Other - Colloquium on Artificial Intelligence Research and Optimization

AI Research and Optimization at Wafer-Scale

Andy Hock, Cerebras Systems, Inc

VP of Product

Virtual- REGISTRATION REQUIRED (SEE ABSTRACT) Zoom
April 07, 2021 - 01:00 pm

Abstract:

Deep learning and artificial intelligence (AI) has great potential for a wide variety of scientific and industry applications. However, modern AI research is constrained by compute: state of the art deep learning models often take days or weeks to train even on large clusters of legacy, general purpose processors like graphics processing units (GPUs). These machines are suitable, but not optimal for AI work. We need a new compute solution to accelerate time to solution, reduce the cost of curiosity, unlock new research and applications.

Cerebras has developed a new processor and system that is able to accelerate AI compute by orders of magnitude beyond GPU, training models in minutes or hours that previously took days or weeks. At the heart of our system is the Wafer-Scale Engine (WSE) – the largest chip ever built and the most powerful processor available for deep learning. The WSE is massive, more than 56x larger than the largest chip built previously. With 400,000 cores, 18GB of fast on-chip SRAM, and a high bandwidth, low latency, software-configurable on-chip network, the WSE delivers cluster-scale compute within a single device programmable via familiar ML frameworks such as TensorFlow. Housed in a standard datacenter-compatible server called the CS-1, our novel solution enables AI research at previously-impossible speeds and scale.

In this talk we will provide an introduction to the Cerebras WSE processor and CS-1 system, and discuss its implications for AI research and optimization.

REGISTRATION IS FREE AND REQUIRED TO RECIEVE ZOOM ID:
https://docs.google.com/forms/d/e/1FAIpQLSdLcIBvX8tiTkxm9HbvBdyPx70fhpKpzsAQmCf0AXgn44cqcQ/viewform

---

Speaker's Bio:

Dr. Andy Hock is VP of Product at Cerebras. Andy came to Cerebras from Google, where he led Data and Analytics Product for the Terra Bella project, using deep learning and AI to create useful data for maps and enterprise applications from satellite imagery. Computation speed was a problem for this work, and at Cerebras, Andy saw an opportunity to help deliver the right compute solution for deep learning and AI. Before Google, Andy was a Senior Scientist and Senior Technical Program Manager at Arete Associates, where he led research for image processing algorithm development. He has a PhD in Geophysics and Space Physics from UCLA and a BA in Astronomy-Physics from Colgate University.

Center for Computation & Technology
2003 Digital Media Center • Telephone: +1 225/578-5890 • Fax: +1 225/578-8957
© 2001–2025 Center for Computation & Technology • Official Web Page of Louisiana State University.