Deep Neural Networks Benchmark Suite for FPGAs Utilizing a TensorFlow to Routing High-Level Synthesis

Amee M.S. Abdelali*, Daniel Holanda Noronha*, Steven J.E. Wilton*, and Lesley Shannon*

* School of Engineering Science; Simon Fraser University; Burnaby BC, Canada, V5A 1S6
* Dept. of Electrical and Computer Engineering, The University of British Columbia, Vancouver BC, Canada V6T 1Z4

Motivation: FPGAs vs. GPUs for DNNs Inference
- More computation for quantized networks
- More efficient for irregular structures
- Higher performance per watts
- Higher acceleration (parallelism, pipelining, etc.)

→ FPGAs are used as inference platforms for deep learning applications

Deep Learning Benchmark Suite for FPGAs
- Enables evaluating the compatibility of different FPGA architectures and CAD algorithms for deep learning application
- Comprises a diverse selection of deep learning applications, e.g., image classification, video classification, speech recognition, and language modeling
- Based on a complete design flow, starting from deep learning application described in python, and ending with an FPGA design

A Representative Collection of Deep Learning Applications

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Application</th>
<th>Affiliation</th>
<th>Input</th>
<th>Parameters</th>
<th>Layers</th>
<th>Functions</th>
<th>Freq (MHz)</th>
<th>Power (W)</th>
<th>Package Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>Image classification</td>
<td>UofT</td>
<td>224×224 RGB image</td>
<td>60</td>
<td>4</td>
<td>✓ ✓ ✓</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inception-V3</td>
<td>Image classification</td>
<td>Google</td>
<td>224×224 RGB image</td>
<td>5</td>
<td>✓ ✓ ✓ ✓</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VGG-16-C</td>
<td>Image classification</td>
<td>Oxford</td>
<td>224×224 RGB image</td>
<td>124</td>
<td>✓ ✓ ✓</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResNet-54</td>
<td>Image classification</td>
<td>Microsoft</td>
<td>224×224 RGB image</td>
<td>30</td>
<td>✓ ✓ ✓ ✓</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3D</td>
<td>Video classification</td>
<td>Facebook AI</td>
<td>16Frame 128×171 RGB</td>
<td>127 ✓ ✓ ✓ ✓</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeepSpeech</td>
<td>Speech recognition</td>
<td>Baidu Research</td>
<td>power normalized audio 20ms windows</td>
<td>38 ✓ ✓ ✓ ✓ ✓</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


This research has been funded in part by the Computing Hardware for Emerging Intelligent Sensory Applications (COHESA) project. COHESA is financed under the National Sciences and Engineering Research Council of Canada (NSERC) Strategic Technologies grant number NETGP485577-15.