The University of Texas at Austin
Chandra Department of Electrical and Computer Engineering
Cockrell School of Engineering

Spring 2026

ECE 361E: Machine Learning and Data Analytics for Edge AI
HW3 Assigned: Feb 17, DUE: Feb 26 (11:59:59pm CST)

Work in teams of two students. At the end of the PDF file, insert a paragraph where you describe each member's contribution and two valuable things you learned from this homework. Only one submission per group is required.

Introduction

In this homework, you will deploy popular deep learning models on two computational platforms, namely Odroid MC1 and RaspberryPi 3B+, using Open Neural Network Exchange (ONNX), one of the most used frameworks for ML models deployment. To compare the inference latency, accuracy and energy consumption during inference, you will use the CIFAR10 dataset. By working on this assignment, you will:

- Understand the process of deploying PyTorch models on edge devices using ONNX (e.g., converting a model to ONNX, deploying it on real edge devices, and doing inference);
- Measure latency, accuracy, energy consumption and CPU temperature variation during inference in order to understand the impact of model inference on edge devices;
- Understand the importance of designing models optimized for edge devices.

Problem 1 [35p]: PyTorch Evaluation of VGG models

Question 1: [10p] Starting from the VGG11 code we provide in ***HW3_files***, create another file called ***vgg16.py*** (under the ***models*** folder) using the VGG16 model architecture (check ***HW3_README*** to see how VGG models are modified for CIFAR10).

Question 2: [15p] Train the VGG11 and VGG16 models ***in parallel*** on Lonestar6 using ***main.py*** (see **Appendix A3.1** for guidelines on running tasks in parallel on TACC). Follow the ***TODO*** parts from the ***main.py*** code to make it work properly and then add the necessary code to complete ***Table 1***.

Table 1

Model	Training accuracy [%]	Test accuracy [%]	Total time for training [s]	Number of trainable parameters	FLOPs	GPU memory during training [MB]
VGG11						
VGG16						

Question 3: [10p] On the same graph, draw a *single* plot consisting of two curves that show the variation of the test accuracy of both VGG11 and VGG16 models vs. the number of training epochs. Based on this plot and the results in **Table 1**, compare VGG11 and VGG16 (accuracy, training time, parameters, FLOPs, etc. and the corresponding tradeoffs); explain which model architecture you would choose for *training* based on the results in **Table 1**.

Problem 2 [45p]: Deployment on Edge Devices Using ONNX

Question 1: [5p] Create a new file named ***convert_onnx.py*** that contains the function needed to convert PyTorch models into the ONNX format (as shown in the HW3 demo) and save the converted models using the ***.onnx*** file extension (check ***HW3_README***).

Question 2: [20p] Use the ***deploy_onnx.py*** file from the ***HW3_files*** folder to deploy your VGG11 and VGG16 ONNX models on the RaspberryPi 3B+ and Odroid MC1 devices. Complete the ***TODO*** parts in the ***deploy_onnx.py*** file. Add some additional code to calculate and report the accuracy of the four ONNX models on the **test dataset**. To evaluate on the test dataset, perform inference on both Odroid and RaspberryPi devices using the *entire* test dataset available in the ***HW3_files/test_deployment*** folder; then, complete ***Table 2*** (check ***HW3_README***).

Table 2

	Total inference time [s]		RAM memory [MB]		Accuracy [%]	
	MC1	RaspberryPi	MC1	RaspberryPi	MC1	RaspberryPi
VGG11						
VGG16						

Question 3: [20p] For the VGG11 model, draw a *single* plot showing the variation of power consumption over time [s] for *both* devices (i.e., one curve for RaspberryPi and one curve for MC1). Then, on a different plot, show the variation of average temperature measurements of the CPU for each device over time [s] (i.e., one curve for RaspberryPi and one curve for MC1). Complete ***Table 3***.

Based on these plots and the data in ***Table 2*** and ***Table 3*** compare the performance of the two edge devices (check ***HW3_README***). Which device would you prefer for *inference*? Explain your choice.

Table 3

Model	MC1 total energy consumption [J]	RaspberryPi total energy consumption [J]
VGG11		
VGG16		

Problem 3 [20p + 10Bp]: MobileNet-v1 on Edge Devices

Question 1: [5p] Train MobileNet-v1 on the CIFAR10 dataset using the ***main.py*** file and the ***mobilenet.py*** model on Lonestar6. Extend ***Table 1*** in Problem 1 by inserting *a new row below the table* to include the results obtained for MobileNet-v1. Insert this modified Table 1 below in your report.

Question 2: [15p] Use ***convert_onnx.py*** (from **Problem 2, Question 1**) to convert the MobileNet-v1 model into ONNX and deploy it using the ***deploy_onnx.py*** file on both Odroid MC1 and RaspberryPi 3B+ devices. Extend ***Table 2*** and ***Table 3*** *with new rows* to include the results you got for MobileNet-v1. Insert these modified tables below in your report.

BONUS Question 3: [10Bp] Analyze and discuss the data in the modified (i.e., extended) ***Table 1***, ***Table 2*** and ***Table 3*** from **Problem 2, Question 3** and **Problem 3, Question 2**. Which model delivers better results (i.e., latency, energy, accuracy) on each edge device? Explain.

Submission Instructions

Include your solutions into a single zip file named <**Team#**>**.zip**. The zip file should contain:

1. A single PDF file containing all your results, tables, plots, and discussions.
2. Your code files, named suggestively (e.g., **p1_q1.py** for **Problem 1 Question 1** code).
3. A *readme.txt* file explaining all your items in the zip file.

Good luck!