obtain sensible predictions, as the overlap between neighboring cells would otherwise be too large. This, however, results in an "artificial" segmentation map for the U-Net which has no direct connection to underlying image features anymore. Nevertheless, the receptive field and learning capabilities of the U-Net are still capable to sensible predictions.

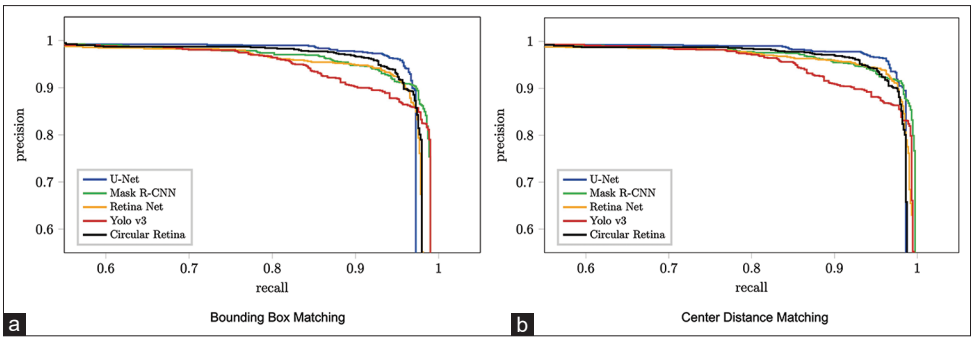

Figure 7: Precision-recall curves for detection using strong annotations with BB matching [Figure 6a, left] and CD matching [Figure 6b, right]

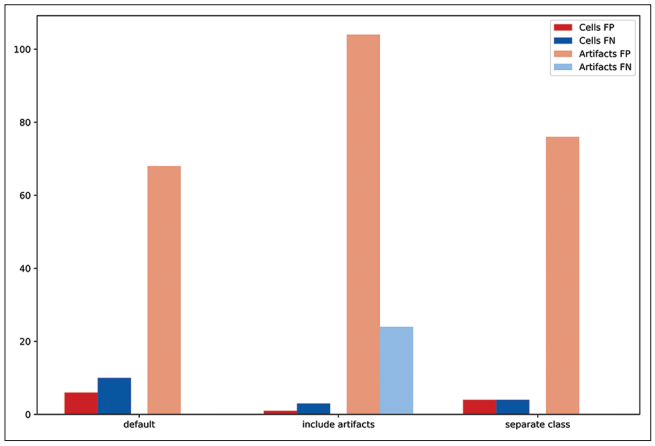

Figure 8: Impact of various methods to handle cell-like artefacts in terms of absolute number of false positive/negative predictions of cells and artefacts

Another major result refers to the difference between BB and center distance matching with " *Don't Care*" labels. When choosing lower IoU thresholds (up to 0.3 instead of 0.5) for BB matching, the results, however, are similar to the other methods. Both can be explained by the results being largely in the right position but with a wrong size estimate. Visualizing the predictions shows that they tend to be too large.

The low performance of the statistical function for the continuous label is also easily explained by comparing it to the binary version. For most of the distance transform values, the statistical function returns values close to one - creating a segmentation map very similar to a barely eroded weak annotation. This leads again to overlapping annotations. The scaled statistical function reduces this effect by penalizing foreground predictions at the annotation border but is not sufficient enough compared to the other methods.

In general, using scaled annotations with a separate class makes the U-Net yield results that even surpass the mask R-CNN, which is the best performing dedicated detection network. Yolo showed similar AP values but had significantly lower F1-scores. This is caused by a relatively low number of false negatives. As already shown in previous work, RetinaNet can be improved using circular anchors while all network architectures benefit from the Advanced NMS.

Furthermore, the influence of precise contours is almost negligible. On the contrary, weak annotations improve detection results slightly for some architectures. This is usually caused by a lower number of false-positive predictions, which likely is a result of shrinking the annotations to avoid overlapping annotations. This results in fewer foreground pixels in the training. Furthermore, weak annotations seem to be particularly robust with small cells such as erythroblasts.

Further experiments will be subject to future work: first of all, it needs to be investigated how well detection architectures perform when simultaneously tasked to classify the cell type. An initial investigation using mask R-CNN with eight classes (seven cell types) suggests that it is possible but requires more training data for underrepresented classes to perform on a similar level. However, even without the classification task, a pure detection pipeline is of great use, for example, for creating a semi-automatic data annotation pipeline. In addition, the work by Chandradevan *et al*.[4] indicates, that a separation between detection and classification can yield sufficient results. Consequently, the detection can be analyzed separately from subsequent tasks.

The low number of unique patients is a limitation of this work. However, since we derive no subject-level characteristics but information about individual objects, the results are meaningful. Already in single slides, the variety in the objects of interest is sufficiently diverse as different regions of a single slide already show great variation in terms of stain and density. In contrast to that, the expected visual changes between patients are relatively low as we do only include slides that rarely show morphological changes caused by a disease.

We additionally performed an evaluation to showcase the generalization capabilities of the best performing U-Net method. To this end, we extended the dataset with slides from four additional patients which specifically contain large visual differences in terms of staining. In total, this includes 8581 cells (excluding artifacts). Reusing well-performing parameters in lieu of hyperparameter optimization yields an AP of 0.923 using centroid distance matching. Despite containing images with strong stain variability, some extreme cases of which would usually be excluded in manual analysis, this is a sufficiently high score compared with the presented results.

Nevertheless, more research with a larger variety of diseases with morphological altered cells is essential for future experiments. This is particularly necessary for the analysis of subsequent or integrated classification tasks.

Lastly, the adaption of other network architectures such as Mask R-CNN to circular anchors might be a useful improvement.

Conclusions

In this work, we investigated the state of the art for the new field of detection of hematopoietic cells in human bone marrow microscopy WSIs. We not only compare dedicated state-of-the-art detection network architectures but also present several ways of utilizing the powerful U-Net architecture for this task. The results show that it is indeed possible to perform accurate detection of blood cells in those images. From all considered detection approaches, the U-Net yields the best results - particularly on weak annotations, which can be obtained more conveniently and efficiently than precise contours.

Acknowledgements

This study was supported by grants of the German Research Foundation (DFG: SFB/TRR57 and SFB/TRR219, BO3755/3-1 and BO3755/6-1), the German Ministry of Education and Research (BMBF: STOP-FSGS-01GM1901A), and the German Ministry of Economic Affairs and Energy (BMWi: EMPAIA project), all to Peter Boor. Furthermore, we received funding from the RWTH Aachen Exploratory Research Space (ERS Seed Fund: OPSF585) to Dorit Merhof and Barbara Klinkhammer.

Financial support and sponsorship

This study was supported by grants of the German Research Foundation (DFG: SFB/TRR57 and SFB/TRR219 and BO3755/6-1), the German Ministry of Education and Research (BMBF: STOP-FSGS-01GM1901A), and the German Ministry of Economic Affairs and Energy (BMWi: EMPAIA project), all to Peter Boor. Furthermore, we received funding from the RWTH Aachen Exploratory Research Space (ERS Seed Fund: OPSF585) to Dorit Merhof and Barbara Klinkhammer.

Conflicts of interest

There are no conflicts of interest.

References

1. Nilsson B, Heyden A. Segmentation of dense leukocyte clusters. In: Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. Kauai, HI, USA; 2001. p. 221-7.
2. Song TH, Sanchez V, Eldaly H, Rajpoot NM. Hybrid deep autoencoder with Curvature Gaussian for detection of various types of cells in bone marrow trephine biopsy images. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017): Melbourne, VIC, Australia; 2017. p. 1040-3.
3. Song TH, Sanchez V, Eldaly H, Rajpoot N. Simultaneous cell detection and classification in bone marrow histology images. IEEE J Biomed Heal Inform 2018;23; p-1469-76
4. Chandradevan R, Aljudi AA, Drumheller BR, Kunananthaseelan N, Amgad M, Gutman DA, *et al.* Machine-based detection and classification for bone marrow aspirate differential counts: Initial development focusing on nonneoplastic cells. Lab Invest 2020;100:98-109.
5. Gräbel P, Crysandt M, Herwartz R, Hoffmann M, Klinkhammer BM, Boor P, *et al.* Evaluating out-of-the-box methods for the classification of hematopoietic cells in images of stained bone marrow. In: Computational Pathology and Ophthalmic Medical Image Analysis. Granada, Spain: Springer; 2018. p. 78-85.
6. Gräbel P, Crysandt M, Herwartz R, Baumann M, Klinkhammer BM, Boor P, *et al.* Refinement of weak annotations for the segmentation of bone marrow leukocytes. In: 2nd MICCAI Workshop on Computational Pathology (COMPAY). Shenzhen, China; 2019.
7. Gräbel P, Özkan Ö, Crysandt M, Herwartz R, Baumann M, Klinkhammer BM, *et al.* Circular anchors for the detection of hematopoietic cells using retinanet. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa City, IA, USA; 2020. p. 249-53.
8. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In: The IEEE International Conference on Computer Vision (ICCV). Venice, Italy; 2017.
9. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA; 2016.
10. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell 2020;42:318-27.
11. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer: Munich, Germany; 2015. p. 234-41.
12. Jaeger PF, Kohl SA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer HP, *et al*. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. arXiv 2018; Proceedings of the Machine Learning for Health NeurIPS Workshop, PMLR; 2020;116:171-83,
13. Schmidt U, Weigert M, Broaddus C, Myers G. Cell detection with star-convex polygons. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer: Munich, Germany; 2018. p. 265-73.
14. Beucher S, Meyer F. The morphological approach to segmentation: The watershed transformation. Opt Eng 1992;34:433.
15. Otsu N. Threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 1979;9:62-6.
16. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. Int J Comput Vis 2010;88:303-38.
17. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, *et al.* Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer: Munich, Germany; 2014. p. 740-55.
18. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, *et al.* Imagenet large scale visual recognition challenge. Int J Comput Vis 2015;115:211-52.