1. Model description
    a. RNN (kaggle public = 14.18079) (feature=fbank)
        Input(shape=[batch, 777, 69])
        BiLSTM(hidden_units=256, activation=tanh)
        BiLSTM(hidden_units=256, activation=tanh)
        Dense(48, activation=softmax)
        loss = cross entropy
        Adam(learning_rate=0.001)
    b. RNN+CNN (kaggle public = 8.77966) (feature=fbank)
        Input(shape=[batch, 777, 69])
        Conv2D(kernel=[5, 5], num_filters=10, activation=relu)
        BiLSTM(hidden_units=256, activation=tanh)
        BiLSTM(hidden_units=256, activation= tanh)
        BiLSTM(hidden_units=256, activation= tanh)
        Dense(48, activation=softmax)
        loss = cross entropy
        Adam(learning_rate=0.001)

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

3. Experimental results and settings
   a. Compare and analyze the results between RNN and CNN
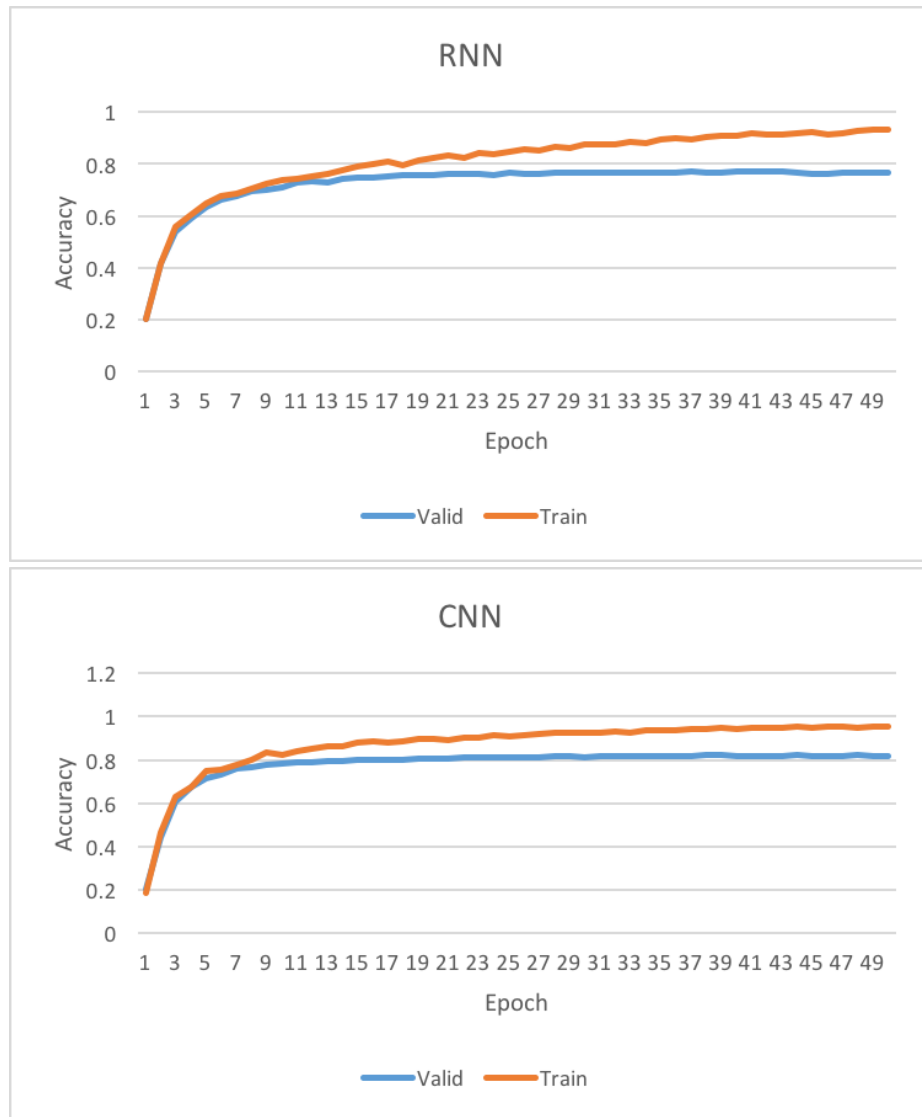
   

   CNN 準確率上升速度和最終 performance 都優於 RNN。
   b. Compare and analyze the results with other models (other models can be variant of basic RNN, like LSTM, or some novel ideas you use)
      i. 測試過 CNN 的 conv2D 改為 kernel shape = [5, feature_dim]，最終表現只略優於 BiLSTM(1.a)
      ii. 嘗試過 RNN、LSTM、BiLSTM 三種不同的 RNN cell 架構，最後表現最好的 是 BiLSTM(1.a)