

Computer Engineering Lab

Philip Leong, David Boland and Barry Flower | Computer Engineering Laboratory
School of Electrical and Information Engineering,
The University of Sydney



THE UNIVERSITY OF
SYDNEY

- › Focuses on how to use FPGAs to solve demanding problems: novel architectures, applications and design techniques for problems combining signal processing and machine learning
- › Expertise
 - Deep neural network acceleration
 - Time series prediction
 - Signal processing
 - FPGA design
- › Collaborations
 - Xilinx, Intel, Exablaze
 - Defence, TASDCRC and DSTG
 - clustertech.com
- › Ex-students
 - Waymo, Intel, Synopsys

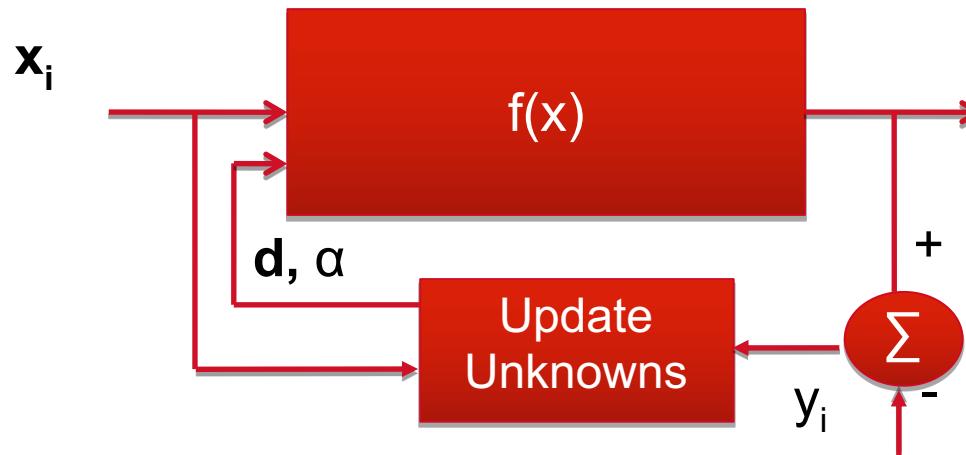


Motivation for FPGAs

- › Field programmable gate arrays (FPGAs) are COTS, user-customisable integrated circuits
- › They offer an opportunity to provide ML algorithms with higher throughput and lower latency through
 - Exploration – easily try different ideas to arrive at a good solution
 - Parallelism – so we can arrive at an answer faster
 - Integration – so interfaces are not a bottleneck
 - Customisation – problem-specific designs to improve efficiency
- › **Describe some of our work on ML hardware implementations that use these ideas**



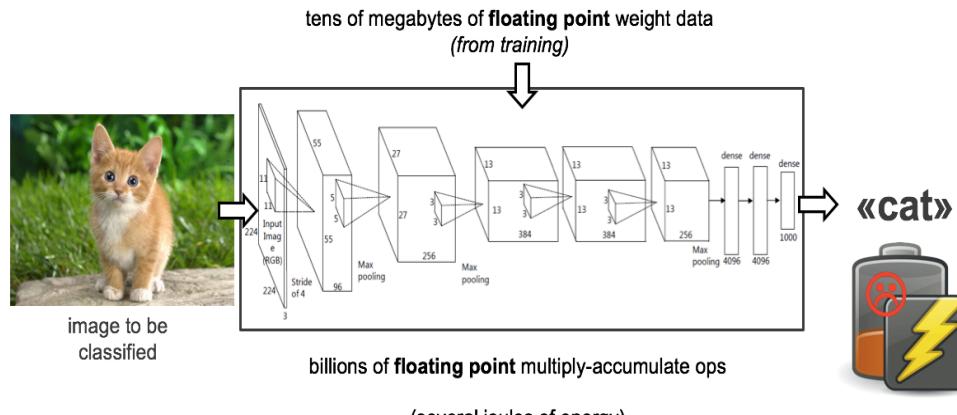
ARC Linkage with Exablaze



- › A family of kernel methods that can do simultaneous learning and inference
 - Highest reported throughput 80 Gbps (TRETS'17)
 - Lowest reported latency 80 ns (FPT'15)
 - Highest capacity (FPGA'18)

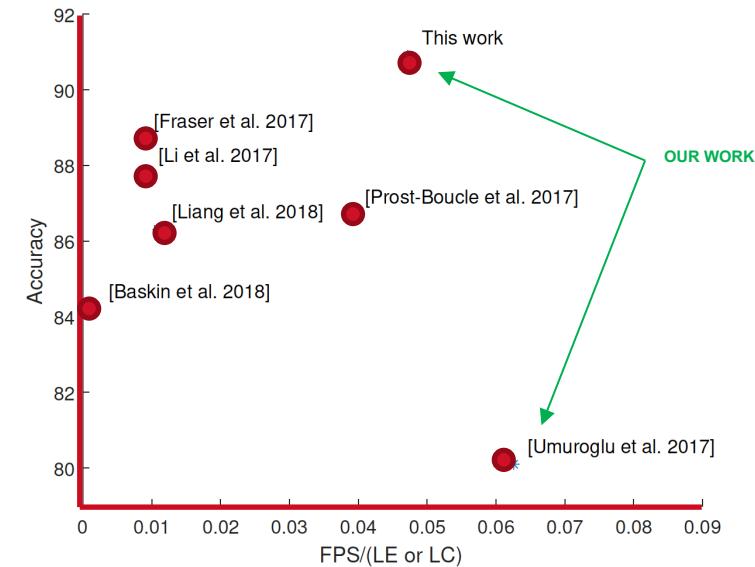
Parallelism: Binarized Neural Networks

Collaboration with Xilinx



$$\begin{matrix}
 -0.4 & -0.4 & 0.9 \\
 0.9 & 0.4 & 0.8 \\
 0.4 & -0.4 & -0.4
 \end{matrix} \approx 0.2 \quad \begin{matrix}
 -1 & -1 & 1 \\
 1 & 1 & 1 \\
 1 & -1 & -1
 \end{matrix}$$

W aW^B

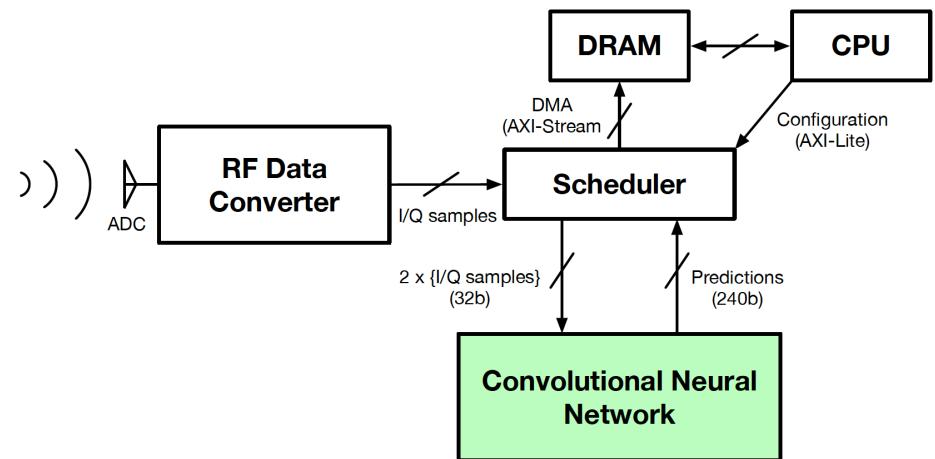
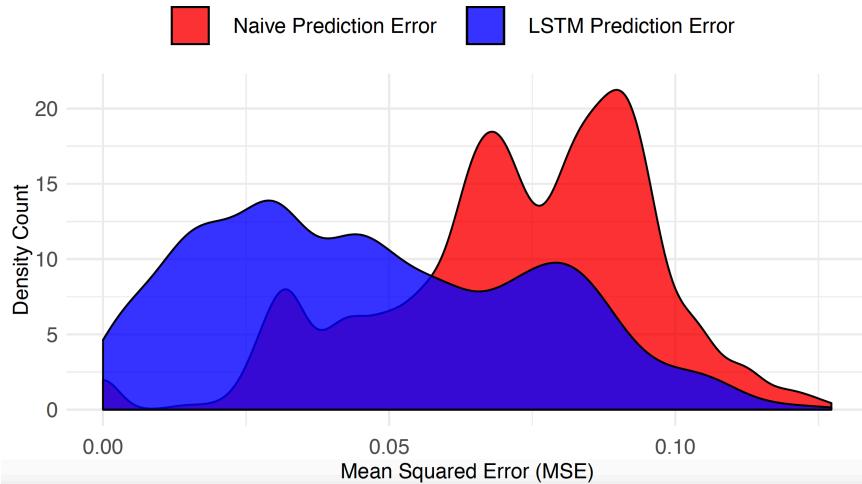


Ours is the most accurate and fastest reported FPGA-based CNN inference implementation
CIFAR10: 90.9% acc, 122K fps (TRETS'19)

Integration: Radio Frequency Signals

Next Generation Technology Fund

- › Processing RF signals remains a challenge
 - FPGAs allow integration of radio, machine learning and signal processing



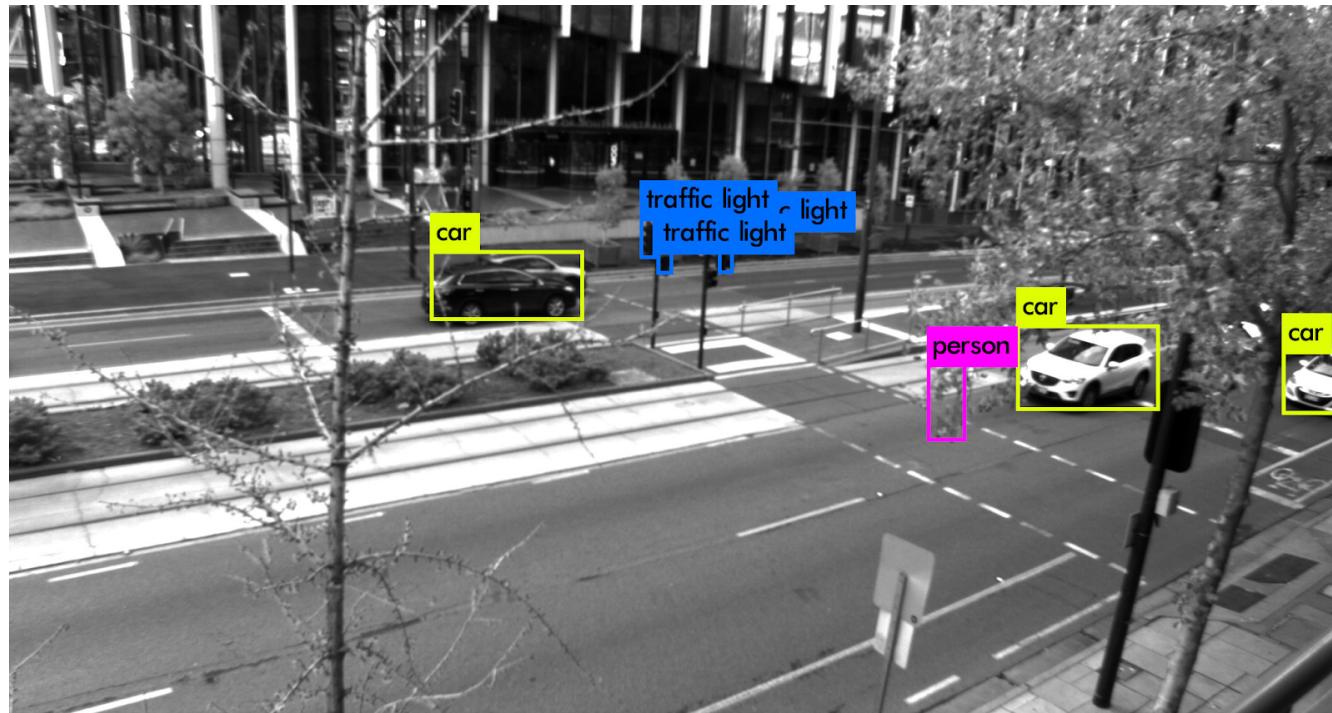
LSTM Spectral prediction: 4.3 μ s latency on Ettus X310 XC7K410T (MILCOM'18)

Ternary Modulation classifier: 488K class/s, 8us latency, Xilinx ZCU111 RFSoC (FPT'19)

Customisation: High Dynamic Range Signals

Defence Innovation Hub

- › Implementation of a neuromorphic high dynamic range camera-based object detector on FPGAs
- › Significantly improved accuracy in high contrast situations



- › Machine learning will enable intelligent sensors and control system crucial to trusted autonomous systems
 - Combine conventional sensors, e.g. radar, lidar, video, radio with powerful object recognition and scene interpretation ability i.e. ML at the edge
- › FPGAs offer EPIC (exploration, parallelism, integration, customization) advantages for miniaturization, reduced energy, and improved performance in such applications

