

A tensorflow backend to SpiNNaker

More precise: a keras backend to SpiNNaker

Jonas Faßbender

Supervisors:

Alan Stokes, Kevin Stratford, Caoimhín Laoide-Kemp

What is SpiNNaker?

What is SpiNNaker?

The project's website [Spi20] states:

- ▶ [SpiNNaker is] a platform for high-performance massively parallel processing appropriate for the simulation of large-scale [spiking] neural networks in real-time, as a research tool for neuroscientists, computer scientists and roboticists

What is SpiNNaker?

The project's website [Spi20] states:

- ▶ [SpiNNaker is] a platform for high-performance massively parallel processing appropriate for the simulation of large-scale [spiking] neural networks in real-time, as a research tool for neuroscientists, computer scientists and roboticists
- ▶ [SpiNNaker is] an aid in the investigation of new computer architectures, which break the rules of conventional supercomputing, but which we hope will lead to fundamentally new and advantageous principles for energy-efficient massively-parallel computing

Why use SpiNNaker as a target for training deep NNs?

Why use SpiNNaker as a target for training deep NNs?

- ▶ I don't like to wait days for training my NN to finish (and it will probably still suck afterwards)

Why use SpiNNaker as a target for training deep NNs?

- ▶ I don't like to wait days for training my NN to finish (and it will probably still suck afterwards)
- ▶ Amount of computation in training deep NNs increases exponentially (double every 3.4 months) [Dar19]

Why use SpiNNaker as a target for training deep NNs?

- ▶ I don't like to wait days for training my NN to finish (and it will probably still suck afterwards)
- ▶ Amount of computation in training deep NNs increases exponentially (double every 3.4 months) [Dar19]
- ▶ We'll run out of available computation (and energy) eventually

Why use SpiNNaker as a target for training deep NNs?

- ▶ I don't like to wait days for training my NN to finish (and it will probably still suck afterwards)
- ▶ Amount of computation in training deep NNs increases exponentially (double every 3.4 months) [Dar19]
- ▶ We'll run out of available computation (and energy) eventually
- ▶ Massively parallel, energy efficient and scalable systems are optimal for training NNs

What is tensorflow? And why the subtitle

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]
- ▶ Represents computation as a graph

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]
- ▶ Represents computation as a graph
- ▶ Basically: abstraction over various hardware and software libraries and APIs (CPUs, GPUs, TPUs, Eigen, MPI, CUDA, OpenCL, ...)

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]
- ▶ Represents computation as a graph
- ▶ Basically: abstraction over various hardware and software libraries and APIs (CPUs, GPUs, TPUs, Eigen, MPI, CUDA, OpenCL, ...)
- ▶ But: no abstraction over any runtime (like SpiNNaker)

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]
- ▶ Represents computation as a graph
- ▶ Basically: abstraction over various hardware and software libraries and APIs (CPUs, GPUs, TPUs, Eigen, MPI, CUDA, OpenCL, ...)
- ▶ But: no abstraction over any runtime (like SpiNNaker)
- ▶ "Easiest" way to add new tensorflow backend? XLA (<https://www.tensorflow.org/xla>)

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]
- ▶ Represents computation as a graph
- ▶ Basically: abstraction over various hardware and software libraries and APIs (CPUs, GPUs, TPUs, Eigen, MPI, CUDA, OpenCL, ...)
- ▶ But: no abstraction over any runtime (like SpiNNaker)
- ▶ "Easiest" way to add new tensorflow backend? XLA (<https://www.tensorflow.org/xla>)
- ▶ Instead: backend for keras (<https://keras.io>)

What is tensorflow? And why the subtitle

- ▶ TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [Mar15]
- ▶ Represents computation as a graph
- ▶ Basically: abstraction over various hardware and software libraries and APIs (CPUs, GPUs, TPUs, Eigen, MPI, CUDA, OpenCL, ...)
- ▶ But: no abstraction over any runtime (like SpiNNaker)
- ▶ "Easiest" way to add new tensorflow backend? XLA (<https://www.tensorflow.org/xla>)
- ▶ Instead: backend for keras (<https://keras.io>)
- ▶ Using high level conceptual graph (the actual layers of the NN) instead of low level computational graph of tensorflow

Challenges

Challenges

- ▶ Writing scientific programs is hard

Challenges

- ▶ Writing scientific programs is hard
- ▶ Writing programs for embedded hardware is hard

Challenges

- ▶ Writing scientific programs is hard
- ▶ Writing programs for embedded hardware is hard
- ▶ Both together? ...

Overcoming them

Overcoming them

- ▶ Is there a linear algebra library for an embedded environment like SpiNNaker? (LAPACKE, CBLAS, ...?)

Overcoming them

- ▶ Is there a linear algebra library for an embedded environment like SpiNNaker? (LAPACKE, CBLAS, ...?)
- ▶ Don't sleep

References



Dario Amodei, Danny Hernandez, Girish Sastry, Jack Clark, Greg Brockman, Ilya Sutskever.

AI and Compute.

<https://openai.com/blog/ai-and-compute/>, 2019.



Martin Abadi et al.

Tensorflow: Large-Scale Machine Learning on Heterogenous Distributed Systems.

<http://download.tensorflow.org/paper/whitepaper2015.pdf>, 2015.



SpiNNaker Team.

SpiNNaker Project.

<http://apt.cs.manchester.ac.uk/projects/SpiNNaker/project/>, 2020.