Göttingen, September 14, 2021

Sartorius Unveils LIVECell, a Deep-Learning Dataset for Label-Free, Quantitative Cell Segmentation in *Nature Methods*

- Open-source, manually annotated, expert-validated dataset includes more than 5000 label-free phase-contrast microscopy images acquired by the Incucyte® Live-Cell Analysis System, consisting of more than 1.6 million cells of eight cell types with distinct morphologies
- Diversity of the dataset enables its use for more effective training of neural networks for creation of customized machine learning algorithms, leading to more accurate segmentation of additional cell types

The Life Science Group Sartorius today announced publication of an article in Nature Methods describing the company's LIVECell (Label-free In Vitro image Examples of Cells) deep-learning dataset for label-free, quantitative segmentation of live cell images. The open-source dataset includes 5000 label-free phase contrast microscopy images consisting of more than 1.6 million cells of eight cell types with distinct morphologies that have been manually annotated. The set of images includes cells grown from initial seeding densities to fully confluent monolayers, resulting in a large variation in cell size and shape.

"The ability to derive physiologically relevant data from label-free microscopy images is a cornerstone of pharmaceutical research and datasets containing images of millions of cells facilitate exploration of biological phenomena with great statistical power," said Rickard Sjögren, PhD, Senior Scientist, Sartorius Corporate Research. "To compensate for a lack of image resolution, however, sophisticated imaging processing pipelines are necessary to generate the accurate cell-by-cell, pixel-by-pixel segmentations necessary to capture subtle changes in cell size, shape and texture, particularly if the goal is to investigate events at the level of cellular subpopulations or individual cells."

While neural networks can learn and adapt to identify and segment a variety of cells, they first require training with high quality datasets representative of the breadth of the cell morphologies to be encountered. Achieving accurate segmentation in microscopy images is essential for quantitative downstream analysis but is a challenging task. Traditional image analysis methods often require tedious algorithm customization and rigorous tuning of parameters specific to the cell morphology of interest.

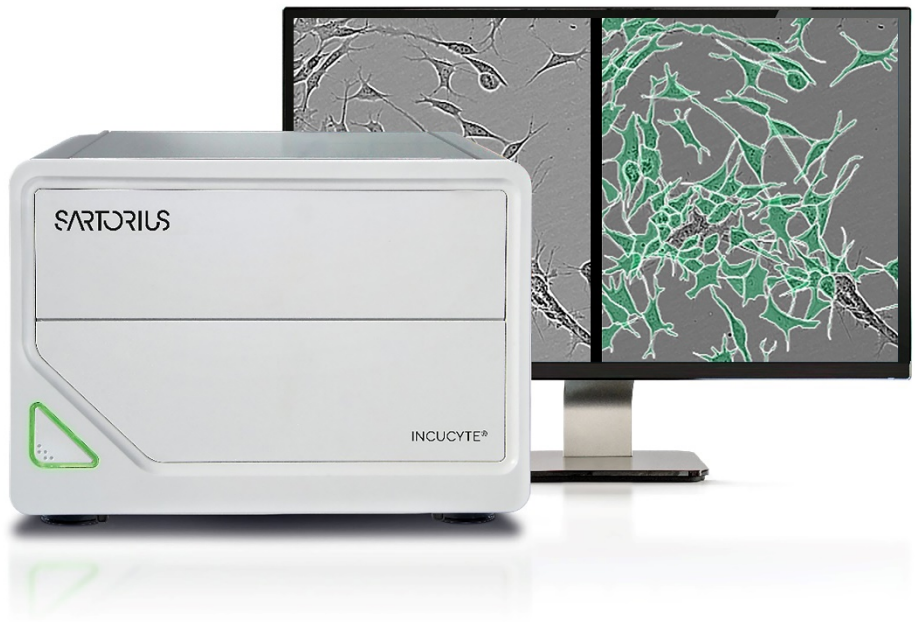
"The diversity of cell types and confluence conditions captured and annotated in the LIVECell dataset overcomes these challenges by facilitating the training of deep learning-based segmentation models," said Tim Jackson, PhD , Senior Image Processing Engineer, Sartorius BioAnalytics Product Development. "Researchers now have an unprecedented, high-quality label-free segmentation resource and starting point for training neural networks. Due to the nature of neural network-based algorithms being orders of

magnitude more complex than traditional image analysis, this data set will allow for more robust segmentation of various cell morphologies, and ultimately minimize user-introduced biases."

Prior to launch of the LIVECell dataset, the largest dataset of label-free images available to researchers consisted of 4,600 images derived from 26,000 cells.

Sartorius collaborated with the German Research Center for Artificial Intelligence (DFKI) to demonstrate the utility of this dataset and plans to continue work with the Center to further advancements in deep learning for the life sciences community.

Images of the eight different cell lines (human breast cancer (3), human glioblastoma, human hepatocyte carcinoma, human neuroblastoma, human ovarian cancer, mouse microglia) were captured every four hours, over the course of three to five days using an Incucyte® Live-Cell Analysis system. The high-throughput Incucyte® system was essential for building the image dataset as it allowed capture of a very high volume of high-quality images. Use of a high-throughput label-free culture system eliminated the risk of biological artifacts, while leads to increased confidence in the output of algorithms based on the dataset.



For further information on the LIVECell data set, please visit: https://sartorius-research.github.io/LIVECell/

For more information on the Incucyte® system, please visit: www.sartorius.com/incucyte

A profile of Sartorius

The Sartorius Group is a leading international partner of life science research and the biopharmaceutical industry. With innovative laboratory instruments and consumables, the Group's Lab Products & Services Division concentrates on serving the needs of laboratories performing research and quality control at pharma and biopharma companies and those of academic research institutes. The Bioprocess Solutions Division with its broad product portfolio focusing on single-use solutions helps customers to manufacture biotech medications and vaccines safely and efficiently. The Group has been annually growing by double digits on average and has been regularly expanding its portfolio by acquisitions complementary technologies. In fiscal 2020, the company earned sales revenue of some 2.34 billion euros. At the end of 2020, nearly 11,000 people were employed at the Group's approximately 60 manufacturing and sales sites, serving customers around the globe.

Contact

Rickard Sjögren, PhD
Senior Scientist, Sartorius Corporate Research
+46736822939
Rickard.Sjoegren@Sartorius.com
Sartorius.com

Follow Sartorius on Twitter **@Sartorius_Group** and on LinkedIn.