

# Machine Learning 2020 - Homework 7 Report

學號：b08902100, 系級：資工一, 姓名：江昱勳

1. 請從 Network Pruning/Quantization/Knowledge Distillation/Low Rank Approximation/Design Architecture 選擇兩者實做並詳述你的方法，將同一個大 model 壓縮至接近相同的參數量，並紀錄其 accuracy。

這裡我搭建了兩個模型。

第一個模型是想實驗模型設計的部份 (把模型變瘦但變深)，總共有 13 層的 CNN 並接上 3 層的全連接層，每一層 CNN 層都是由一組 Depthwise Convolution + Pointwise Convolution 構成，前 5 層的活化函數皆採用 ReLU6，後面 8 層則採用 swish，而僅只有前 4 層有使用 max pooling，channel 的部份分別為 16,32,32,64,64,64,64,128,128,128,128,128,128,最後全連接層的部分則為 160,160,11，總共有 186123 個參數，不過此 model 實際訓練的時候似乎訓練不起來，有可能是過深的原因，訓練了 200 個 epoch 後在 validation 上的 accuracy 仍只有 0.205539，與其在 training set 上的 accuracy 差不多。

第二個模型主要是嘗試 pruning 的部份，原本的模型總共有 8 層 CNN 並接上 3 層的全連接層，每一層 CNN 層都是由一組 Depthwise Convolution + Pointwise Convolution 構成，前 5 層的活化函數皆採用 ReLU6，後面 3 層則採用 swish，而僅只有前 4 層有使用 max pooling，channel 的部份分別為 24,48,96,96,192,192,192,192，最後全連接層的部分則為 160,160,11，該模型在 validation 上的正確率為 0.846064。使用 0.95 的 pruning rate 後參數減少為 186750，而 validation 的正確率也減少至 0.834694。

2. [Knowledge Distillation] 請嘗試比較以下 validation accuracy (兩個 Teacher Net 由助教提供) 以及 student 的總參數量以及架構，並嘗試解釋為甚麼有這樣的結果。你的 Student Net 的參數量必須要小於 Teacher Net 的參數量。

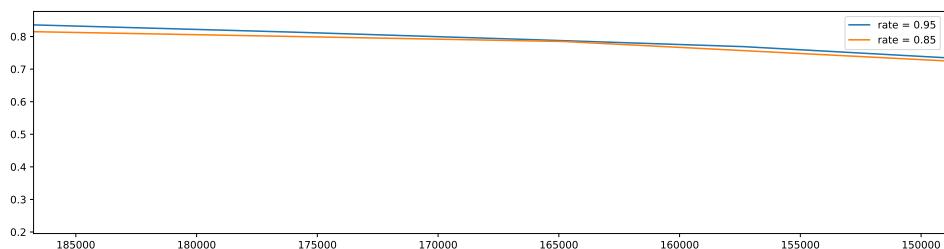
- x. Teacher net architecture and # of parameters: torchvision's ResNet18, with 11182155 parameters.
- y. Student net architecture and # of parameters: 144389
  - a. Teacher net (ResNet18) from scratch: 80.09%
  - b. Teacher net (ResNet18) ImageNet pretrained & fine-tune: 88.41%
  - c. Your student net from scratch: 78.8630 %
  - d. Your student net KD from (a.): 79.1254 %
  - e. Your student net KD from (b.): 80.6414 %

我的 student net 模型總共有 9 層 CNN 並接上 3 層的全連接層，每一層 CNN 層都是由一組 Depthwise Convolution + Pointwise Convolution 構成，前 5 層的活化函數皆採用 ReLU6，後面 4 層則採用 swish，而僅只有前 4 層有使用 max pooling，channel 的部份分別為 19,38,76,76,76,152,152,152,152，最後全連接層的部分則為 128,128,11。

以上訓練結果皆為訓練 200 epoch 後的結果，可以發現原先 train from scratch 即可達到 78% 左右的準確率，而因此讓他跟 80% 的模型學習則可以讓他在 200 epoch 時就幾乎將 teacher 學會的東西學走達到 79% 的準確率，而若是讓他與更加厲害的老師學習則可以看到在 200 epoch 時就能學到 80% 的準確率，這甚至超過了 train from scratch 的 ResNet18，因此可以看到老師的準確率會影響模型學習的成果，如果與準確率高的模型學習可以較快達到一定的準確率。

3. [Network Pruning] 請使用兩種以上的 pruning rate 畫出 X 軸為參數量，Y 軸為 validation accuracy 的折線圖。你的圖上應該會有兩條以上的折線。

我分別使用了 0.95 的 pruning rate 與 0.85 的 pruning rate 做實驗，以下是參數量對 validation accuracy 的折線圖。



此外我也把訓練過程記錄下來了，我這裡是以每 200 個 epoch 就 prune 一次，以下是訓練時的 validation accuracy 與 training accuracy，以及 loss 的部份。

