Barracuda Introduction

Tracy Sharpe
Principal Software Engineer

What is Barracuda?

→ Lightweight cross-platform C# API for Unity framework

→ Runs on client devices ranging from phones, to game consoles, to PCs

→ Supports CPU or GPU inferencing

→ GPU inferencing targets compute shaders (HLSL, OpenGL, Metal, etc) and pixel shaders (notably WebGL)

→ Imports ONNX models

Unity ®

Where is Barracuda used?

→ Internally used by Unity APIs for adding intelligent behavior to objects

→ Unity Labs and other teams exploring neural rendering techniques (either at content creation time or during a runtime experience)

→ External customers use Barracuda as general inference engine primarily for convolutional image models

Unity®

Use case for Barracuda and WebML?

→ Unity authoring tools have been built as web pages using Unity WebGL player

→ WebGL performance is limiting for models running under Barracuda

→ WebGPU alternative backend would use more of our compute shader investments

→ WebML would offer native inference engine performance

→ Experimental at this time

Unity®