學號：R05922096 系級： 資工碩二 姓名：李哲安

1. Model description
   1. RNN (kaggle public = 14.18079) (feature=fbank)

Input(shape=[batch, 777, 69])

Bi-LSTM(hidden\_units=256, activation=tanh)

Bi-LSTM(hidden\_units=256, activation=tanh)

Dense(48, activation=softmax)

loss = cross entropy

Adam(learning\_rate=0.001)

* 1. RNN+CNN (kaggle public = 8.77966) (feature=fbank)

Input(shape=[batch, 777, 69])

Conv2D(kernel=[5, 5], num\_filters=10, activation=relu)

Bi-LSTM(hidden\_units=256, activation=tanh)

Bi-LSTM(hidden\_units=256, activation= tanh)

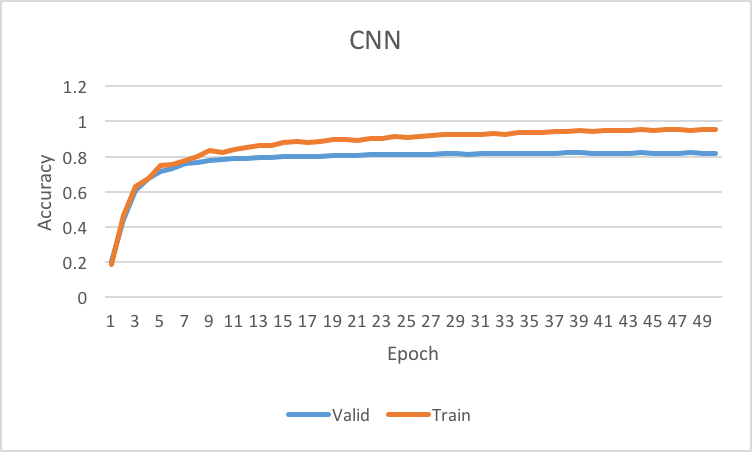
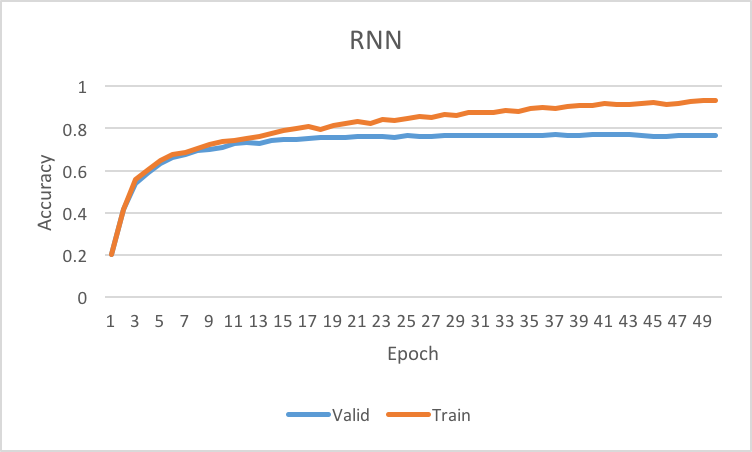
Bi-LSTM(hidden\_units=256, activation= tanh)

Dense(48, activation=softmax)

loss = cross entropy

Adam(learning\_rate=0.001)

1. How to improve your performance?
   1. Write down the method that makes you outstanding
      1. LSTM分數14.18079
      2. 加了CNN on acoustic features後分數-> 8.77966
      3. 去除連續label中的非連續label (aaaabcccc -> aaaacccc) -> 8.65536
   2. Describe the model or technique
      1. 同1.b題
   3. Why do you use it?
      1. 看了作業投影片的paper後覺得很厲害
2. Experimental results and settings
   1. Compare and analyze the results between RNN and CNN



CNN準確率上升速度和最終performance都優於RNN。

* 1. Compare and analyze the results with other models (other models can be variant of basic RNN, like LSTM, or some novel ideas you use)

測試過CNN的conv2D改為kernel shape = [5, feature\_dim]，最終表現只略高於LSTM(1.a)