

SlideRunner - A Tool for Massive Cell Annotations in Whole Slide Images

Marc Aubreville^a, Christof Bertram^b, Robert Klopffleisch^b, Andreas Maier^a

^a*Pattern Recognition Lab, Computer Sciences, Friedrich-Alexander-Universität
Erlangen-Nürnberg*

^b*Institute of Veterinary Pathology, Freie Universität Berlin, Germany*

1. Introduction

Whole-slide images enable automated image analysis in histopathology and is a well-recognized and legally approved method for quantification of immunohistochemistry in research and diagnosis. It has, beyond that, a great potential to further support the pathologists' work in monotoneous, time-consuming tasks such as counting rare events [1]. The availability of large amounts of data has also paved the way for pattern recognition methods that can be used to automatically pre-annotate and analyze whole-slide images. Especially since the advent of deep learning, the need for high-quality data annotation at a large scale has increased significantly, but so have the prospects of high-quality detection results.

The demand for those large-scale data sets poses two major problems to the annotation task, that are currently unsolved for many medical image recognition domains: The lack in quantity in data, and the lack in quality of labels. To overcome the limitations in quantity of expert label data, Albarqouni *et al.* used an aggregation of expert labels and crowdsourcing with non-experts [2], showing effectively the power of an enhanced data set. However, it can be stated that expert-labelled data is generally of higher quality than non-expert labelled data.

High rating variance in expert annotations

Medical images in general, and histology images in particular, often have a high inter-observer variance in rating. As Boiesen *et al.* report, subjective differences in grading in breast cancer histology tumor diagnostics can lead to a significantly high variance in the tumor classification [3], which in term has implications on an individual, targeted curative treatment. This, of course, calls for computer-based assistance systems, which could significantly contribute to a highly standardized diagnosis and thus could lead to a more uniform classification.

Further, histo-pathology images are often annotated in an unsuitable way for supervised machine learning on a segmentation task (e.g. just by an arrows pointing at structures, using circles or as per-slide annotation). While this

might be a suitable and sufficient annotation for a human observer, it is often too ambiguous for pattern recognition methods.

Prospects of multi label data sets

For algorithmic approaches, the quality of input data is a major bottleneck for the quality of the overall outcome. In a machine learning sense, we have to consider the labels provided by experts, usually considered as ground truth, to be noisy as well. Besides human error, also the difficulty of a certain data subset will play a role. The closer a single object is to a prototype, the easier the recognition for a human as well as for a machine learner, an effect which already successfully employed in other machine learning domains [4].

Bengio *et al.* suggest to exploit this difficulty in order to improve on generalization of machine learning approaches, effectively mimicking human learning, which is why they coined the term curriculum learning [5]. While there are other approaches to provide this from the confidence of machine learning approaches inherently, a very strong indicator is human error. Multi label expert data sets provide thus the possibility to differentiate prototypical from difficult samples, while at the same time supplying the learning system with a better estimate for the ground truth.

Existing software

Besides commercially available digital histopathology software, e.g. provided by the manufacturers of the scanner hardware, there is a high number of open source software solutions available for slide viewing and analysis [1], with some products supporting whole-slide images and others using standard graphics formats. Many solutions provide not only annotation capabilities, but also plug-in systems for automated analysis or pre-processing (e.g. Icy [6] or CellProfiler [7]). As de Chaumont *et al.* reported, the emphasis of these projects was collaborative design and evaluation of tools and algorithms within the bioinformatics community. To extend the collaborative approach to the annotation, Marée *et al.* released Cytomine, an internet-based general-purpose annotation tool [8], with full integration of whole-slide images and respective on-demand loading. Since within the tool, annotations are represented in layers, it is possible to do a blind annotation with multiple experts, which is valuable for shape estimations (e.g. for tumor regions) in a blinded manner. For cell type annotations, however, multiple opinions on one cell are of great importance to judge the prototypical character of an occurrence and hence the expected difficulty.

In order to improve on current data sets, we can formulate requirements on software that would be used to aggregate data sets with multiple labels on the same cells and using big data amounts, which we do not find entirely fulfilled in state of the art solutions (see Table 1):

- The user interface must be intuitive to use and annotations can be set with very little interactions, using one click only if possible.
- The software must be able to support blinded class labeling, where the annotated cell is displayed but the label of previous annotators is hidden.

Table 1: Comparison of a selection of open source software packages suited for cell annotations.

Product	Icy [6]	Qupath [9]	Cytomine [8]	SlideRunner
Whole-Slide Image support	Yes (bio-formats)	Yes (openslide)	Yes (openslide)	Yes (openslide)
Very large image support	No (limit at 2.1 Gigapixels)	Yes	Yes	Yes
Multi-user Annotation	No	No	Yes	Yes
Single Click Cell Annotations	No	Yes	No	Yes
Blind Annotations	No	No	Yes	Yes
Blind Multi-labels	No	No	No	Yes
Guided Screening Process	No	No	No	Yes

2. Methods

SlideRunner is a GPL-licensed tool¹, written in Python 3.5 and using OpenSlide [10] as image loading backend. It provides several modes of navigation, and two major modes of operation for annotation:

- **Center annotation:** In this mode a single mouse click is needed to add an annotation in the center of an object, e.g. a cell.
- **Outline annotation:** Either using multiple clicks or click-and-drag, polygon curves can be added and annotated.

Data model

As depicted in Fig. 1(b), the major entity of our database model is the annotation. Each annotation can have multiple coordinates, with their respective x and y coordinates, and the order they were drawn (for polygons). Further, each annotation has a multitude of labels that were given by one person each and are belonging to one class, respectively.

Blinded Annotation

For blinded multi-annotation, it is important that the viewer is provided with a view that does present the area or object annotations of other experts, yet hides the information about the class the object was given but those other raters. For this, SlideRunner provides a mode where only own annotations are provided with (color-based) class informations in the image (see Fig. 1(a)).

To leverage more gains in annotation performance, this mode is enriched by a *discovery mode*, in which the user is automatically presented with a random new image section upon completion of the currently visible view until all annotated objects have been classified successfully. This mode is expected to considerably reduce required time to reach for the next unclassified cells to annotate.

Guided Screening

For the first annotation of the image, a guided screening mode is being provided that guides the expert at maximum optical zoom over the complete

¹available at: <https://github.com/maubreville/SlideRunner>

Table 2: Confusion matrix between two pathologists. Most significant disagreement was between the ambiguous class and the clear classes.

	granulocytes	mitotic cells	normal tumor cells	ambiguous
granulocytes	10318	395	327	2249
mitotic cells	147	30623	202	458
normal tumor cells	27	546	18445	387
ambiguous	257	2949	1331	2420

image. This mode will ensure that the observer will definitely examine every field of view of the image and is therefore especially helpful if complete annotation coverage is desired. In order to only display fields of interest containing tissue sections, white/empty areas need to be excluded from the image, which is performed by the algorithmic chain depicted in Fig. 2. After thresholding and morphologic closing, a grid is projected on the image, and non-empty slide partitions are shown to the user.

3. Results

With the means described in this text, a double-annotated database of sparsely annotated cell types from canine mast cell tumor slides was established. Histo-pathologically, canine cells show great similarity to human cells including the appearance and diversity of mitotic figures. As Table 2 shows, we found a good agreement between the raters in general, with Cohen’s $\kappa = 0.815$. Most disagreement can be found between clear decisions for one of the cell classes and the ambiguous class. For a curriculum learning-based approach, this may be a good hint towards the difficulty of detection. Mean annotation times (measured as time difference between annotation events) were 6.6 s and 6.3 s for both raters for first annotations, and 2.0 s and 2.6 s for second annotations, respectively (evaluated on N=71,561 labels).

For one particular slide, one pathologist performed full annotation of mitotic figures manually, resulting in 2,252 single mitotic events distributed over the slide. For the same slide, a second pathologist performed an annotation using the guided screening mode, resulting in 4,233 mitotic events.

4. Discussion

While minor differences were expected, the significant increase in mitotic figure annotations for usage of guided screening is surprising at first. Potentially, this effect can be attributed to a more thorough annotation in this mode, where the expert’s attention is not focused on the center of the image. In general, statistics between both pathologists working in the same mode did not differ significantly for the whole data set.

We find that using the methods realized in SlideRunner lead to a fast annotation process of mitotic figures and other cell types. In part, however, this

might also be related to a generally high mitotic count in the slides that have been labeled. In general, yet, using the tool provided means to build up a database of mitotic and other cell annotations that is unprecedented in size and could leverage precision gains provided by machine learning methods.

References

- [1] C. A. Bertram, R. Klopffleisch, The Pathologist 2.0: An Update on Digital Pathology in Veterinary Medicine, *Vet Pathol* 54 (2017) 756–766.
- [2] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, N. Navab, AggNet - Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images., *IEEE Trans Med Imaging* 35 (2016) 1313–1321.
- [3] P. Boiesen, P. O. Bendahl, L. Anagnostaki, H. Domanski, E. Holm, I. Idvall, S. Johansson, O. Ljungberg, A. Ringberg, G. Östberg, M. Fernö, Histologic grading in breast cancer: reproducibility between seven pathologic departments, *Acta Oncol* 39 (2000) 41–45.
- [4] A. Batliner, S. Steidl, C. Hacker, E. Nöth, H. Niemann, Tales of Tuning - Prototyping for Automatic Classification of Emotional User States (2005) 489–492.
- [5] Y. Bengio, J. Louradour, R. Collobert, J. Weston, Curriculum learning (2009) 41–48.
- [6] F. de Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, T. Lagache, A. Dufour, J.-C. Olivo-Marin, Icy: an open bioimage informatics platform for extended reproducible research, *Nat Methods* 9 (2012) 690–696.
- [7] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, D. M. Sabatini, CellProfiler: image analysis software for identifying and quantifying cell phenotypes, *Genome Biology* 7 (2006) R100.
- [8] R. Marée, B. Stévens, L. Rollus, N. Rocks, X. Lopez, I. Salmon, D. Cataldo, L. Wehenkel, A rich internet application for remote visualization and collaborative annotation of digital slides in histology and cytology, *Diagn Pathol* 8 (2013) S26.
- [9] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, P. W. Hamilton, QuPath: Open source software for digital pathology image analysis, *Sci Rep* 7 (2017) 16878.
- [10] A. Goode, B. Gilbert, J. Harkes, D. Jukic, M. Satyanarayanan, OpenSlide: A vendor-neutral software foundation for digital pathology., *J Pathol Inform* 4 (2013) 27.

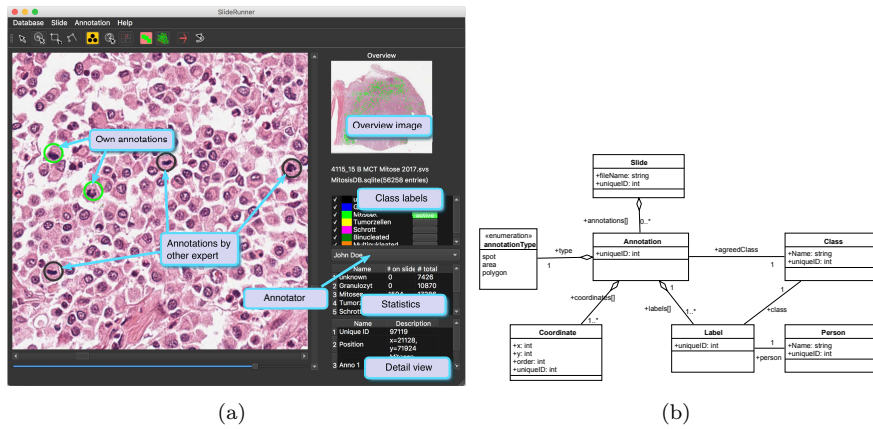


Figure 1: a) GUI overview in blinded mode. In this mode, annotations by other experts are only visible as unknown classes (in black). b) UML diagram of database structure. Every annotation may contain multiple labels from different persons and multiple coordinates.

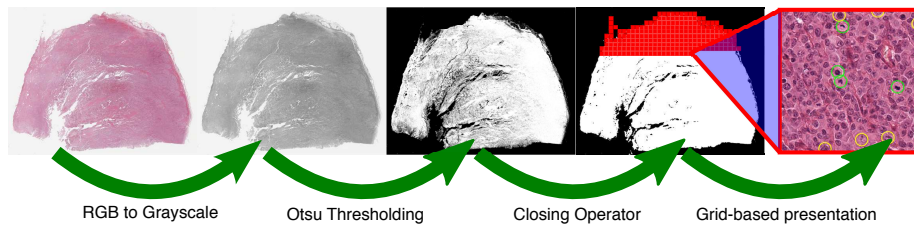


Figure 2: Algorithmic toolchain for guided screening. Grid-based segments are presented from left to right, top to bottom.