Web图形和机器学习的一些探索

Yang Gu
Manager of Intel Web Graphics and Machine Learning Team
Sep 6, 2022
Agenda

- Web图形
- Web机器学习
- 英特尔“Web图形和机器学习”团队
Web图形
- Legacy native APIs: OpenGL, OpenGL ES, D3D11
- Modern native APIs: Vulkan, D3D12, Metal
- WebGL is based on legacy native APIs, while WebGPU is based on modern ones
Web Graphics Architecture

Usage

Browser API

Native API

Hardware

CPU

GPU

WebGPU

WebGPU API

WebGPU Shader

(VGSL)

Vulkan/D3D12/Metal

SPIR-V/(HLSL/DXIL)/MSL

Hardware

CPU

SwiftShader

GPU

Native API

OpenGL

Microsoft

DirectX11

Vulkan

Microsoft

DirectX12

Intel

Browser API

WebGL

WebGPU

Ps

Autocad 3D

App

Zoom

Tencent

WebXR

Usage

WebGPU

WebGPU API

WebGPU Shader

(VGSL)

Vulkan/D3D12/Metal

SPIR-V/(HLSL/DXIL)/MSL

Hardware

CPU

SwiftShader

GPU

Native API

OpenGL

Microsoft

DirectX11

Vulkan

Microsoft

DirectX12

Intel

Browser API

WebGL

WebGPU

Ps

Autocad 3D

App

Zoom

Tencent

WebXR
# Major Diff between WebGL and WebGPU

<table>
<thead>
<tr>
<th>Aspect</th>
<th>WebGL</th>
<th>WebGPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native APIs Based on</td>
<td>Old</td>
<td>Modern (Explicit -&gt; performant and predictable)</td>
</tr>
<tr>
<td>CPU Multi-threading</td>
<td>NA</td>
<td>Supported</td>
</tr>
<tr>
<td>Validation</td>
<td>Heavy</td>
<td>Lightweight (validation layer)</td>
</tr>
<tr>
<td>Compute</td>
<td>NA</td>
<td>Supported</td>
</tr>
</tbody>
</table>
WebGPU Status

- **Spec**
  - Both API and WGSL specs are in good shape
  - Finished TAG review
  - Going through WG (Media, WebML, Immersive Web, WASM, etc.) reviews

- **Implementation**
  - Chrome implementation is almost ready
  - Other browsers (Firefox, Safari) are catching up

- **Early adopters**
  - Three.js, Babylon.js, Unity, Unreal, Coco2, Snap, Node.js, Deno, TensorFlow.js, etc.
Intel’s Efforts on WebGPU

- 2000+ patches in Spec, CTS and Chrome implementation, covering MSAA, Dynamic Buffer Offset, Extension Infrastructure, Coordinate System, Timestamp Query, Pipeline Creation, FP16, Video Uploading, WGSL-to-DXIL, DXC Support, DP4A and so on
- Maintain and improve conformance and performance on Intel platforms
- Attend regular Working Group sync meeting
- Host frequent sync meetings with Google and Microsoft
- Presented at SIGGRAPH and WebGL/WebGPU Meetup
Web 机器学习
Web Machine Learning Architecture

Use cases
- Image Classification
- Object Detection
- Background Segmentation
- Noise Suppression
- Natural Language

Frameworks
- TensorFlow.js / TFLite Web
- ONNX Runtime Web
- OpenCV.js
- MediaPipe (live streaming)

Web API
- WASM
- WebGPU
- WebNN

Web Engines
- Web Browser (e.g., Chrome/Edge)
- JavaScript Runtime (e.g., Electron/Node.js)

Hardware
- CPU
- GPU
- Other Accelerators
TensorFlow.js WebGPU Backend

- Began the effort in Sep 2019 and own the WebGPU backend
- Work closely with Google TensorFlow Team
- Alpha release together with other backends
- 2/3 ops are supported, and all the ops in benchmarks are supported
- Much better perf than WASM backend with middle or large sized models
- 1.5x - 3x perf of WebGL backend (1/3 in 2019)
- Provide feedbacks and requests to WebGPU, like Video Uploading, Timestamp Query, Workgroup Memory Init, Uniformity Analysis, etc.
WebNN

- Began the effort in Oct 2018, initiated the CG and chair it
- Work closely with Microsoft, Google, etc.
- Based on native ML APIs in OSes, brings near-to-native performance
- Implementing in Chrome, starting from CPU solution based on XNNPACK
英特尔“Web图形和机器学习”团队
About The Team

- Started the effort on the web in Nov 2011
- Intel rep of Khronos WebGL Working Group and W3C WebGPU Working Group
- Spec editor and key contributor to W3C Web Machine Learning WG
- 4000+ patches in Specs, CTS (Conformance Test Suite), Chromium and related projects (ANGLE, Dawn, etc.)
- Default owner of all Web Graphics issues on Intel platforms
Stay Connected

- Email: yang.qu@intel.com
- For Web Graphics questions, feel free to submit them to https://github.com/webatintel/webgraphicsforum/issues, if you couldn’t find a good fit in global communities
Backup
WebGPU Compute

Random Global Memory Access

Workgroup Memory & Barriers

Atomic Functions
**Conformance on Intel Platforms**

- **Intel driver CI (hardware, driver)**
  - Daily on 10+ platforms, including pre-production ones
  - Monthly update with Chromium, WebGL CTS, ANGLE end2end tests, Dawn end2end tests
- **Chrome CI (browser, test suites)**
  - Mature Chrome, including test framework, to better understand Intel GPU generations and driver versions
  - Owner of Intel specific graphics issues in Chrome, add regression test cases
- **Team’s test**
  - Daily on 2 machines, one for Windows and one for Linux
  - On demand tests against latest hardware