MAGMA 2.6, Future Research in AI

Cade Brown <cade@utk.edu>

ICL INNOVATIVE COMPUTING LABORATORY

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Time @ ICL

2 years ago (2019-08), I joined ICL as a Research Assistant to Stan Tomov. Since then, I've primarily worked on MAGMA

This year, I've continued my work, and expanded as well:

- Merged HIP and CUDA support, as well as new/old cuSPARSE functionality
- Release: MAGMA 2.6 (no more 'hipMAGMA', it is all MAGMA proper now), and patches resolving issues found thus far
- Exploring new research areas in ML/AI, including visualization techniques

MAGMA

With MAGMA 2.6, researchers are installing and using MAGMA, and reporting issues. MAGMA 2.6.1 was released as a result

Not much has changed in our methodology to port MAGMA (it has worked well), but as AMD fixes bugs, we revert our hot-fixes

Performance for most dense routines has continually improved. However, some sparse functionality has degraded, which is something we're going to continue working on

AI/ML Research

Since the release of MAGMA 2.6, I have been researching AI/ML methods, to see where there is potential for new methods, papers, or software.

There are now huge models, with billions of parameters (for example, ViT-G/14). It is difficult to even analyze a network of such size, and a simple measure like loss/accuracy is not helpful enough to guide design changes.

Network Visualization (Dreaming)

Deep Dreaming, as it is commonly known, is a visualization technique for neural networks, which amplifies certain features in an image (according to a pre-trained network)

The process is similar to training network weights, except that the image is ‘trained’ instead of the weights

It can visually demonstrate patterns that a network sees, which can be extremely useful for debugging and optimizing network architectures

Example: DenseNet201

DenseNet201 is a classification network designed to classify any image to 1 of 1000 classes. It simply takes an image, and has been trained to output the probability that it is each of the classes.

Using deep dreaming, we can use that network to explain to us the patterns it sees, visually, by modifying the image to maximize those activations (via gradient ***ascent***)

https://www.mathworks.com/help/deeplearning/ref/densenet201.html

Beyond Images

Currently, the only implementations of ‘dreaming’ are on images. But, the basic idea can be applied to audio, sensory data, and spatial models.

For example, models that detect disease off of a medical scan could try to modify the readings and calculate the difference to see where the problem might lie. However, since there could be many ‘solutions’, this method may not find the ‘true’ cause

Summary

- HPC is requiring software to support many hardware configurations
  - AMD GPUs, Intel GPUs, …
- MAGMA has been successfully ported and released to support HIP
- Machine learning will continue to grow, especially in HPC
- ML/AI models are getting bigger, too big to debug easily
  - An opportunity for new software/methods

Questions?