數位語音處理 Final Project

程式實作

Speech enhancement using Deep Neural Networks

學生：凌于凱

學號：b04501127

簡介：

在這次實作中，先將 HW2_1 的訓練、測試資料，隨機與 babble、factory1、volvo、white 噪音依照三種訊號雜訊比(-5、0、5db)混合，在透過 Deep Neural Networks，將加入雜訊的訓練資料當作模型輸入，沒有加入雜訊的當作輸出，訓練出一個能夠將雜訊訊號恢復成原始訊號的模型，再透過 HW2_1 訓練出來的模型進行測試，比較加噪音、沒加噪音、經過語音增強的結果，並分析其頻譜圖。

實作方法：[1][2]

Feature extraction

1. 假設一音訊 Y。
2. 將音訊切成一個個音框，且每個音框重疊率為 50%，計算 Discrete time Fourier transform(DFT)得到 Y^f。
3. 求取 Y^f 的 phase 為 $\angle Y^f$。
4. 求取 Y^f 的平方再取對數為 Y^1。

Training stage

1. 假設 noise data 為 Y，clean data 為 X。
2. 進行 Feature extraction 得到 Y^1、X^1。
3. 將 Y^1 做標準化，將其連同左右 N(此次實作取 2)個 frame 一起輸入為 DNN input layer。
4. 再將 X^1 當作 output layer 的輸出。
5. 設計 hidden layer，如 Fig2，此次實作為 3 層 hiddem layer，每層各有 2048 個神經元。
6. 訓練模型，讓 predict 與 X^1 的均方差最小，直到收斂。\

Enhancement stage

1. 假設一 noise data 為 Y。
2. 進行 Feature extraction 得到 Y^1、$\angle Y^f$。
3. 將 Y^1 用剛剛訓練完的 model 測試得到 X^1。
4. 將 X^1 取指數再開根號，並與 $\angle Y^f$ 結合，並計算其 inverse Fourier Transform 為 X^t，即為 Speech enhancement 後的訊號。

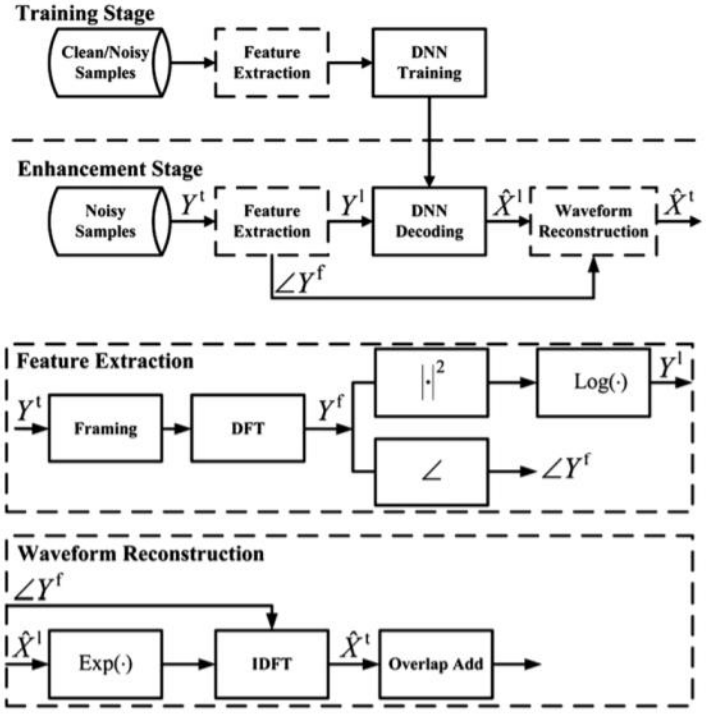


Fig. 1. A block diagram of the proposed DNN-based speech enhancement system.

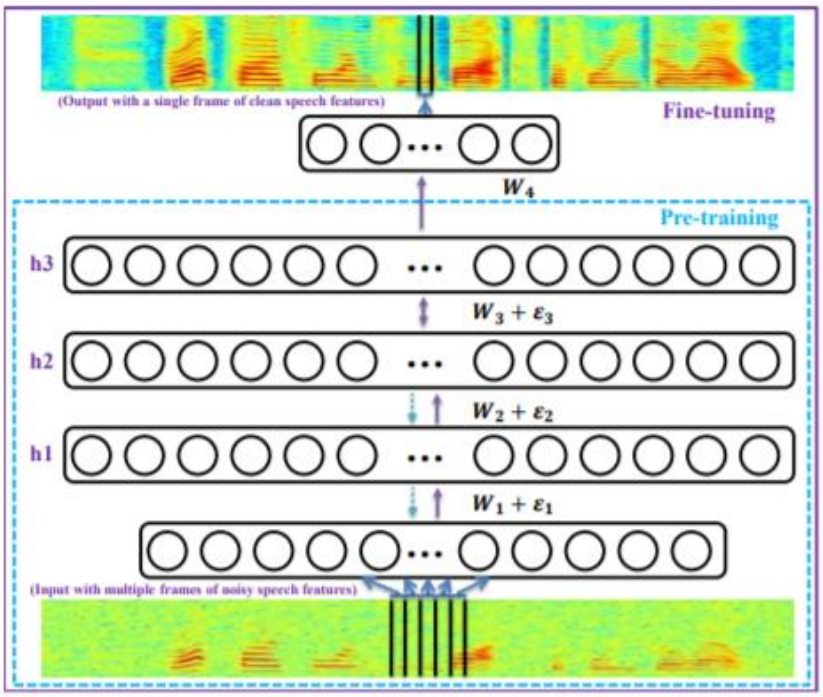


Fig. 2. Illustration of the basic DNN training procedure.

From [1][2]

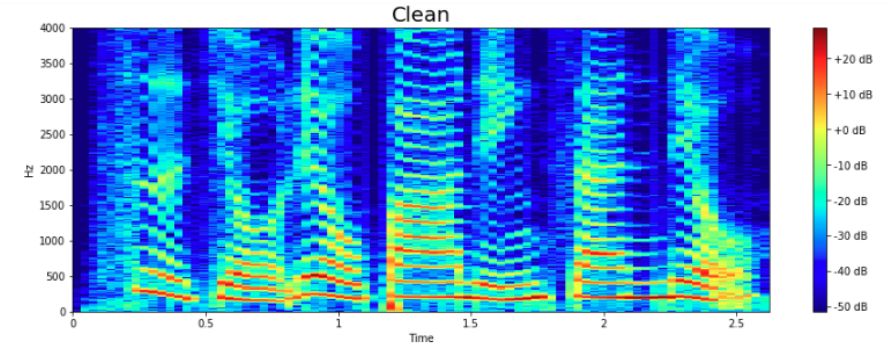
實作結果：

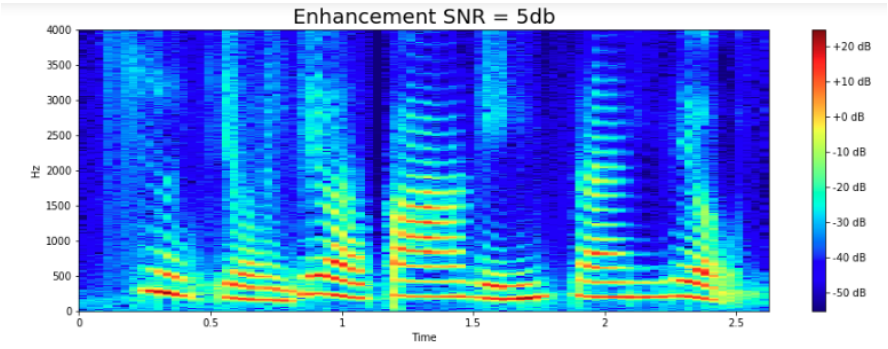
Accuracy(%)

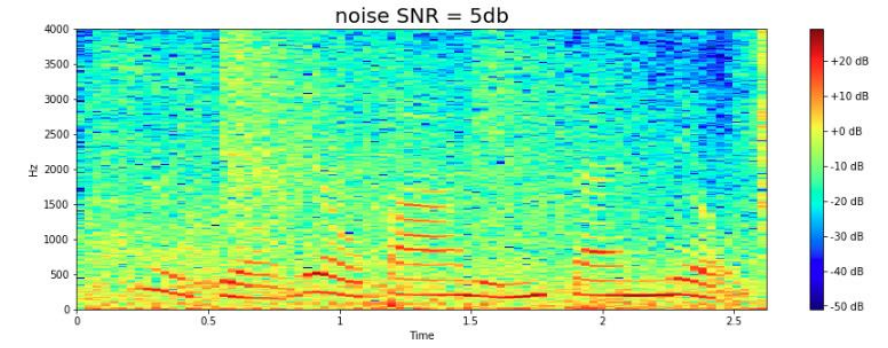
Type \ SNR	-5db	0db	5db	Clean
Noisy	32.11	41.77	58.63	97.12
After enhancement (1 frame as input)	39.59	48.56	60.93	
After enhancement (5 frames as input)	49.88	63.58	75.20	

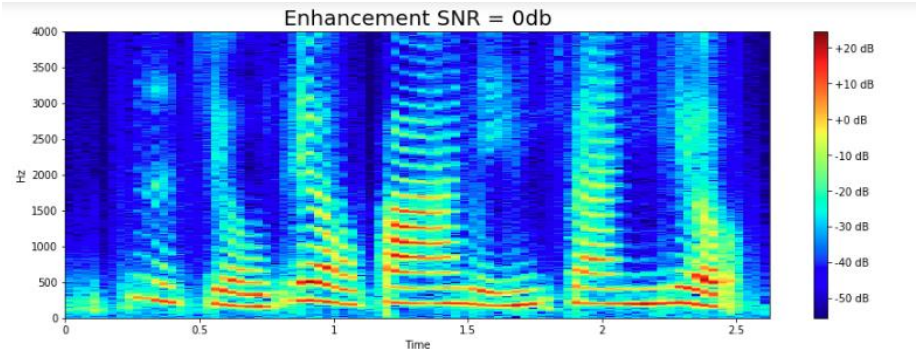
成果大概會比沒做處理的高 20%，但成果還是很普通，且可以發現若只取一個 frame 當 input 的話，成效並不是很好，如果連旁邊幾個 frames 都一起拿進去 train 的話，成效會更好。

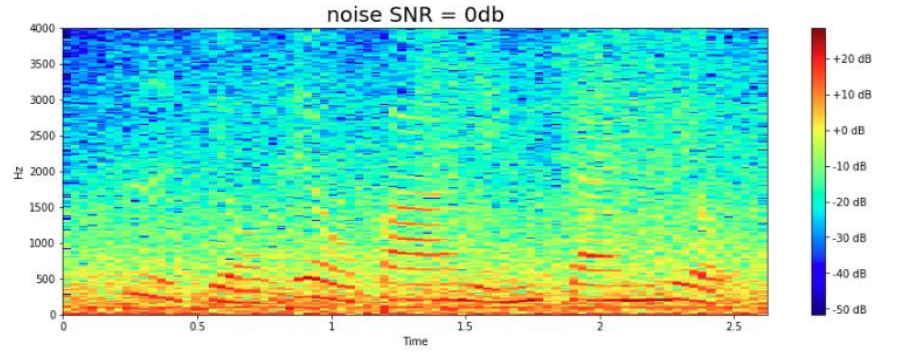
Spectral

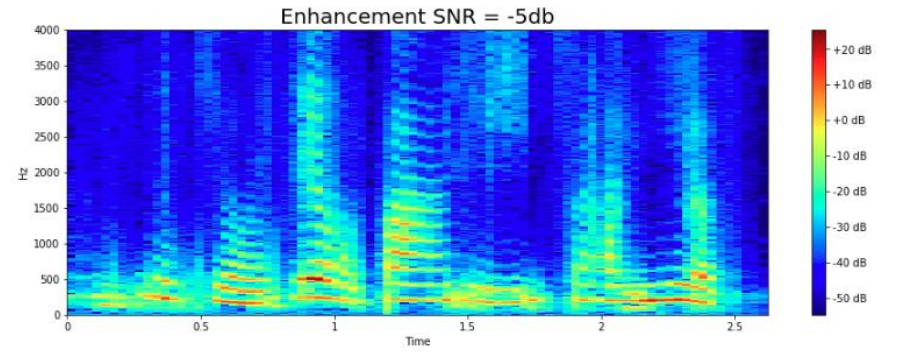


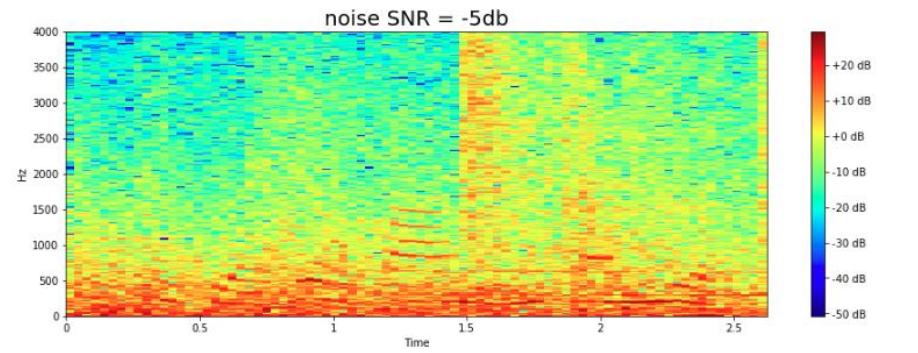












可以發現在高頻中，高能的黃點大概都被過濾掉了，而在最開始與最後的那一條錯誤率比較高，可能是因為 model 是使用左右加起來 5 個 frames 去 train 的，所以在一開始與最後的左右兩邊會沒有周圍的 frame，故沒有取進去，因此最後左右兩側邊緣的結果是使用預測出來左右邊緣值所求得的，左右兩邊會有 3 個頻率高低相等。

結論

雜訊對於語音辨識的影響非常的大，而該如何去濾掉這些雜訊，自古至今，都是很重要的議題，從 Spectral Subtraction 開始有最初 speech enhancement 的觀念，到現在因為硬體進步迅速，許多問題都開始研究如何深度學習的方法解決，在這次期末報告中，閱讀一些相關論文之後，深刻了解這方面的知識，透過實作，也讓我在更加了解如何使用 keras 去 train NN model，以及該如何使用 librosa 套件讀取音訊檔，使用 DFT 將時間軸轉換為空間軸，產生頻譜圖，這些都是很寶貴的經驗。

Reference

[1] A Regression Approach to Speech Enhancement Based on Deep Neural Networks, Yong Xu, Jun Du, Li-Rong Dai, and Chin-Hui Lee

[2] An Experimental Study on Speech Enhancement Based on Deep Neural Networks, Yong Xu, Jun Du, Li-Rong Dai, and Chin-Hui Lee

How to Use

Prerequisites:

Python 3.5+

Keras 2.0.8

Librosa

Scipy

Numpy

matplotlib

Download dataset

Speechdata and hw2_1

$ bash ./get_data.sh

Train model

$ python3 train.py

Generate Enhancement File(wav in ./speechdata/testing_enh_snr*/)

$ python3 test.py

Test HMM

Copy enhancement file(in ./speechdata/testing_enh_snr*/) or noisy file
(in ./speechdata/testing_noise_snr*) to ./hw2_1/speechdata/test, then run

$ 01_run_HCopy.sh

$ 04_testing.sh

Generate spectral(Need enhancement data, noisy data, test data)

$ python3 gen_spectral.py