Introducing WASI-NN

Andrew Brown, Mingqiu Sun

What is WASI?

Web Assembly System Interface

- Sub group of the WASM CG
- Official group to define WASI API
- https://github.com/WebAssembly/WASI
- Video meeting biweekly
- Witx is the interface definition language

What is Bytecode Alliance

- The Bytecode Alliance is an open source community dedicated to creating secure new software foundations, building on standards such as WebAssembly and WebAssembly System Interface (WASI).
- https://bytecodealliance.org/
- Founding members: Mozilla, Fastly, Intel and RedHat
- Intel Involvement:
  - WebAssembly Micro Runtime (WAMR)
  - SIMD implementation in Wasmtime
  - WASI-NN interface specification

Why WASI-NN?

- Trained machine learning models are often need to be deployed on a variety of devices with different architecture and operating systems.
- WASM provide an ideal portable form of deployment of those models.
- Initial focus is on inferencing, inspired by WebNN's model loader API
- Framework and model format agnostic
- Expect to support a variety of devices: CPU, GPU, FPGA, TPU, …

API Walkthrough

https://github.com/abrown/WASI-nn/blob/master/phases/ephemeral/witx/wasi_ephemeral_nn.witx

Call for action

Your feedback is highly appreciated, as the proposal is in its initial stage, changes are easier to incorporate!