Hardware Implementation of Image Classifiers

Ansh Vivek (210020002)
Sankalp Nagaonkar (210020031)
Yash Meshram (210020053)

Supervisor: Prof. Samba Raju

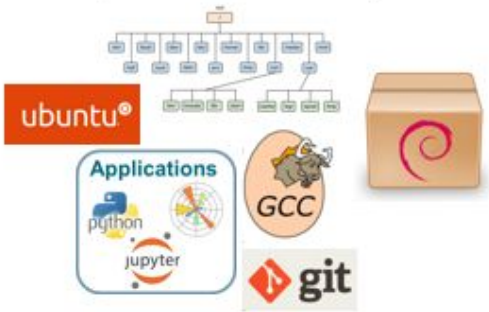
Previous Evaluation

- Matlab
  - Accuracy of 19.691% after 15 epochs
  - One major reason for this was the unavailability of pretrained model
- ZCU 104
  - We faced issues with zynet
  - Error in compiling Vivado Project. Hence, couldn’t get a bitstream

PYNQ

- PYNQ is an open source software framework based on Linux and Ubuntu as well as Python language, mainly providing a development platform based on Python for Xilinx boards
- Based on Hardware Overlay and API in PYNQ and xfOpenCV hardware acceleration library provided by Xilinx, hardware acceleration can be achieved by making full use of (Programmable Logic) resources in Zynq boards
- The PYNQ Image is mainly composed of three parts: Ubuntu Root FS + Python 3.x Package + Jupyter Notebook

Ubuntu Bionic AArch64
Root FS with **Python 3.6**



+

2018.3 Xilinx
aarch64 PetaLinux

+

Xilinx PYNQ Python
module and classes

=

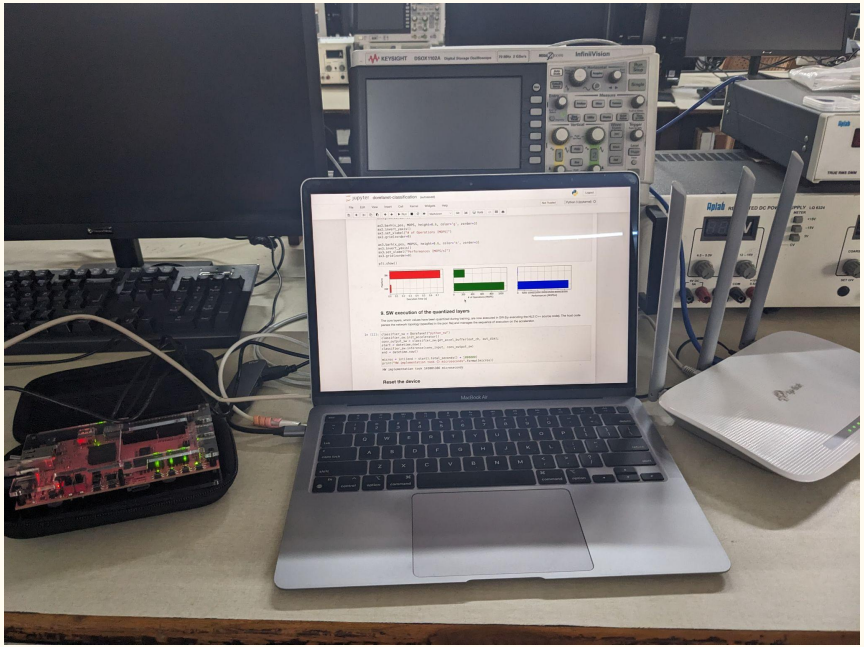
PYNQ

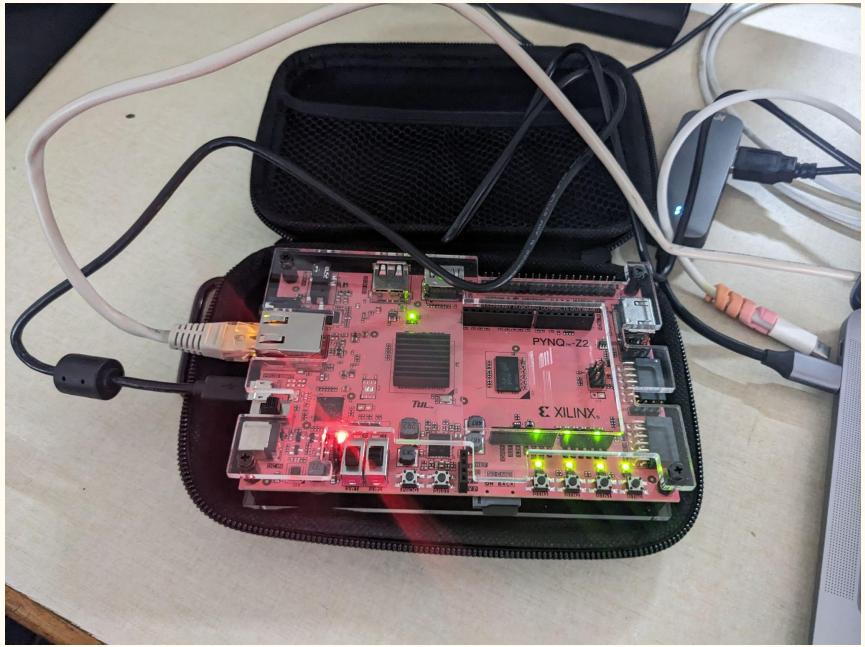
(.img)

PYNQ -Z2 : Setup

- Need the boot image of the board
- Flash the image of the board on an SD card using Windows Disk Imager
- Power the board using a USB cable
- To connect the board to internet, plug it into the router's LAN cable
- To access the board, connect to the IP address of the board which is provided by the router
- Paste the router IP in the browser as an address and append it with ":9090"
- This will open a Jupyter Notebook interface
- Now we can run programs on the board using the Jupyter Notebook interface

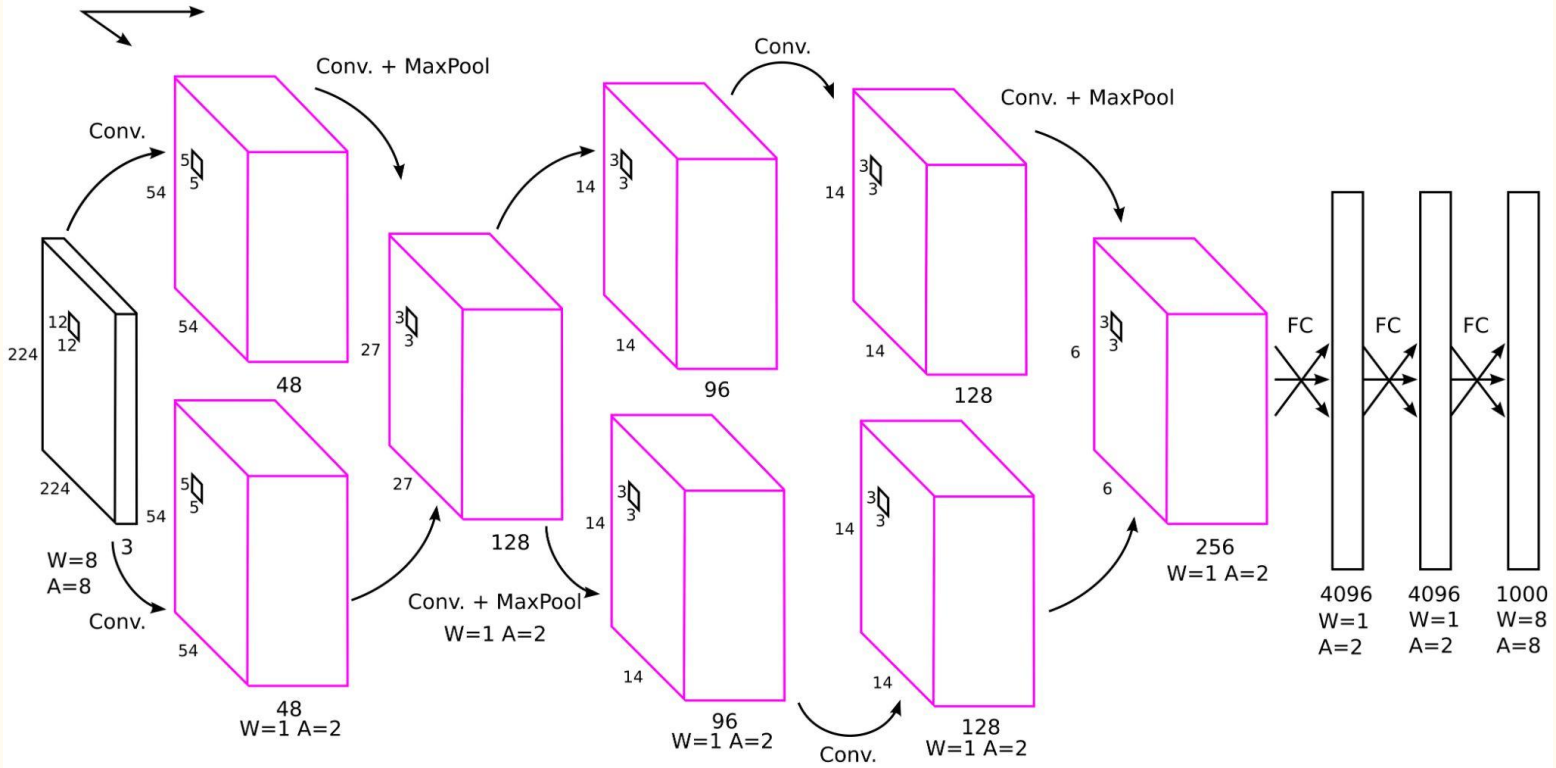
PYNQ -Z2 : Setup





DoReFa Net

- The network we used for inference is a variant of DoReFaNet.
- The pink layers are executed in the Programmable Logic at reduced precision (1 bit for weights, 2 bit for activations) while the other layers are executed in python.
- It quantizes the neural network by reducing the precision of weights and activation values.


Conv.
Conv. + MaxPool
Conv.
Conv. + MaxPool
FC
FC
FC
224
224
3
W=8
A=8
12
12
54
54
5
5
48
54
54
5
5
48
W=1 A=2
27
27
3
3
128
Conv. + MaxPool
W=1 A=2
14
14
3
3
96
14
14
3
3
96
W=1 A=2
14
14
3
3
128
14
14
3
3
128
W=1 A=2
Conv.
6
6
3
3
256
W=1 A=2
4096
W=1
A=2
4096
W=1
A=2
1000
W=8
A=8

Results

- Accuracy: Top-5- 78%
- Time: HW-0.322, SW-2.009
- MOPs: HW-1073.856, SW-238.176
- MOPs per sec: HW-3330.615, SW-118.514