IT Peer Network

THREE APPROACHES TO HPC AND AI CONVERGENCE

Written by IT Peer Network | September 11, 2018

Artificial Intelligence (AI) is by no means a new concept. The idea has been around since Alan Turing's publication of "Computing Machinery and Intelligence" [https://www.csee.umbc.edu/courses/471/papers/turing.pdf] in the 1950s. But until recently, the computing power and the massive data sets needed to meaningfully run AI applications weren't

IT Peer Network

performance they need to run AI applications that further drive their mission needs.

Many organizations that already rely on a high-performance computing (HPC) infrastructure to support applications like modeling and simulation are now looking for ways to benefit from AI capabilities. Given that AI and HPC both require strong compute and performance capabilities, existing HPC users who already have HPC-optimized hardware are well placed to start taking advantage of AI. They also have an opportunity to gain efficiency and cost benefits by converging the two applications on one infrastructure.

APPROACH 1: USING EXISTING HPC INFRASTRUCTURE TO RUN EXISTING AI APPLICATIONS

This usually involves running AI applications developed on infrastructure-optimized AI frameworks, such as TensorFlow*, Caffe* and MXNet* on an HPC system. Companies looking to add AI capabilities to an existing HPC system based on Intel® Xeon® processors should ensure they use the latest optimized framework [https://ai.intel.com/framework-optimizations/] that best supports their planned use case [https://ai.intel.com/] .

An example of this type of use case can be seen in a recent collaboration project between Intel and Novartis [https://newsroom.intel.com/news/using-deep-neural-network-acceleration-image-analysis-drug-discovery/] , which used deep neural networks (DNN) to accelerate high content screening capabilities within image analysis. High content screening is fundamental in early drug discovery as it enables the analysis of microscopic images to see the effects of thousands of genetic or chemical treatments on different cell cultures. This is done through classical image-processing techniques to extract information on thousands of pre-defined features such as size, shape and texture. Applying deep learning to this process means the system is automatically learning the features that can distinguish one treatment from another.

By applying DNN acceleration techniques to process multiple images, the team cut the time to train image analysis models from 11 hours to 31 minutes – an improvement of greater than 20 times[1] [#_edn1] . This was done using a typical HPC infrastructure—eight CPU-based servers and

IT Peer Network

and to make full use of the server platform's large memory support. As a result, they were able to scale more than 120 3.9-megapixel images per second with 32 TensorFlow workers.

APPROACH 2: ADDING AI TO THE MODELING AND SIMULATION WORKFLOW TO ACCELERATE INNOVATION AND DISCOVERY

Organizations already using HPC to run modeling and simulation can introduce AI to their existing workflows to gain insights from their results faster. While existing visualization techniques enable scientists to derive insights from simulation results, some of this process can be automated using a continuous workflow that runs a simulation and modeling HPC workload and then feeds the data it creates into an AI workflow for improved insight.

Here is an example of how Princeton University Neuroscience Institute [https://iq.intel.com/mapping-mind-artificial-intelligence/] used a similar approach with HPC, machine learning (ML) and AI to analyze data coming from functional magnetic resonance imaging (fMRI) scans to determine what's going on inside the brain. The study involved using an ML system that has been being trained on real-life scans to create a model of the brain that would be able to recognize different cognitive processes.

he model was then used to look at real-time fMRI brain images of patients reacting to conflicting stimuli to 'guess' which cognitive processes were going on (and which stimuli were being paid more attention). This information was then used for immediate feedback by updating the stimuli presented. This ability to quickly analyze fMRI data using HPC and react using ML and AI systems is helping scientists better understand cognitive processes with a view to eventually improving the diagnosis and treatment of psychiatric disorders.

APPROACH 3: COMBINING HPC AND AI MODALITIES

A more ambitious approach is to embed HPC simulations into AI, where AI uses simulation to augment training data or provide supervised labels for generally unlabeled data. Alternatively, AI

IT Peer Network

In the field of astronomy—typically a heavy user of HPC—numerous new use cases have emerged for accelerating space research by combining HPC and AI modalities. One use case involves using AI to study gravitational lenses [https://www.theverge.com/2017/11/15/16654352/ai-astronomy-space-exploration-data] , a rare phenomenon that happens when a massive object like a galaxy or black hole comes between a light source and an observer on earth, bending the light and space around it. This means astronomers can see more distant (and much older) parts of the universe that they wouldn't usually be able to see.

Gravitational lenses are hard to find and traditionally have been identified by manually processing space images. In 2017 researchers from the universities of Bonn, Naples, and Groningen [https://arxiv.org/abs/1702.07675] used a Convolutional Neural Network (CNN) to accelerate detection. They started by creating a dataset to train the neural network by feeding six million images of fake gravitational lenses to the AI network, and leaving it to identify patterns. After this training, the AI system was set loose on real images from space, analyzing them to identify gravitational lenses at greater speed than human examination and with incredibly high rates of accuracy.

Another recent use case demonstrated that AI-based models can potentially replace computationally expensive tasks in simulation. In this example, Intel collaborated with High-Energy Physics (HEP) scientists to study what happens during particle collisions. The study used a huge number of CPUs to power its most complex and time-consuming simulation tasks. This included processing information from high-granularity calorimeters—the apparatus that measure particle energy. The team aimed to accelerate their ability to study collision data from these devices in preparation for greater data volumes coming from future collisions.

The team wanted to see if Generative Adversarial Networks (GANs) trained on the calorimeter images could act as a replacement for the computationally expensive Monte Carlo methods currently used to analyze them. GANs were seen as a suitable AI application as they are excellent at generating new variations based on the data studied. GANs were used to generate realistic

IT Peer Network

After training the GAN, the team found strong agreement between the images it generated and those produced by the simulation-based Monte Carlo approach. They reviewed both high-level qualities like energy shower shapes, and detailed calorimeter responses at a single-cell level and found that the agreement was incredibly high. This opens a promising avenue for further investigation for machine-learning-generated distributions in place of costly physics-based simulations.

GETTING STARTED WITH AI APPLICATIONS

When taking your first steps towards converged AI and HPC, it is important to understand different AI capabilities and how they can help solve the particular problems your organization is working on. The next step is to find AI frameworks that support your use case. During framework selection, it is best to look for ones that are already optimized for your current HPC infrastructure. For companies wanting to run AI on existing Intel® technology-based infrastructure we've created this overview [https://ai.intel.com/framework-optimizations/] of resources optimized for popular AI frameworks.

The next step is to run an AI workload pilot on your existing HPC infrastructure. At Intel, we work with customers across academia, government and enterprise to help them scope, plan and implement AI capabilities into their HPC environments. To find out more about how to optimize HPC architectures for AI convergence read this solution brief [https://www.intel.com/content/www/us/en/analytics/high-performance-computing/optimize-hpc-architectures-for-ai-convergence.html] .

For organizations wanting to optimize their existing infrastructure for specific workloads such as professional visualization or simulation and modeling, Intel® Select Solutions for HPC [https://www.intel.com/content/www/us/en/products/solutions/select-solutions/hpc.html] offer easy and quick-to-deploy infrastructure. Optimized for specific HPC applications, Intel® Select Solutions help to accelerate time to breakthrough, actionable insight, and new product design.

IT Peer Network

Omni-Path Switch. Single node configuration: CPU: Intel® Xeon® Phi Processor 7290F, 192GB DDR4 RAM, 1x 1.6TB Intel® SSD DC S3610 Series SC2BX016T4, 1x 480GB Intel® SSD DC S3520 Series SC2BB480G7, Intel® MKL 2017/DAAL/Intel Caffe.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks [http://www.intel.com/benchmarks]

Intel® technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com [https://www.intel.com] .

🗓 September 11, 2018 [https://itpeernetwork.intel.com/three-approaches-hpc-ai-convergence/]   👤 IT Peer

Network   📂 Artificial Intelligence   🏷 Artificial Intelligence, Convergence, HPC, Intel Select Solutions