

Accelerating Deep Learning for visual understanding under a Framework CAFFE by OpenCL



Background

Visual understanding is a broad term used to describe efforts to extract meaning and knowledge from images computationally. Much effort has been spent by researchers to develop methods to allow for information to be gained from learning skills humans do. However, it has proven to be a very difficult problem.

In recent years some major advances in the field are seeing vision systems come about that are performing well at general recognition in images for large sets of objects. All of these new systems have been based around the concepts of convolutional neural networks (CNN) and deep learning. Especially since the coming out of an open source software framework Caffe (see here: <http://caffe.berkeleyvision.org/>), designing and testing CNN networks become much easier and faster. During its first year, Caffe has been forked over 1000 times. In addition it has been used for many of the ImageNet competition entries. Caffe currently supports both CPU and GPU for forward/backward passes, as well as options for using several different CPU BLAS libraries and CUDA/cuDNN based GPU implementations.

Motivation and Goals

Much of the research and software development focus related to CNNs has been focused so far on developing better and faster methods for training networks. Because of this most software framework for CNN are targeted only at high end discrete GPUs. That focus however is beginning to expand, as networks are now available that can perform well on many vision tasks, it is now desirable to start creating applications which use CNN based vision systems. This presents new challenges however as the devices which are most likely to be used for classifying and detecting objects in a video stream are mobile power constrained platforms that typically do not have a large discrete GPU available.

We see this as a great opportunity to leverage smaller, more power efficient GPUs which are often found in mobile devices. One of the hurdles we identified to this use case however is that almost all efforts and software frameworks for CNNs thus far have focused solely on discrete GPUs and proprietary APIs for accessing them.

As a first target to explore the advantages of utilizing integrated GPUs to accelerate CNN processing we decided to target a very popular open source framework, Caffe, and create an OpenCL accelerated branch.

Tasks

- Enabled OpenCL GPU support in Caffe.
- Passed all existing unit tests in Caffe.

What you can benefit through this topic

You can benefit far more than your grade in this topic.

- You will become a skillful OpenCL programmer.
- You will become an expert in the current hot technique: machine learning.

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Echtzeitsysteme und Robotik

Supervisor:
Prof. Dr.-Ing Alois Knoll

Advisor:
Dr. Kai Huang
Biao Hu M.Sc.

Research project:
-

Type:
BA/MA

Research area:
Machine Learning and Parallel Programming

Programming language:
OpenCL, C++

Required skills:
Good at C++, Machine Learning

Language:
englisch

Date of submission:
15. Juli 2016

For more information please contact us:

Phone: +49.89.289.18134

E-Mail: huangk@tum.de
hub@in.tum.de

Internet: www6.in.tum.de