Abstract

Deep Neural Network Algorithms on Graphics Processors for Embedded Systems
by David Cronin
MAI Computer Engineering, 2018

Deep neural networks are becoming an increasingly popular method for tackling a range of problems, particularly those of recognition. A lot of focus is given to improving improving their performance by utilising graphics processors to perform key computations. However this focus is primarily on using computers with very high performance components. This thesis tackles the idea of performing these tasks on low powered embedded devices, often made with components that have comparatively very low performance. As such this thesis describes the implementation of an algorithm that performs a key computation associated with deep neural networks, specifically matrix multiplication.

The algorithm described can target graphics processors using OpenGL on a large range of devices, scaling from small low powered embedded devices with ARM architectures and integrated graphics processors to large high powered desktop workstations with Intel architectures and discrete graphics processors. The implementation is also portable and cross platform, supporting both macOS and Linux. This is achieved by repurposing OpenGL functionality, originally intended for graphics processing, to perform more general purpose computations.

This thesis demonstrates that it is possible to achieve improved execution times when performing matrix multiplication by efficiently using using OpenGL to target the graphics processor. This thesis also demonstrates as a proof on concept that this algorithm can compile and run on a small low powered embedded device for more limited computations. Further work is also described that could be done to achieve improved results. By being able to perform the work of deep neural networks on these embedded devices a whole range of new applications could become possible, as these devices would be more capable of interpreting their environments using image and audio processing.