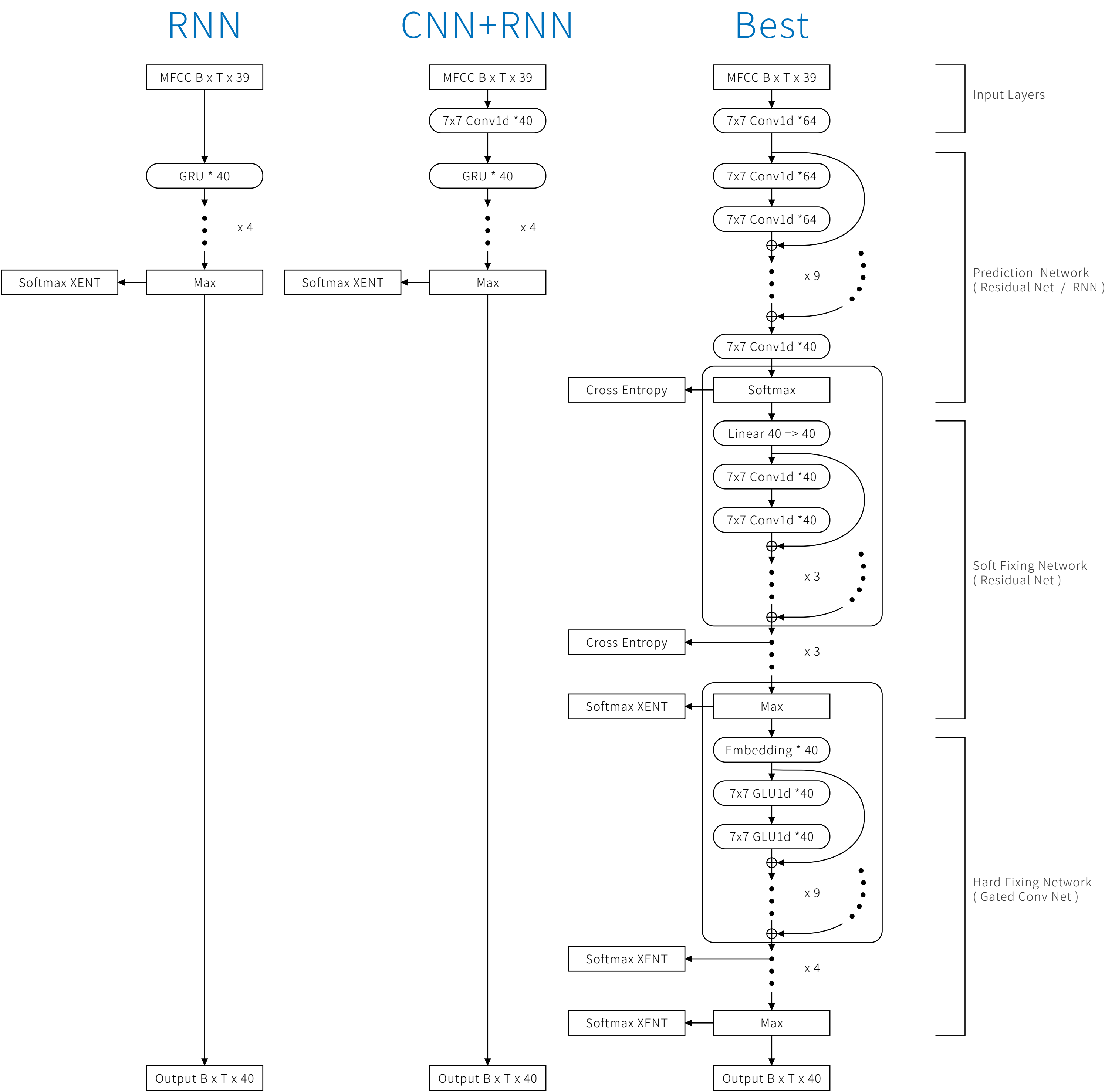


# ADL Homework #1 Report

b04902013 鄧逸軒

## 1. Model Description

Input 使用 MFCC , Output 39 phones + 1 padding , 總共 40 類 . RNN 的模型就是簡單的疊 4 層bi-GRU . 而 RNN + CNN 的模型則是在 RNN 的前面加上一層 Conv1d . 表現最好的模型可以分為四個部份 , Input Layers 負責調整 Channel 數量 , Prediction Net 包含 9 組 ResNetCell , 共 18 層 Conv1d , 後面的兩個Network 用來修正前面的預測 , Soft 和 Hard 的差別在於 Soft 是和 Prediction Network 一起訓練的 , 而 Hard 則是在前面的網路訓練完後 , 拿預測的結果再進行訓練的 .



## 2. How to improve your performance

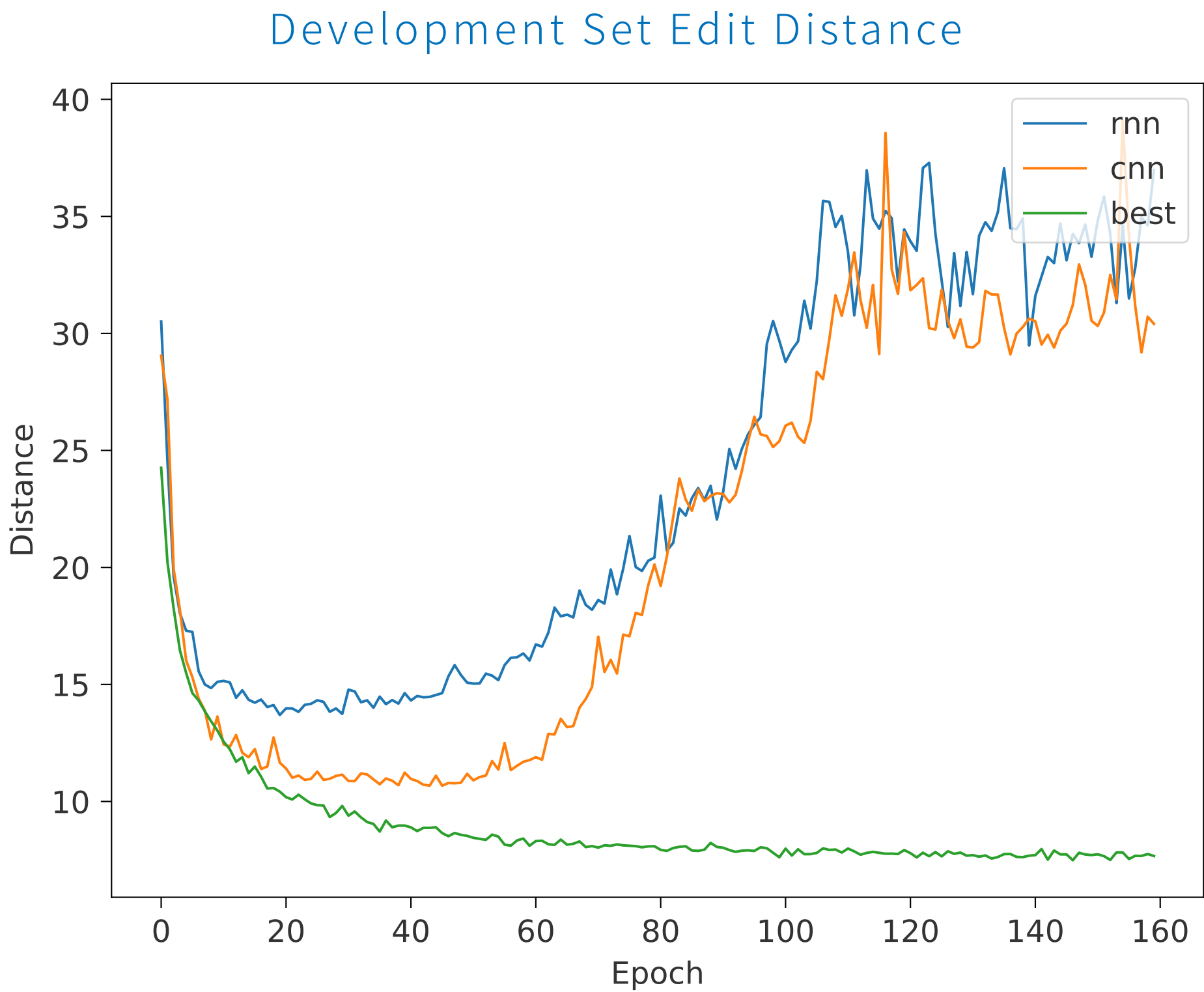
修正網路可以讓 edit distance 下降 2 左右，這個想法是源自於看到 Output 是有著大量的重複，應該可以看附近的 Output 來稍微修正預測錯誤的地方，一開始先做出最下面的 Hard Fixing Network，發現效果不錯之後我讓他重複修四次。之後把他改成 end-to-end 的版本，接在 Prediction Network 上面一起訓練。比較意外的是這樣訓練出來的結果再接上修正網路還是可以下降 1 個 edit distance 左右。

最後在去除重複的部份的時候，我會跳過 Prediction Net 預測的機率沒超過 0.3 的部份。

## 3. Experimental Results and Settings

在 RNN 前面加上一層 CNN 增進蠻多的，我認為是因為 phone 之間比較沒關聯性，而 CNN 的 window 會讓他聚焦在現在的 phone，所以獲得了比較好的效果。

後期 distance 衝上去是因為 RMSProp learning rate 沒調整好，best 是用 Adam 剛好沒遇到這個問題。



RNN Cell 的部份 GRU 和 LSTM 效果差不多，不過 SRU 的表現以及速度都比前兩個好很多，CNN 用 GLU 比用一般的 Conv1d + SELU 效果好一點。

Input Layers 如果換成 Fbank + Conv2d 在 Private Set 和 Development Set 有更好的分數，不過如果全部換成 2D 疊 ResNet-18 的架構反而會變差。

Development Set Edit Distance

Net	LSTM	GRU	SRU	C+RNN	Best	Conv2D	ResNet18
Distance	13.9	13.6	9.3	10.6	6.5	6.3	8.0