Name of Certifying Engineer(s): Ben Hawks

Email of Certifying Engineer(s): bhawks@fnal.gov

Name(s) of System Under Test: Xilinx Pynq-Z2

Division (check one):

**✓** Open

* Closed

Category (check one):

**✓** Available

* Preview
* Research, Development, and Internal (RDI)

Benchmark(s) (check all that apply):

* Visual Wake Words
* Keyword Spotting

**✓** Anomaly Detection

**✓** Image Classification

Please fill in the following table adding lines as necessary:

|  |  |  |
| --- | --- | --- |
| System Under Test Name | Benchmark | Accuracy/AUC |
| Pynq Z2 | Anomaly Detection | 74.6%/82% |
| Pynq Z2 | Image Classification | 76.5%/88% |

For each SUT, is the benchmark Accuracy/AUC target met? (Not a requirement for the Open division) (check all that apply):

* Yes (Visual Wake Words … 80% Accuracy)
* Yes (Keyword Spotting … 90% Accuracy )
* Yes (Anomaly Detection … 0.85 AUC)
* Yes (Image Classification ... 85% Accuracy)
* No, for some combination of benchmark, scenario and SUT

For each SUT and benchmark, did the submission run on the whole validation set in accuracy mode? (check one):

**✓** Yes

* No

For each SUT and benchmark, does the submission use the EEMBC Runner? (check one)

**✓** Yes

* No

For each SUT and benchmark, is the same code run in accuracy and performance modes? (check one)

**✓** Yes

* No

Are the weights calibrated using data outside of the official calibration set? (check one)

* Yes

**✓** No

What numerics does the submission use? (check all that apply)

* INT4

**✓** INT8

* INT16

**✓** UINT8

* UINT16
* FP11
* FP16
* BF16
* FP32

**✓** Other, please specify: ap\_fixed <32,6>, <16,6>, <14,6>, <11,11>, <11,6>, <11,1>, <12,6>, <9,6>, <8,8>, <8,6>, <8,3>, ap\_ufixed <8,0>

What backend does the submission use? (check all that apply)

* Vendor backend, please name:
* TF-Lite Micro
* Micro TVM

**✓** Other, please specify: hls4ml

Which of the following caching techniques does the submission use? (check all that apply, ideally none):

* Caching Inputs between iterations
* Caching responses between iterations
* Caching intermediate computations between iterations

Which of the following techniques does the submission use? (check all that apply, ideally none if submitting to the closed division.)

**✓** Quantization aware training

* Wholesale weight replacement
* Weight supplements
* Discarding non-zero weight elements
* Pruning
* Modifying weights during the timed portion of an inference run
* Hard coding the total number of queries
* None of the above

Is the submission congruent with all relevant MLPerf rules?

**✓** Yes

* No

If the answer to the above question is no, please explain:

For each SUT, have you filled out the JSON system description file?

**✓** Yes

* No

For each SUT, does the submission accurately reflect the real-world performance of the SUT?

**✓** Yes

* No

Does your submission include the following: (check all that apply)

**✓** System description file

**✓** Code that implements the benchmarks

**✓** Code/scripts that train the model(s) (Open Division)

**✓** Metadata that describes each system-implementation combination tested

**✓** Scripts that set up and execute each system implementation tested

**✓** Result logs for each system implementation tested

**✓** This Checklist