The Journal of Open Source Software

InspectorCell: Finding Ground Truth in Multiplexed Microscopy Images

Andre Gosselink [1,2*], **Tatsiana Hofer**[3*], **Elvira Criado-Moronati**[2], **Arndt von Haeseler**[3,4], **and Jutta Kollet**[2]

1 Institute of Medical Statistics and Computational Biology, University of Cologne, Bachemer Str. 86, 50931 Cologne, Germany **2** R&D Department, Miltenyi Biotec B.V. & Co. KG, Friedrich Ebert Straße 68, 51429 Bergisch Gladbach, Germany **3** Center for Integrative Bioinformatics Vienna, Max Perutz Labs, University of Vienna, Medical University of Vienna, Dr. Bohr Gasse 9, 1030 Vienna, Austria **4** Faculty of Computer Science, University of Vienna, Währinger Str. 29, 1090 Vienna, Austria * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Multiplexed immunofluorescence microscopy produces large image stacks of immunologic tissue sections. Cells in these stacks can be segmented and classified by supervised machine learning methods. However, training these models requires high-quality labeled datasets. With recent increases in image stack sizes, the generation of labeled datasets for supervised machine learning has become a major bottleneck. InspectorCell alleviates this bottleneck by providing an intuitive, graphical interface for synchronized manual segmentation and annotation of cells in highly multiplexed microscopy images. The modular implementation of InspectorCell in Python enables tight integration into existing applications such as Orange3 or CellProfiler. A image dataset with exampelary annotations is available at: https://doi.org/10.7303/syn37910913.2

Statement of need

The cellular composition of tumors is of great scientific interest as the presence of tumor-infiltrating lymphocytes correlates with the survival of cancer patients (Idos et al., 2020; Santoiemma & Powell, 2015). The cellular phenotypes can be investigated with multiplexed tissue imaging methods such as CODEX (Goltsev et al., 2018) or MACSima™ (Kinkhabwala et al., 2022). These techniques generate large stacks of images of the same tissue slice, where each image covers the intensity profile of a different marker. Segmentation then enables single-cell analysis to identify cell types by spatially co-localized markers. Once the datasets are segmented and annotated for the presence of the markers, they can be used in supervised machine learning for feature extraction and classification tasks. Nevertheless, this workflow has a bottleneck already in generating segmented and annotated training data because current software does not provide synchronized viewing, editing, and annotation of multiple images.

While it is possible to edit cell segments in CellProfiler (Carpenter et al., 2006), only a single channel can be evaluated simultaneously. In FIJI (Schindelin et al., 2012) cells can be annotated when several immune staining images are displayed in parallel. However, the different views of the sample are not synchronized. An annotation or change of a cell segment must be performed in each window individually, but finding the same location on all images is very difficult. In ilastik (Sommer et al., 2011) the generation of training datasets are fused to the training of a machine learning classifier. However, the user can evaluate only one image at a time and therefore misses information necessary for annotation. Hence, a synchronized overview of all channels of the image stack is crucial for evaluating or editing a cell segmentation. Such an

overview is also needed for annotating the localized marker expression as high or low intensity for each image of the stack. With InspectorCell we provide a solution for efficient manual segmentation and annotation of large image stacks. The modular implementation in Python 3 enables extension to existing software solutions such as CellProfiler.

The primary benefit of using InspectorCell is the ability to view cells within the context of multiple immunomarkers. It accelerates manual segmentation and annotation of cells in multiplexed immunoflu- orescence images. Expert immunologists can use it to evaluate immunological and morphological information of cells at a glance to rapidly generate high-quality cell segmentations with annotations. This can be saved as a JSON file for downstream applications or stored in databases. The application addresses the need for software solutions to generate ground truth training datasets of highly multiplexed immunofluorescence images. InspectorCell accelerates the ad-hoc creation of high-quality training and validation datasets needed in biological image analysis by machine learning.

Results

We used InspectorCell to generate an exemplary training dataset from multiplexed immunofluorescence microscopy images of an ovarian cancer tissue section, obtained with the MACSima™ imaging platform 2 (Miltenyi Biotec B.V. & Co. KG). The tissue was stained with Hoechst and 98 antibodies against various cluster of differentiation (CD) proteins, conjugated with phycoerythrin. For the generation of the training dataset CD103, CD3, CD326, CD4, CD45, and CD8 were evaluated. A pixel-based segmentation map was generated with CellProfiler. The immunofluorescence images and segmentations were imported into InspectorCell (fig. Figure 1a). A typical problem in the CellProfiler output was oversegmentation, which was corrected with Inspec- torCell by merging (fig. Figure 1b). Furthermore, some segments were extended to encompass the complete area of distinctly stained cells (fig. Figure 1c). Finally, annotations were directly applied to the cell segments, using keyboard shortcuts. The segmentation map that resulted from this workflow was compared to the initial CellProfiler segmentation. From the initial 1960 segments, only 1750 remained after correction. Thus, at least 10 % of the segments in this example have originally been over-segmented. Since we elimi- nated the bias of oversegmentation, we anticipate better performance in machine learning applications with our InspectorCell derived ground truth dataset than with the original oversegmented single cell data as a training dataset.

The Journal of Open Source Software

Figures

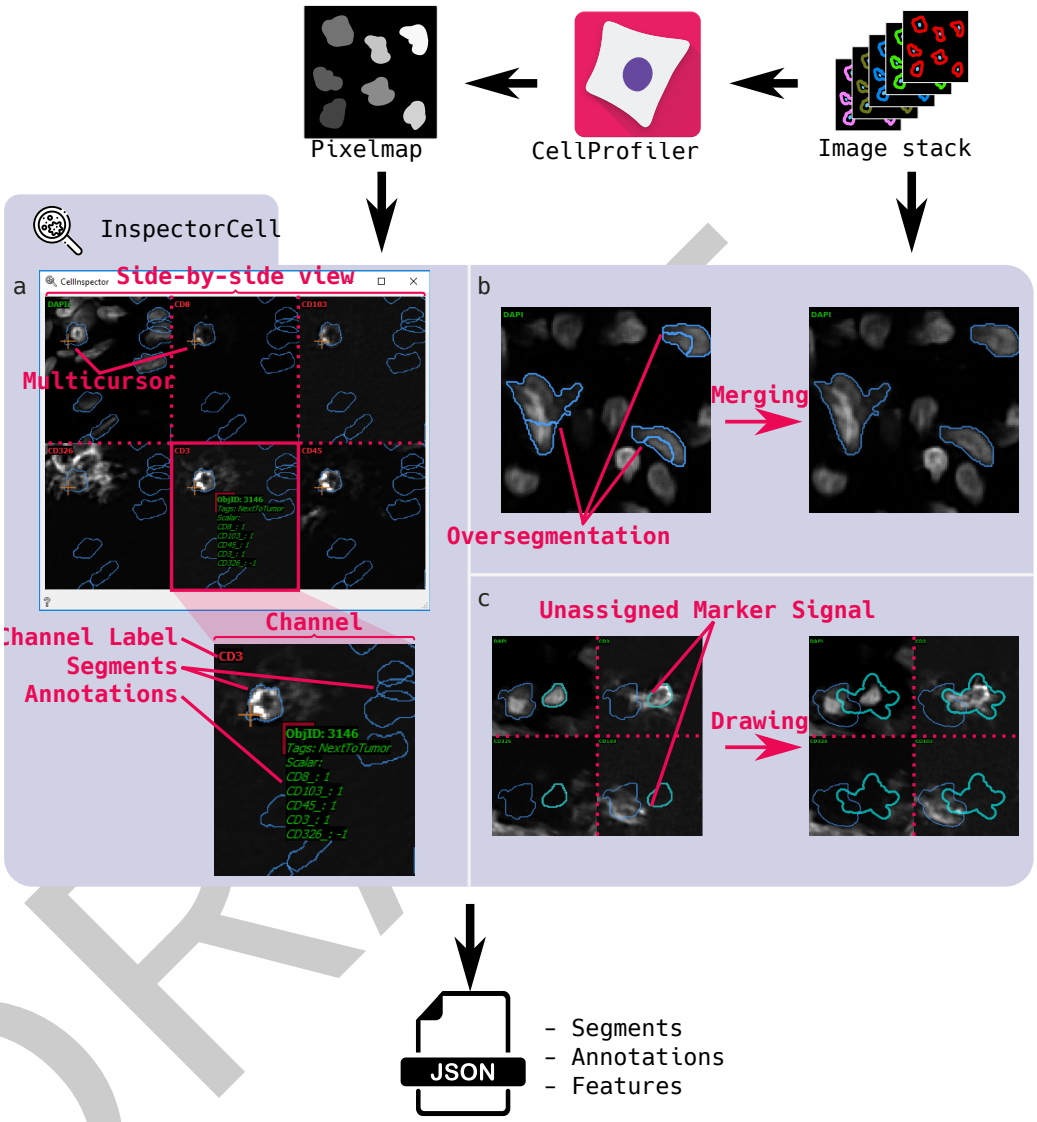

Figure 1: Exemplary use cases of InspectorCell. Segmentation of image stacks are generated with CellProfiler and opened with InspectorCell. (a) Six immunofluorescence images of the image stack are displayed side- by-side in a 3×2 grid. Cell segments are displayed as blue polygons (top). A synchronized multi cursor (orange) is a visual anchor in all channels. The CD3 channel (lower middle) is enlarged below the main window. The cell segment annotation is editable and displayed for the active cell segment in green font. (b) Over-segmentation can be merged with a single keystroke after mouse selection. (c) A cell segment (light blue) can be edited to embrace the marker signal area attributable to a distinct cell. Multiple segments can enclose the same areas to reflect cell overlaps and interactions. The manual edits of cell segmentations and annotations are saved in a single JSON file. Additionally, the JSON file can store extracted cell features, for example, mean pixel intensities. The JSON file is in plain text and can be readily used in downstream analysis.

Acknowledgements

We thank Ali Kinkhabwala for his help during development, Bianca Heemskerk for releasing the data, Paurush Praveen, Werner Müller and Achim Tresch for reviewing the manuscript.

This project was partially funded by Miltenyi Biotec B.V. & Co. KG.

References

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P., & Sabatini, D. M. (2006). CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, *7*(10), R100. https://doi.org/10.1186/gb-2006-7-10-r100

Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., & Nolan, G. P. (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*, *174*(4), 968–981. https://doi.org/10.1016/j.cell.2018.07.010

Idos, G. E., Kwok, J., Bonthala, N., Kysh, L., Gruber, S. B., & Qu, C. (2020). The prognostic implications of tumor infiltrating lymphocytes in colorectal cancer: A systematic review and meta-analysis. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60255-4

Kinkhabwala, A., Herbel, C., Pankratz, J., Yushchenko, D. A., Rüberg, S., Praveen, P., Reiß, S., Rodriguez, F. C., Schäfer, D., Kollet, J., Dittmer, V., Martinez-Osuna, M., Minnerup, L., Reinhard, C., Dzionek, A., Rockel, T. D., Borbe, S., Büscher, M., Krieg, J., … Bosio, A. (2022). MACSima imaging cyclic staining (MICS) technology reveals combinatorial target pairs for CAR T cell treatment of solid tumors. *Scientific Reports*, *12*(1), 1911. https://doi.org/10.1038/s41598-022-05841-4

Santoiemma, P. P., & Powell, D. J. (2015). Tumor infiltrating lymphocytes in ovarian cancer. *Cancer Biology & Therapy*, *16*(6), 807–820. https://doi.org/10.1080/15384047.2015.1040960

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Sommer, C., Strähle, C., Köthe, U., & Hamprecht, F. A. (2011). *Ilastik: Interactive learning and segmentation toolkit* (pp. 230–233). https://doi.org/10.1109/ISBI.2011.5872394