Real Time, GPU-Accelerated Analysis and Visualization in the Life Sciences
Ken Kennedy Institute Virtual Data Science Conference, Rice University
October 26 - 27

Michelle Gill, Ph.D.
Senior Scientist, Deep Learning and Proteomics, NVIDIA

Avantika Lal, Ph.D.
Senior Scientist, Deep Learning and Genomics, NVIDIA

Abstract:

The drug discovery process is being transformed by the increased availability of datasets, with multi-omics data accelerating target identification and validation, and large chemical databases improving lead discovery and optimization times. However, the growing volume of data can be a bottleneck for analysis, as is the increased resolution and complexity of these datasets. DNA and RNA sequencing experiments are now routinely performed at the scale of individual cells, with hundreds of thousands of features measured in every cell of a biological sample; and analogous techniques in proteomics are also increasing in popularity. Virtual screening, which is commonly used for lead discovery, now utilizes features derived from libraries containing billions of compounds.

To extract scientific insights from such datasets, we need methods that take advantage of the complexity of the data, and can scale efficiently to massive datasets. The increased degree of parallelism afforded by GPUs has made them an ideal foundation for the acceleration of high performance computing, including those in data science. Furthermore, GPU-accelerated data science can be combined with deep learning to develop analysis pipelines that are both faster and more accurate than the existing state of the art. The NVIDIA omics team develops tools that use GPUs to both accelerate and improve the analysis of omics and cheminformatics data.

We show how GPUs can accelerate the analysis of data from single-cell DNA and RNA sequencing experiments, as well as features derived from libraries of small molecules, enabling analyses that previously required minutes or hours to be completed in seconds. This magnitude of acceleration makes the prospect of interactive, real-time analysis a reality. Finally, we demonstrate unique GPU-powered interactive visualization tools which enable scientists to interactively explore, analyze and discover novel biological and chemical insights from massive datasets.