



User-Accessible Machine Learning Approaches for Cell Segmentation and Analysis in Tissue

Seth Winfree*

Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, United States

Advanced image analysis with machine and deep learning has improved cell segmentation and classification for novel insights into biological mechanisms. These approaches have been used for the analysis of cells *in situ*, within tissue, and confirmed existing and uncovered new models of cellular microenvironments in human disease. This has been achieved by the development of both imaging modality specific and multimodal solutions for cellular segmentation, thus addressing the fundamental requirement for high quality and reproducible cell segmentation in images from immunofluorescence, immunohistochemistry and histological stains. The expansive landscape of cell types—from a variety of species, organs and cellular states—has required a concerted effort to build libraries of annotated cells for training data and novel solutions for leveraging annotations across imaging modalities and in some cases led to questioning the requirement for single cell demarcation all together. Unfortunately, bleeding-edge approaches are often confined to a few experts with the necessary domain knowledge. However, freely available, and open-source tools and libraries of trained machine learning models have been made accessible to researchers in the biomedical sciences as software pipelines, plugins for open-source and free desktop and web-based software solutions. The future holds exciting possibilities with expanding machine learning models for segmentation via the brute-force addition of new training data or the implementation of novel network architectures, the use of machine and deep learning in cell and neighborhood classification for uncovering cellular microenvironments, and the development of new strategies for the use of machine and deep learning in biomedical research.

OPEN ACCESS

Edited by:

Bruce Molitoris,
Indiana University, United States

Reviewed by:

Noriko F. Hiroi,
Keio University Shonan Fujisawa
Campus, Japan

*Correspondence:

Seth Winfree
swinfree@unmc.edu

Specialty section:

This article was submitted to
Computational Physiology
and Medicine,
a section of the journal
Frontiers in Physiology

Received: 13 December 2021

Accepted: 12 January 2022

Published: 10 March 2022

Citation:

Winfree S (2022) User-Accessible
Machine Learning Approaches
for Cell Segmentation and Analysis
in Tissue. *Front. Physiol.* 13:833333.
doi: 10.3389/fphys.2022.833333

Keywords: machine learning, deep learning—artificial neural network, segmentation, classification, neighborhoods, microenvironment, bio-imaging tools

INTRODUCTION

Image use in the biomedical sciences varies from demonstrative and representative to data for quantitative interrogation. Quantitative analyses of tissue and cells, the basic building blocks in biology, requires the accurate segmentation of cells or surrogates of cells and methods for classifying cells and quantitative analysis of cell type, cell states and function. Cellular segmentation has been an intense focus in biomedical image analysis for decades and has evolved from largely *ad hoc* approaches to generalizable solutions (Meijering, 2012). The classification strategies for cell type (e.g., immune cell, epithelium, stromal, etc.) and state (e.g., injured, repairing, dividing) have

Abbreviations: ML, machine learning; DL, deep learning; GUI, graphical user interface; API, application programming interface.

developed rapidly (Meijering, 2012, 2020; Meijering et al., 2016). How different cell types organize into microenvironments or neighborhoods is important for our understanding of pathogenesis and biology. The identification and classification of these neighborhood or microenvironments is of significant interest to the bioimaging community (Allam et al., 2020; Stoltzfus et al., 2020; Solorzano et al., 2021). This mini review will cover the current state of quantitative analysis of tissues and cells in imaging data, with a discussion of segmentation, classification, and neighborhood analysis, specifically highlighting the application of machine learning, including recent advancements, challenges, and the tools available to the biomedical researchers.

SEGMENTATION

A cornucopia of segmentation approaches has been developed for specific experimental situations, tissue types or cell populations including clusters of cells, specific cell types, etc. (Meijering, 2012; Meijering et al., 2016). Often these approaches are built as pipelines in image processing software, enabling the sharing of segmentation methods (Berthold et al., 2007; de Chaumont et al., 2012; Schindelin et al., 2012; Bankhead et al., 2017; McQuin et al., 2018; Berg et al., 2019). A common approach is to first differentiate foreground, the cell, from background in a semantic segmentation step. Secondly, objects of interest in the image are isolated, or instance segmentation, by identifying and then splitting touching cells. Meijering outlined five fundamental methods for segmentation: intensity thresholding (Otsu, 1979), feature detection, morphologically based, deformable model fitting and region accumulation or splitting (Meijering, 2012). These methods are often combined sequentially. For instance, cell segmentation might include *semantic* segmentation of a foreground of all nuclei with pixel intensity, followed by a second *instance* segmentation for identifying an individual nucleus using a region accumulation approach like watershed (Beucher and Lantuejoul, 1979). A common limitation is the *ad hoc* nature of segmentation approaches: the applicability of a segmentation method may be limited by constraints in the datasets including differences in staining or imaging modality (fluorescence vs. histology staining), artifacts in image capture (out-of-focus light or uneven field illumination) or morphological differences (spherical epithelial vs. more cylindrical muscle nuclei). These constraints, and others, have limited the development of generalizable segmentation algorithms.

Cell segmentation with machine learning is well established—a popular approach is to perform semantic pixel segmentation with a Random Forest Classifier (Hall et al., 2009; McQuin et al., 2018; Berg et al., 2019). Segmentation with a Random Forest Classifier, as with all machine learning approaches, requires training data. In cell segmentation this is data that has been annotated to indicate which pixels in images are foreground, nuclei, vs. background. *ilastik* provides an intuitive and iterative solution for generating training data with a GUI that allows a user to: (1) highlight pixels to indicate nuclei vs. background-training data, (2) test classification and segmentation, (3) repeat and add or subtract highlighted pixels, to improve the classification and segmentation. This process is powerful but can become

labor-intensive in different tissues where there may be a variety of nuclei (e.g., shape, texture, size, clustering, etc.) in smooth muscle, epithelium, endothelium, and immune cells in varying densities and distributions. Unfortunately, while high quality cell culture nuclei training datasets and tissue image datasets exist, 2D training data of nuclei in tissue is limited or fractured across multiple repositories (Ljosa et al., 2012; Williams et al., 2017; Ellenberg et al., 2018; Kume and Nishida, 2021). Furthermore, while 3D electron microscopy data is readily available, 3D fluorescence image or training datasets of nuclei is limited (Ljosa et al., 2012; Iudin et al., 2016; El-Achkar et al., 2021; Lake et al., 2021). The availability of training data is one of the most significant barriers to the application of machine learning to cell image segmentation (Ching et al., 2018). Fortunately, the number of venues to share imaging datasets should not limit the dissemination of training datasets as they are generated (**Table 1**, Datasets and Repositories).

Recently, three novel approaches were developed to address the dearth of segmentation training data for the variety of cell-types and imaging modalities. The first, and most direct approach has been the concerted effort of a number of groups including the Van Valen and Lundberg laboratories to establish “human-in-the-loop” pipelines and infrastructure of software and personnel, including collaborative crowd sourcing, to generate ground truth from imaging datasets (Sullivan et al., 2018; Moen et al., 2019; Bannon et al., 2021). A limitation of this approach is the requirement for on-going support for personnel; on-going support is critical to long term success. To ease the generation of high-quality training data with a “human-in-the-loop” approach, methods have also been established around segmentation refinement (Sullivan et al., 2018; Lutnick et al., 2019; Moen et al., 2019; Govind et al., 2021; Lee et al., 2021). An alternative to these brute-force approach has been to generate synthetic training data by combining “blob” models of cells with real images using generative adversarial networks (Dunn et al., 2019; Wu et al., 2021). Further, to leverage training data across imaging modalities *NucleAIzer*¹ relies on style transfer with a generative-adversarial-network to generate synthetic data using prior training data from other modalities (fluorescence, histological stains, or immunohistochemistry). Thus, this approach can expand training data by mapping to a common modality, giving a nearly general solution to segmentation across 2D imaging modalities (Hollandi et al., 2020).

The on-going search for generalizable segmentation is an area of active research in deep learning and is critical to establishing rigorous and reproducible segmentation approaches. To this end, a pipeline that requires little to no tuning on multiple datasets and modalities was demonstrated recently (Waibel et al., 2021). In the interim, the field will continue to make progress with generalizable segmentation, existing approaches, networks, etc. can provide the foundation for novel segmentation solutions. For instance, deep learning approaches to address 2D and 3D cell segmentation are often based on existing networks (Haberl et al., 2018; Schmidt et al., 2018; Falk et al., 2019; Weigert et al., 2020; Minaee et al., 2021; Stringer et al., 2021), using training data augmentation (Moshkov et al., 2020), or transfer learning

¹<https://www.nucleaizer.org/>

TABLE 1 | End user accessibility of tools supporting machine and/or deep learning for bioimage analysis.

User	Application	Name	Support or demonstrated		Description	Software type	URL	References		
			Classical learning	Deep learning (DL)						
Novice	Datasets and repositories	Image data resource (IDR)	Not determined	Yes, ex. Idr0042	Tissue and cell images with cell based training datasets	Repository	https://idr.openmicroscopy.org	Williams et al., 2017		
		Broad bioimage benchmark collection	Yes	Yes	Cell images training datasets	Repository	https://bbbc.broadinstitute.org	Ljosa et al., 2012		
		Cell image library	Not determined	CDeep3M	Multimodal cell images, linked to CDeep3M for testing	Repository	http://cellimagelibrary.org/pages/datasets	NA		
		BioImageDbs	Yes	Yes	R package and repository for images	Bioconductor package	https://kumes.github.io/BioImageDbs/	Kume and Nishida, 2021		
		EMPIAR	Yes, ex. EMPIAR-10069	Yes, ex. EMPIAR-10592	Electron microscopy images	Repository	https://www.ebi.ac.uk/empiar/	Ludin et al., 2016		
		SciLifeLab	Not determined	Yes	Scientific data, images and figure	Repository	https://www.scilifelab.se/data/repository/	NA		
		BioImage Archive	Yes	Yes	Archive of IDR and EMPIAR	Repository	https://www.ebi.ac.uk/bioimage-archive/deepcell.org	Ellenberg et al., 2018		
		DeepCell Kiosk		Establishing a cellwise dataset	Tool for segmentation in the cloud	Web interface		Moen et al., 2019		
		Cellpose		Segmentation	Tool for segmentation in the cloud and python GUI	Web interface, application	https://github.com/MouseLand/cellpose	Stringer et al., 2021		
		NucleAIzer		Transfer learning	Tool for segmentation in the cloud	Web interface	www.nucleaizer.org	Hollandi et al., 2020		
Practitioner	Segmentation	CDeep3M		Electron microscopy segmentation	Multiple trained networks for distinct structures in EM images	Web interface, model zoo	https://cdeep3m.crbs.ucsd.edu/	Haberl et al., 2018		
		QuPath	Feature design for segmentation	Inference with StarDist	ML segmentation with GUI	Application, plugin	qupath.github.io	Bankhead et al., 2017		
		DeepImageJ		Inference in ImageJ with BioImage.IO	Tool for inference on the desktop	Plugin	deepimagej.github.io/deepimagej/	Gómez-de-Mariscal et al., 2021		
		Ilastik	Feature design for segmentation	Interfaces with BioImage.IO	Segmentation with GUI	Application, plugin	www.ilastik.org	Berg et al., 2019		
		CellProfiler and CellAnalyst	Feature design for classification	Unet Segmentation	Pipeline Based image processing tool with ML and DL support	Application	cellprofiler.org	Dao et al., 2016; McQuin et al., 2018		
		StarDist		Segmentation	Python and Java (ImageJ/FIJI) tool for segmentation	Plugin	https://github.com/stardist/stardist	Weigert et al., 2020		
		HistoricML2		Model for training and tools for inference	Framework for training and inference on imaging data	Web interface	https://historicml2.readthedocs.io/	Lee et al., 2021		
		CSBDeep		Image restoration, segmentation	FIJI plugins and python for image restoration and segmentation	Python, plugin	https://csbdeep.bioimagecomputing.com/	Schmidt et al., 2018; Weigert et al., 2020		
		Developer	Cellwise Classification Neighborhoods	CytoMAP	Feature design for neighborhoods		Cell classification and neighborhood analysis with GUI	Application	gitlab.com/gernerlab/cytomap/-/wikis/home	Stoltzfus et al., 2020
				Volumetric tissue exploration and analysis	Feature design for classification and segmentation		Cell segmentation, classification and neighborhood analysis with GUI	Plugin	https://vtea.wiki	Winfree et al., 2017
modelzoo.co				Models for many datatypes	Open source and pretrained networks	Web repository	modelzoo.co	NA		
InstantDL				Segmentation and classification	Broadly applicable segmentation and classification framework	Python, CoLab	https://github.com/marrlab/InstantDL/	Waibel et al., 2021		
BioImage.IO				Models for specifically for bioimaging	DL networks for the bioimaging community	Web repository	bioimage.io	NA		
Segmentation	Segmentation	ZeroCostDL4Mic		Training and inference with BioImage.IO	Tool for training and inference in the cloud	Cloud based, CoLab	github.com/HenriquesLab/ZeroCostDL4Mic	von Chamier et al., 2021		
		OpSeF		DL network training and inference	Python framework in Jupyter notebooks	Python	github.com/trasse/OpSeF-IV	Rasse et al., 2020		
		Weka	Extensive library of classifiers and tools		ML frame work for Java, Plugin for ImageJ	API, application, plugin	www.cs.waikato.ac.nz/ml/weka/	Hall et al., 2009		

(Zhuang et al., 2021). Thus, until there is a generalizable solution, new deep learning segmentation approaches can be developed quickly by building on existing work with focused training datasets specific to tissue, cell-type and imaging modality.

CLASSIFICATION

Using specific protein or structural markers is a common way to determine cell-types in cytometry approaches like flow and image cytometry. Image cytometry is complicated by defining which pixels are associated with which cells. While a nuclear stain can be used to identify the nucleus, membrane, and cytoplasmic markers may be inconsistent across cell-types, cell-states, and tissues. A common solution is to measure markers in pixels proximal to segmented nuclei. These pixels can be defined by using a limited cell-associated region-of-interest that wraps around an existing nuclear segmentation or by performing a tessellation with a Voronoi segmentation (Winfree et al., 2017; Goltsev et al., 2018; McQuin et al., 2018).

The mean-fluorescence intensity (or other intensity measure, mode, upper-quartile mean, etc.) of markers in cell-associated segmented regions is frequently used for classification. A common supervised approach is to perform a series of sequential selections or gates based on marker intensities like flow cytometry. This “gating strategy” can easily identify specific cell-types with a predefined cell-type hierarchy. Cell classification can be semi-automated with unsupervised or semi-supervised machine learning using classifiers and clustering approaches. Popular approaches include Bayesian and Random Forest classifiers and clustering with k-means or graph based community clustering like the Louvain algorithm (Hall et al., 2009; Dao et al., 2016; McQuin et al., 2018; Phillip et al., 2021; Solorzano et al., 2021). Importantly, analyzing highly multiplexed image datasets, more than twenty markers, with a supervised “gating” approach may prove intractable necessitating machine learning approaches (Levine et al., 2015; Goltsev et al., 2018; Neumann et al., 2021).

Deep learning has been broadly applied to classification of images (Gupta et al., 2019). One of the strengths of a deep learning classification approach, as with segmentation, is that it is possible to start with a pretrained network-potentially reducing training set sizes. For instance, in 2D image classification, a convolutional neural network (CNN) like ResNet-50 initially trained on natural images (e.g., animals, vehicles, plants, etc.) can be retrained with a new label structure and training data that might include, for instance, cell nuclei (Woloshuk et al., 2021). Some deep learning models can further simplify workflows, like regional-CNNs, performing both segmentation and classification (Caicedo et al., 2019).

One image dataset that presents an interesting challenge and unique opportunity in both segmentation and classification is multiplexed fluorescence *in situ* hybridization (FISH). These approaches can, through combinatorial labeling of fluorophores, generate images of nearly all putative transcripts (Coskun and Cai, 2016). Although a semantic and instance segmentation approach can be used to identify and classify cells using associated FISH probes (Littman et al., 2021), a

recent pixelwise-segmentation free approach has been proposed. This approach organizes the detected FISH-probes into spatial clusters using graphs from which signatures of cells and cell-types are determined (Shah et al., 2016; Andersson et al., 2020; Partel and Wählby, 2021).

MICROENVIRONMENTS AS NEIGHBORHOODS

The classification of microenvironments in tissues informs our understanding of the role of specific cells and structures in an underlying biology. This has led to the development of neighborhood analysis strategies that involve the segmentation of groups of cells or structures which are then classified with machine learning using neighborhood features such as cell-type census and location (Stoltzfus et al., 2020; Solorzano et al., 2021; Winfree et al., 2021). This process mirrors the segmentation and classification of single cells by protein and RNA markers where the types of cells or the distributions of cell types in neighborhoods are the markers used to classify the neighborhoods. The segmentation strategies for defining neighborhoods usually rely on either regular sampling of a tissue or cell centric approaches (e.g., distance from a cell or the *k*-nearest neighborhoods) (Jackson et al., 2020; Stoltzfus et al., 2020; Lake et al., 2021; Winfree et al., 2021). The impact of neighborhood size and defining it variably and locally (e.g., microenvironments may be different near arterioles vs. microvasculature) are under explored avenues in the analysis of cellular microenvironments in bioimaging datasets. Importantly, further development of neighborhoods analyses is critical as it has demonstrated mechanistic insight in human disease when used with highly multiplexed chemical and fluorescence imaging (Jackson et al., 2020; Schürch et al., 2020; Stoltzfus et al., 2021).

TECHNOLOGY AND TOOL ACCESSIBILITY

Minimizing the exclusivity of segmentation and classification advancements with the development of user accessible tools, is critical to the democratization of image analysis. In the above discussions of both segmentation and classification, most researchers and developers paid careful attention to providing tools for use by biomedical scientists. Example tools include web interfaces, stand-alone applications, or plugins for open-source image processing software (**Table 1**). These tools provide access to users that are novices in image analysis, day-to-day practitioners and super-users or developers across the three fundamental tasks of cell segmentation, cell classification and neighborhood analysis (**Table 1**). Furthermore, deep learning networks are available through online repositories such as github.com and modelzoo.co. An exciting development is the recent set of publications that have defined one-stop-shops for deep learning models and accessible tools for using and training existing deep learning networks (Iudin et al., 2016; Berg et al., 2019; Rasse et al., 2020; Gómez-de-Mariscal et al., 2021; von Chamier et al., 2021, p. 4). This includes the integration of segmentation tools

with online repositories of trained deep learning networks that can be easily downloaded and tested on cells and modalities of interest. With this added accessibility, there is a risk of misuse and possible abuse. However, the ease of reproducibility may outweigh this risk.

CONCLUSION

The bioimaging community has recognized for decades that image data is more than a picture. Mining imaging data collected in the biomedical sciences has blossomed in the past 20 years, pushed by advancements in multiplexed tissue labeling, image capture technologies, computational capacity, and machine learning. It will be exciting to see the next developments in image analysis with machine learning approaches. Perhaps we will witness: (1) a fully generalizable

multidimensional cell segmentation approach; (2) novel approaches to cell-segmentation independent of pixelwise classification (as with some FISH data), or (3) new models of neighborhoods to characterize cellular microenvironments and niches. Furthermore, with web-based repositories to share datasets and tools that are suitable for all levels of expertise, these and other developments will be accessible to both experts, practitioners, and researchers new to imaging and image analysis. The broad accessibility of image data and tools could facilitate the adoption of common and rigorous processes for meaningful biological insight from image datasets across fields of study for so much more than a picture.

AUTHOR CONTRIBUTIONS

SW conceived and outlined the manuscript.

REFERENCES

- Allam, M., Cai, S., and Coskun, A. F. (2020). Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics. *NPJ Precis. Oncol.* 4:11. doi: 10.1038/s41698-020-0114-1
- Andersson, A., Partel, G., Solorzano, L., and Wahlby, C. (2020). “Transcriptome-Supervised Classification of Tissue Morphology Using Deep Learning,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, (Piscataway: IEEE), 1630–1633. doi: 10.1109/ISBI45749.2020.9098361
- Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., and McArt, D. G. (2017). QuPath: open source software for digital pathology image analysis. *Sci. Rep.* 7:16878. doi: 10.1038/s41598-017-17204-5
- Bannon, D., Moen, E., Schwartz, M., Borba, E., Kudo, T., Greenwald, N., et al. (2021). DeepCell Kiosk: scaling deep learning-enabled cellular image analysis with Kubernetes. *Nat. Methods* 18, 43–45. doi: 10.1038/s41592-020-01023-0
- Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., et al. (2019). ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* 16, 1226–1232. doi: 10.1038/s41592-019-0582-9
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., et al. (2007). “KNIME: The Konstanz Information Miner,” in *Studies in Classification, Data Analysis, and Knowledge Organization*, eds C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (Berlin: Springer).
- Beucher, S., and Lantuejoul, C. (1979). “Use of Watersheds in Contour Detection,” in *International Workshop on Image Processing, Real-Time Edge and Motion Detection*, (Rennes: Centre de Morphologie Mathématique). doi: 10.1016/s0893-6080(99)00105-7
- Caicedo, J. C., Goodman, A., Karhohs, K. W., Cimini, B. A., Ackerman, J., Haghghi, M., et al. (2019). Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* 16, 1247–1253. doi: 10.1038/s41592-019-0612-7
- Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 15:20170387. doi: 10.1098/rsif.2017.0387
- Coskun, A. F., and Cai, L. (2016). Dense transcript profiling in single cells by image correlation decoding. *Nat. Methods* 13, 657–660. doi: 10.1038/nmeth.3895
- Dao, D., Fraser, A. N., Hung, J., Ljosa, V., Singh, S., and Carpenter, A. E. (2016). CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* 32, 3210–3212. doi: 10.1093/bioinformatics/btw390
- de Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., et al. (2012). Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* 9, 690–696. doi: 10.1038/nmeth.2075
- Dunn, K. W., Fu, C., Ho, D. J., Lee, S., Han, S., and Salama, P. (2019). DeepSynth: three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. *Sci. Rep.* 9:18295. doi: 10.1038/s41598-019-54244-5
- El-Achkar, T. M., Eadon, M. T., Menon, R., Lake, B. B., Sigdel, T. K., Alexandrov, T., et al. (2021). A multimodal and integrated approach to interrogate human kidney biopsies with rigor and reproducibility: guidelines from the Kidney Precision Medicine Project. *Physiol. Genomics* 53, 1–11. doi: 10.1152/physiolgenomics.00104.2020
- Ellenberg, J., Swedlow, J. R., Barlow, M., Cook, C. E., Sarkans, U., Patwardhan, A., et al. (2018). A call for public archives for biological image data. *Nat. Methods* 15, 849–854. doi: 10.1038/s41592-018-0195-8
- Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., and Marrakchi, Y. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16, 67–70. doi: 10.1038/s41592-018-0261-2
- Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhat, S., Hale, M., Vazquez, G., et al. (2018). Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 174, 968–981.e15. doi: 10.1016/j.cell.2018.07.010
- Gómez-de-Mariscal, E., García-López-de-Haro, C., Ouyang, W., Donati, L., Lundberg, E., Unser, M., et al. (2021). DeepImageJ: a user-friendly environment to run deep learning models in ImageJ. *Nat. Methods* 18, 1192–1195. doi: 10.1038/s41592-021-01262-9
- Govind, D., Becker, J. U., Miecznikowski, J., Rosenberg, A. Z., Dang, J., Tharoux, P. L., et al. (2021). PodoSighter: a Cloud-Based Tool for Label-Free Podocyte Detection in Kidney Whole-Slide Images. *J. Am. Soc. Nephrol.* 32, 2795–2813. doi: 10.1681/ASN.2021050630
- Gupta, A., Harrison, P. J., Wieslander, H., Pielawski, N., Kartasalo, K., Partel, G., et al. (2019). Deep Learning in Image Cytometry: a Review. *Cytometry A*. 95, 366–380. doi: 10.1002/cyto.a.23701
- Haberl, M. G., Churas, C., Tindall, L., Boassa, D., Phan, S., Bushong, E. A., et al. (2018). CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation. *Nat. Methods* 15, 677–680. doi: 10.1038/s41592-018-0106-z
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explor. Newsl.* 11, 10–18. doi: 10.1145/1656274.1656278
- Hollandi, R., Szkalitsy, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., et al. (2020). nucleAIzer: a Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. *Cell Syst.* 10, 453–458.e6. doi: 10.1016/j.cels.2020.04.003
- Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J., and Patwardhan, A. (2016). EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* 13, 387–388. doi: 10.1038/nmeth.3806
- Jackson, H. W., Fischer, J. R., Zanutelli, V. R. T., Ali, H. R., Mechera, R., Soysal, S. D., et al. (2020). The single-cell pathology landscape of breast cancer. *Nature* 578, 615–620. doi: 10.1038/s41586-019-1876-x
- Kume, S., and Nishida, K. (2021). BioImageDBs: Bio- and biomedical imaging dataset for machine learning and deep learning (for ExperimentHub). R package version 1.2.2. Available online at: <https://bioconductor.org/packages/release/data/experiment/html/BioImageDBs.html> (accessed January 21, 2022).
- Lake, B. B., Menon, R., Winfree, S., Qiwen, H., Ricardo Melo, F., Kian, K., et al. (2021). An Atlas of Healthy and Injured Cell States and

- Niches in the Human Kidney. *bioRxiv* [Preprint]. doi: 10.1101/2021.07.28.454201
- Lee, S., Amgad, M., Mobadersany, P., McCormick, M., Pollack, B. P., Elfandy, H., et al. (2021). Interactive Classification of Whole-Slide Imaging Data for Cancer Researchers. *Cancer Res.* 81, 1171–1177. doi: 10.1158/0008-5472.CAN-20-0668
- Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir el, A. D., and Tadmor, M. D. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* 162, 184–197.
- Littman, R., Hemminger, Z., Foreman, R., Arneson, D., Zhang, G., Gómez-Pinilla, F., et al. (2021). Joint cell segmentation and cell type annotation for spatial transcriptomics. *Mol. Syst. Biol.* 17:e10108. doi: 10.15252/msb.202010108
- Ljosa, V., Sokolnicki, K. L., and Carpenter, A. E. (2012). Annotated high-throughput microscopy image sets for validation. *Nat. Methods* 9, 637–637. doi: 10.1038/nmeth.2083
- Lutnick, B., Ginley, B., Govind, D., McGarry, S. D., LaViolette, P. S., Yacoub, R., et al. (2019). An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nat. Mach. Intell.* 1, 112–119. doi: 10.1038/s42256-019-0018-3
- McQuin, C., Goodman, A., Chernyshev, V., Kamensky, L., Cimini, B. A., Karhohs, K. W., et al. (2018). CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol.* 16:e2005970. doi: 10.1371/journal.pbio.2005970
- Meijering, E. (2012). Cell Segmentation: 50 Years Down the Road [Life Sciences]. *IEEE Signal Process. Mag.* 29, 140–145. doi: 10.1109/MSP.2012.2204190
- Meijering, E. (2020). A bird's-eye view of deep learning in bioimage analysis. *Comput. Struct. Biotechnol. J.* 18, 2312–2325. doi: 10.1016/j.csbj.2020.08.003
- Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A., and Olivo-Marin, J. C. (2016). Imagining the future of bioimage analysis. *Nat. Biotechnol.* 34, 1250–1255. doi: 10.1038/nbt.3722
- Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., and Terzopoulos, D. (2021). Image Segmentation Using Deep Learning: a Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–1. [Online ahead of print] doi: 10.1109/TPAMI.2021.3059968
- Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. *Nat. Methods* 16, 1233–1246. doi: 10.1038/s41592-019-0403-1
- Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R., and Horvath, P. (2020). Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Sci. Rep.* 10:5068. doi: 10.1038/s41598-020-61808-3
- Neumann, E. K., Patterson, N. H., Allen, J. L., Migas, L. G., Yang, H., Brewer, M., et al. (2021). Protocol for multimodal analysis of human kidney tissue by imaging mass spectrometry and CODEX multiplexed immunofluorescence. *STAR Protoc.* 2:100747. doi: 10.1016/j.xpro.2021.100747
- Otsu, N. A. (1979). Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* 9, 62–66. doi: 10.1109/TSMC.1979.4310076
- Partel, G., and Wählby, C. (2021). Spage2vec: unsupervised representation of localized spatial gene expression signatures. *FEBS J.* 288, 1859–1870. doi: 10.1111/febs.15572
- Phillip, J. M., Han, K. S., Chen, W. C., Wirtz, D., and Wu, P. H. (2021). A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. *Nat. Protoc.* 16, 754–774. doi: 10.1038/s41596-020-00432-x
- Rasse, T. M., Hollandi, R., and Horvath, P. (2020). OpSeF: open Source Python Framework for Collaborative Instance Segmentation of Bioimages. *Front. Bioeng. Biotechnol.* 8:558880. doi: 10.3389/fbioe.2020.558880
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019
- Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). “Cell Detection with Star-Convex Polygons,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Lecture Notes in Computer Science*, eds A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Berlin: Springer International Publishing), 265–273. doi: 10.1007/978-3-030-00934-2_30
- Schürch, C. M., Bhate, S. S., Barlow, G. L., Phillips, D. J., Noti, L., Zlobec, I., et al. (2020). Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* 182, 1341–1359.e19. doi: 10.1016/j.cell.2020.07.005
- Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 92, 342–357. doi: 10.1016/j.neuron.2016.10.001
- Solorzano, L., Wik, L., Olsson Bontell, T., Wang, Y., Klemm, A. H., Öfverstedt, J., et al. (2021). Machine learning for cell classification and neighborhood analysis in glioma tissue. *Cytometry A* 99, 1176–1186. doi: 10.1002/cyto.a.24467
- Stoltzfus, C. R., Filipek, J., Gern, B. H., Olin, B. E., Leal, J. M., Wu, Y., et al. (2020). CytoMAP: a Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues. *Cell Rep.* 31:107523. doi: 10.1016/j.celrep.2020.107523
- Stoltzfus, C. R., Sivakumar, R., Kunz, L., Olin Pope, B. E., Menietti, E., Speziale, D., et al. (2021). Multi-Parameter Quantitative Imaging of Tumor Microenvironments Reveals Perivascular Immune Niches Associated With Anti-Tumor Immunity. *Front. Immunol.* 12:726492. doi: 10.3389/fimmu.2021.726492
- Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* 18, 100–106. doi: 10.1038/s41592-020-01018-x
- Sullivan, D. P., Winsnes, C. F., Åkesson, L., Hjelmare, M., Wiking, M., Schutten, R., et al. (2018). Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* 36, 820–828. doi: 10.1038/nbt.4225
- von Chamier, L., Laine, R. F., Jukkala, J., Spahn, C., Krentzel, D., Nehme, E., et al. (2021). Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* 12:2276. doi: 10.1038/s41467-021-22518-0
- Waibel, D. J. E., Shetab Boushehri, S., and Marr, C. (2021). InstantDL - An easy-to-use deep learning pipeline for image segmentation and classification. *BMC Bioinformatics* 22:103. doi: 10.1186/s12859-021-04037-3
- Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020). “Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy,” in *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, (Snowmass: IEEE). doi: 10.1109/WACV45572.2020.9093435
- Williams, E., Moore, J., Li, S. W., Rustici, G., Tarkowska, A., Chessel, A., et al. (2017). Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* 14, 775–781. doi: 10.1038/nmeth.4326
- Winfree, S., Al Hasan, M., and El-Achkar, T. M. (2021). Profiling immune cells in the kidney using tissue cytometry and machine learning. *Kidney360* 10.34067/KID.0006802020. [Online ahead of Print] doi: 10.34067/KID.0006802020
- Winfree, S., Khan, S., Micanovic, R., Eadon, M. T., Kelly, K. J., Sutton, T. A., et al. (2017). Quantitative Three-Dimensional Tissue Cytometry to Study Kidney Tissue and Resident Immune Cells. *J. Am. Soc. Nephrol.* 28, 2108–2118. doi: 10.1681/ASN.2016091027
- Woloshuk, A., Khochare, S., Almulhim, A. F., McNutt, A. T., Dean, D., Barwinska, D., et al. (2021). In Situ Classification of Cell Types in Human Kidney Tissue Using 3D Nuclear Staining. *Cytometry A* 99, 707–721. doi: 10.1002/cyto.a.24274
- Wu, L., Han, S., Chen, A., Salama, P., Dunn, K. W., and Delp, E. J. (2021). RCNN-SliceNet: a Slice and Cluster Approach for Nuclei Centroid Detection in Three-Dimensional Fluorescence Microscopy Images. *ArXiv* [Preprint]. Available online at: <http://arxiv.org/abs/2106.15753> (Accessed October 15, 2021).
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2021). A Comprehensive Survey on Transfer Learning. *Proc. IEEE* 109, 43–76. doi: 10.1109/JPROC.2020.3004555

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Winfree. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.