UPISC UNIVERSITY OF PITTSBURGH INFRASTRUCTURE SENSING COLLABORATION WORKSHOP

NATIONAL ENERGY TECHNOLOGY LABORATORY

Machine Learning on Intermittently Powered Microcontrollers

Paul Kyros, Yukai Song, Christopher Brubaker, Inhee Lee, Jingtong Hu - Swanson School of Engineering - Electrical and Computer
Contact: PJK33@pitt.edu, JTHU@pitt.edu

Motivation

Can we power small DNN models with ambient energy?

- Hard to reach areas: underwater, under bridge
- Harsh environment: battlefield, space
- Scientific research: wild animals, volcano

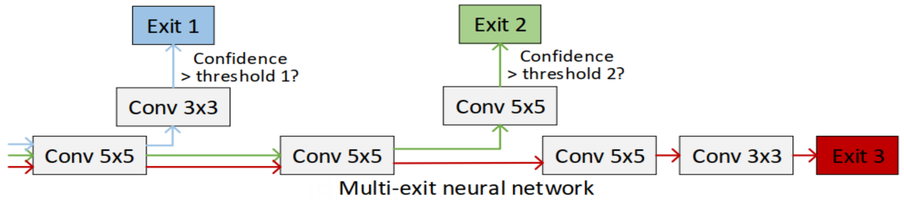
Challenges

- How to fit the multi-exit network on MCUs while keeping a high accuracy of each exit
  - Large model size vs. small storage
  - High computation cost vs. low available energy
- How to select the exit for each event to achieve
  - A high average accuracy
  - A low average latency in the long-term

Contributions:

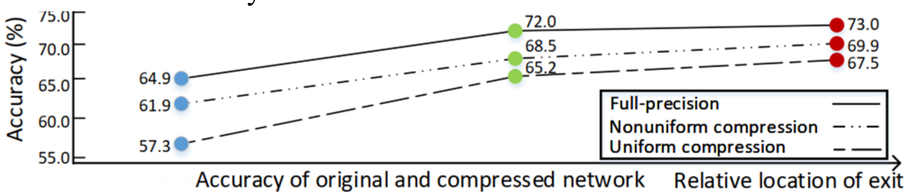
Intermittent Inference Model

- Guarantee an inference result before power failure occurs



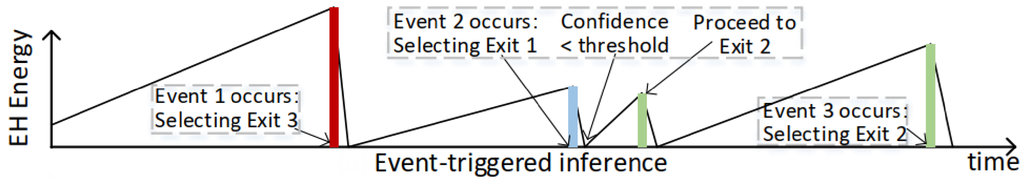
Power Trace-Aware Compression

- Compress multi-exit networks to fit onto MCUs while maximizing the average inference accuracy



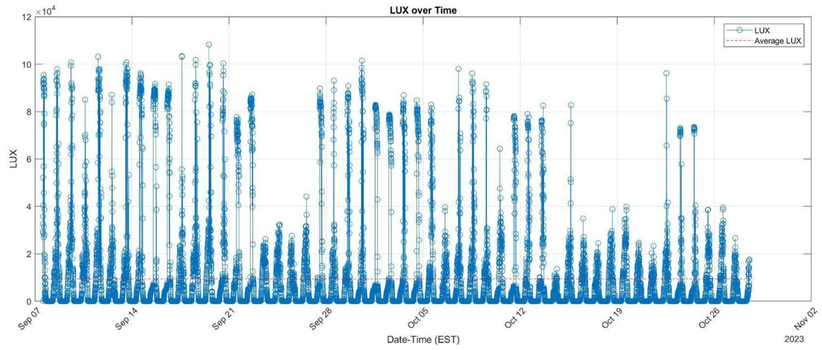
Runtime Adaptation

- Runtime exit selection selects the exit for each event, considering the EH environment and difficulty of processing each event



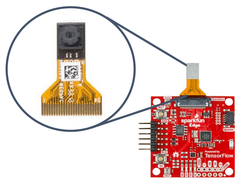
Measuring Power Trace

- **HOBO sensor measures light intensity (LUX)**
  - Analyzed and plotted in MATLAB



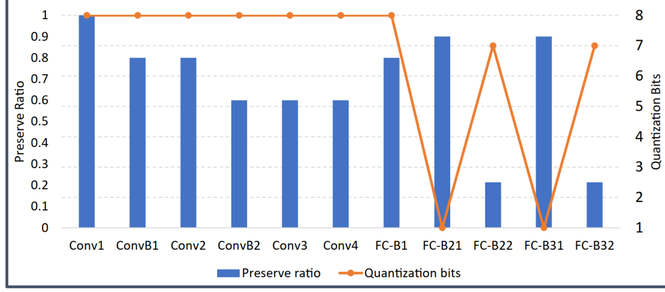
- **Energy profile of solar panel determined via source meter**
  - Keithley 2401 SMU Low Voltage SourceMeter

- **Measuring MAX achievable inferences powered by supercapacitor**
  - Powered by solar panel and charge and fire circuit using 5.5V 4F supercapacitor
  - Inferences run using a Multi-exit Convolutional Neural Network
  - STM32 Nucleo-64 board to run inferences
  - Uses a SparkFun EDGE board to capture images
  - Exit chosen based upon power condition of the system



Experimental Results

- Generated Pruning and Quantization Policy



- **IEpmJ and FLOPS**
  - **I**nteresting **E**vents **p**er **m**illiJoule and FLOPS
  - Maximizing the average accuracy of all events is equivalent to maximizing *IEpmJ*