學號：B07902034　系級：資工二　　姓名：王昱凱

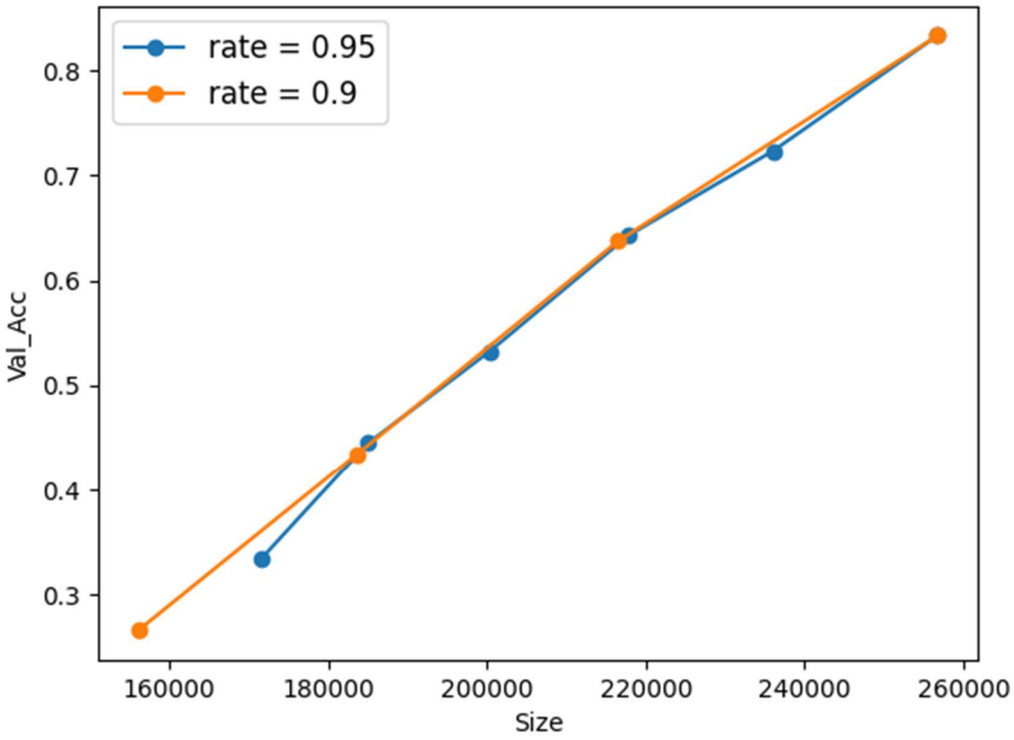
1. 請從 Network Pruning/Quantization/Knowledge Distillation/Low Rank Approximation 選擇兩個方法(並詳述)，將同一個大 model 壓縮至同等數量級，並討論其 accuracy 的變化。 (2%)

   我挑選的兩個方法為Knowledge Distillation和Quantization，首先我Knowledge Distillation是按照助教在colab的實作，student net的架構也是按照colab上的架構，而teacher net是使用resnet18，在validation set上，實作Knowledge Distillation前，也就是使用resnet18的準確率為0.88左右，而實作後的準確率降低為0.83左右，而我Quantization的實作方法也和助教在colab上的相同，先將32-bit的tensor轉換成16-bit的float，接著再將16-bit做min-max正規化轉成8-bit，而實作Quantization後在validation set上的準確率從0.83降低至0.82，降低的幅度其實不大

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。(2%)
   x. Teacher net architecture and # of parameters: torchvision’ s ResNet18, with 11,182,155 parameters.
   y. Student net architecture and # of parameters: 架構與colab上相同，有8層的cnn，其中第二層以後有使用depthwise&pointwise，參數量為256779
   a. Teacher net (ResNet18) from scratch: 80.09%
   b. Teacher net (ResNet18) ImageNet pretrained & fine-tune: 88.41%
   c. Your student net from scratch: 76.12%
   d. Your student net KD from (a.): 80.87%
   e. Your student net KD from (b.): 83.32%
   我在(c)(d)(e)中獲得的validation accuracy是利用training set訓練70個epoch後取準確率最高的結果，而會有這樣的結果是因為(d)(e)有使用Knowledge Dis

tillation，是利用teacher net得到的label來計算loss，因此準確率會比起自己從頭開始訓練的(c)要有更好的準確率，而(d)(e)的差別在於他們分別學習的大model本身就有不同的準確率，因此(e)的準確率會比(d)來的更好

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。(2%)



在本題，我是以colab上student net為原始架構，並且分別以0.95和0.9的pruning rate來做為觀察，而我得到的結果發現兩種折線接近重疊的情況，但rate = 0.9的折線明顯準確率下降得更快，因此我推測pruning rate越低，準確率下降的速度也會越快