ECE 361E: Homework 3

Sidharth Babu, SNB2593 Tianda Huang, TH32684

February 23, 2023

Problem 1

Question 2

Model	Training Accuracy(%)	Test Accuracy (%)	Total time for training (s)	Number of Trainable Params	Floating Point Operations	GPU memory during training (mb)
VGG11	97.57	76.48	3011.79	9,750,922	306587648	1215
VGG16	97.86	78.89	3622.42	14,655,050	551954432	1425

Question 3

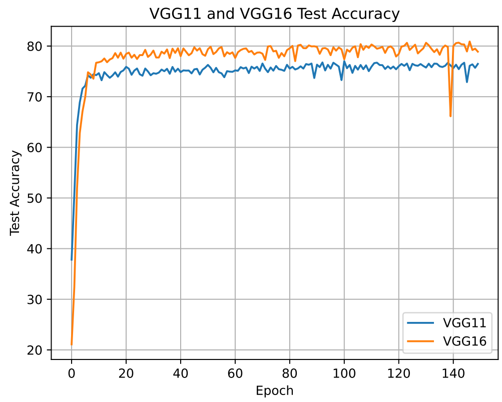

Figure 1: Test Accuracy of VGG11 and VGG16

VGG16 performs incrementally better than VGG11 on both train and test. However it has 1.5x the amount of trainable parameters and 1.8x the amount of floating point operations. The small accuracy boost is not worth the increase in computational complexity, and therefore we would choose VGG11 to train.

Problem 2

Question 2

	Total Inference Time [s]		RAM memory [MB]		Accuracy[%]	
	MC1	RaspberryPi	MC1	RaspberryPi	MC1	RaspberryPi
VGG11	658.23	680.61	477	171	76.48	76.48
VGG16	990.92	1172.01	353	192	78.89	78.89
MobileNet	491.65	329.29	302	140	77.75	77.75

Model	MC1 Total Energy Consumption [J]	RPI total Energy Consumption
VGG11	6574.78	3739.29
VGG16	10106.78	6381.64
MobileNet	4196.58	1877.44

Question 3

Problem 3

BONUS

The MobileNet model performs the best in terms of composite performance across both platforms. It is the most energy efficient model, is the fastest model in terms of inference, and has accuracy on par with the other two models. MobileNet's key advantage is the usage of depthwise convolution, which significantly reduces the computation cost of the forward pass on convolutional layers. Depthwise convolution works by splitting the convolution operation into two steps. First, a single filter is applied to each input channel, which is followed by a pointwise convolution to combine the output channels into one output. This is significantly faster than convolving one learned filter with the entire input tensor.