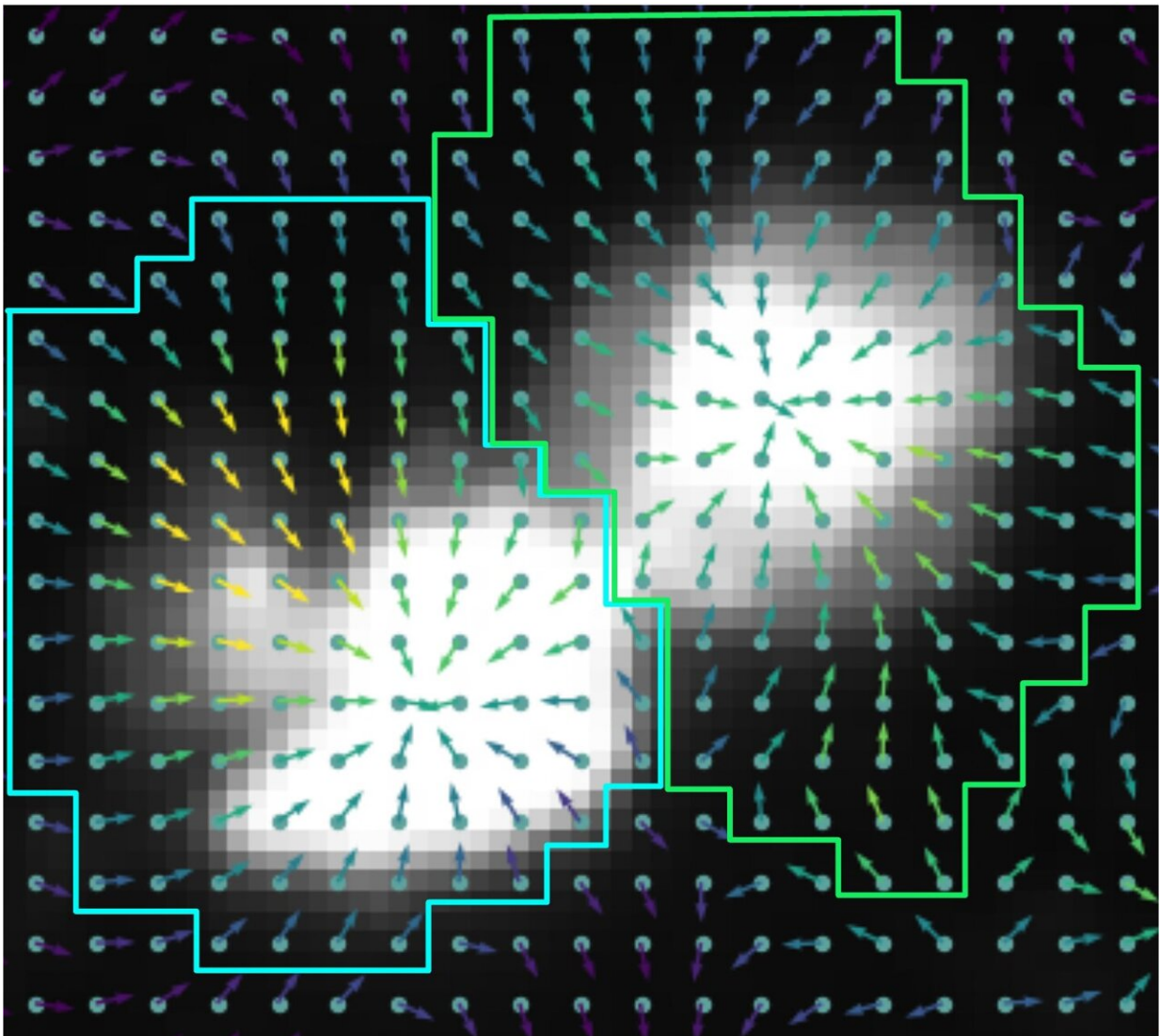# Newly-developed AI method uses transformer models to study human cells

August 15 2023, by Adam Kohlhaas

A new algorithm developed by researchers in Carnegie Mellon University's Computational Biology Department called subcellular spatial transcriptomics cell

segmentation (SCS) harnesses AI and advanced deep neural networks to adaptively identify cells and their constituent parts. Credit: Carnegie Mellon University

Researchers in Carnegie Mellon University's School of Computer Science have developed a method that uses artificial intelligence to augment how cells are studied and could help scientists better understand and eventually treat disease.

Images of organ or tissue samples contain millions of cells. And while analyzing these cells in situ is an important part of biological research, such images make it nearly impossible to identify individual cells, determine their function and understand their organization. A technique called spatial transcriptomics brings these cells into focus by combining imaging with the ability to quantify the level of genes in each cell—giving researchers the ability to study in detail several key biological mechanisms, ranging from how immune cells fight cancer to the cellular impact of drugs and aging.

Many current spatial transcriptomics platforms still lack the resolution required for closer, more detailed analysis. These technologies often group cells in clusters that range from several to 50 cells for each measurement, a resolution that may be sufficient for well-represented large cells but that is problematic for small cells or ones that aren't well represented. These rare cells may be the most critical for the disease or condition being studied.

In a new paper published in *Nature Methods*, Computational Biology Department researchers Hao Chen, Dongshunyi Li and Ziv Bar-Joseph have unveiled a method that uses artificial intelligence to augment the latest spatial transcriptomics technologies.

The CMU research focuses on more recent technologies that produce images at a much closer scale, allowing for subcellular resolution (or multiple measurements per cell). While these techniques solve the resolution issue, they present new challenges because the resulting images are so close-up that rather than capturing 15 to 50 cells per image, they capture only a few genes. This reversal of the previous problem creates difficulties in identifying the individual components and determining how to group these measurements to learn about specific cells. It also obscures the big picture.

The algorithm developed by the CBD researchers, called subcellular spatial transcriptomics cell segmentation (SCS), harnesses AI and advanced deep neural networks to adaptively identify cells and their constituent parts. SCS uses transformer models, similar to those used by large language models like ChatGPT, to gather information from the area surrounding each measurement. Just as ChatGPT uses the entire context of a sentence or paragraph for word completion, the SCS method fills in missing information for a specific measurement by incorporating information from the cells around it.

When applied to images of brain and liver samples with hundreds of thousands of cells, SCS accurately identified the exact location and type of each cell. SCS also identified several cells missed by current analysis approaches, such as rare and small cells that may play a crucial role in specific diseases or processes, including aging. SCS also provided information on location of molecules within cells, greatly improving the resolution at which researchers can study cellular organization.

"The ability to use the most recent advances in AI to aid the study of the human body opens the door to several downstream applications of spatial transcriptomics to improve human health," said Ziv Bar-Joseph, the FORE Systems Professor of Machine Learning and Computational Biology at CMU. Such downstream applications are already being

investigated by several large consortiums, including the Human BioMolecular Atlas Program (HuBMAP), that are using spatial transcriptomics to create a detailed, 3D map of the human body.

"By integrating state-of the-art biotechnology and AI, SCS helps unlock several open questions about cellular organization that are key to our ability to understand, and ultimately treat, disease," added Hao Chen, a Lane Postdoctoral Fellow in CBD.

SCS is available free on GitHub.

  More information: Hao Chen et al, SCS: cell segmentation for high-resolution spatial transcriptomics, *Nature Methods* (2023). [DOI: 10.1038/s41592-023-01939-3](DOI)

Provided by Carnegie Mellon University