學號：B05902010　系級：資工四　　姓名：張頌平

Hw3 Report

1. 請說明你實作的 CNN 模型(best model)，其模型架構、訓練參數量和準確率為何？(1%)

Ans:
我的模型架構為***VGG-16***，詳細層數如下：
(64, 64, MaxPool, 128, 128, MaxPool, 256, 256, 256, MaxPool, 512, 512, 512, MaxPool, 512, 512, 512, MaxPool)
每層皆有加上Batch Normalization，最後接上fully connected layer時的每層也有加上dropout(0.4) 防止訓練時參數過多而導致overfitting，詳細層數如下：
(25088->4096, 4096->1024, 1024->128, 128->11)

模型總訓練參數量為：121815627
batch_size: 32
learning rate: 7e-5
訓練過程中我有使用cross-validation，將train和validation的data分成三組，最後將其三次訓練的model做ensemble
準確率為：
validation score: 86.706, 87.485, 88.715
public test score: 89.300, 88.942, 89.479

ensemble public test score: 90.854

而以下的實驗皆會以第一組(原本的train和validation)訓練後的結果作為report的紀錄呈現

2. 請實作與第一題接近的參數量，但 CNN 深度（CNN 層數）減半的模型，並說明其模型架構、訓練參數量和準確率為何？(1%)

Ans:
CNN層數減半後的詳細層數如下：
(64, MaxPool, 128, MaxPool, 256, 256, MaxPool, 512, 512, MaxPool)
每層皆有加上Batch Normalization，最後接上fully connected layer時的每層也有加上dropout(0.4) 防止訓練時參數過多而導致overfitting，fully connected layer層數如下：
(100352->1024, 1024->128, 128->11)

模型總訓練參數量為：107398411
準確率為：
validation score: 82.886
public test score: 86.013

3. 請實作與第一題接近的參數量，簡單的 DNN 模型，同時也說明其模型架構、訓練參數和準確率為何？(1%)

Ans:
DNN層數如下(每層也有加上dropout(0.4))：
(150528->768, 768->128, 128->11)

模型總訓練參數量為：115706123
準確率為：
validation score: 31.458
public test score: 30.902

4. 請說明由 1 ~ 3 題的實驗中你觀察到了什麼？(1%)

Ans:
從上面的三個實驗比較後，滿明確的知道圖片的訓練是很看重每個pixel的方位關係，在純DNN的情況下，因為會先把圖片的dimension攤平，所以導致每個pixel之間的關係會被模糊化，儘管參數量差不多，但仍會導致訓練結果相差很多；而將CNN層度減半的情況下，因為原本的best model(VGG 16)的層數很深，因此在訓練時間上快上一些(200->160 sec/epoch)，也更快收斂，雖然最後的準確率稍微比較低，但也仍有一定水準(80%up)，如果在訓練時間、空間皆有限的情況下，將CNN層數減半，會是一個讓準確率維持一定水準的一個好選擇。

5. 請嘗試 data normalization 及 data augmentation，說明實作方法並且說明實行前後對準確率有什麼樣的影響？(1%)

Ans:
我是利用torchvision的transform來做data normalization和data augmentation
方法如下：
data normalization:
Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
(mean & std of ImageNet dataset)

data augmentation:
RandomresizedCrop(224): 將原本的圖檔用(256*256)讀進來，在訓練時隨機切(224*224)的影像來進行訓練(原因：目標物品可能並不佔據全部的圖片，縮小目標有機會更容易辨認種類)
RandomHorizontalFlip(): 將圖片隨機水平翻轉(p=0.5)
RandomRotation(30): 將圖片隨機轉-30~30度
ColorJitter(brightness=0.1, contrast=0.1): 亮度和對比隨機調整至0.9~1.1倍
以上的data augmentation的目的皆在於能夠讓同張圖片有更多方式呈現，避免過深的model和過長的訓練次數而導致overfitting。

沒有加data normalization & data augmentation
準確率為：
validation score: 70.466
public test score: 73.102

有加data normalization ＆ 沒有加data augmentation
準確率為：
validation score: 71.137
public test score: 73.998

沒有加data normalization ＆ 有加data augmentation
準確率為：
validation score: 85.656
public test score: 87.507

有加data normalization & data augmentation
準確率為：
validation score: 86.706
public test score: 89.300

6. 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析](1%)

Ans:
從下圖validation set的confusion matrix發現，大部分的結果預測都相當不錯，唯一比較容易搞混的是class 1 & class 2，總共會有0.1875原本是class 1的物品被辨認成class 2，而class 1 是乳製品，class 2 是甜點，他們在各自class的差異性比較大(牛奶和乳酪)，兩個class之間也有滿多相似處(乳酪長得滿像長條蛋糕?)，因此導致模型在部分情況會出現搞混的情況。