

DLHLP HW3 Report

組長 Github ID: ChuanYouLin

組員(姓名+學號):

學號：r08944008 姓名：簡 義

學號：r07922104 姓名：林傳祐

學號：r08944024 姓名：陳品媛

學號：r06922089 姓名：邱淳浩

1. (5%)請記錄 evaluate.log 裡面的SiSNR 數值，和當時所用的 hyperparameter(這一題請3-1不用PIT, 3-2用PIT)

data	N	L	B	H	P	X	R	Norm	Causal	Mask nonlinear	SI-SNRi
3-1	128	40	128	256	3	7	2	gLN	0	relu	18.89
3-2	512	16	128	512	3	8	3	gLN	0	relu	10.41

2. (5%)嘗試調整不同的hyperparameter，比較其差異，並試著分析結果 (至少針對2種不同的hyperparameter進行實驗)

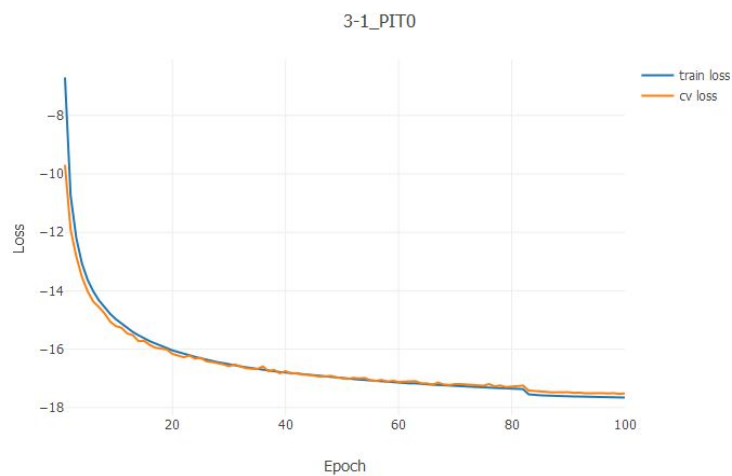
data	pit	N	L	B	H	P	X	R	Norm	Causal	Mask nonlinear	SI-SNRi
3-1	0	128	40	128	256	3	7	2	gLN	0	relu	18.89
3-1	0	32	16	32	64	3	4	2	gLN	0	relu	13.48
3-2	1	128	40	128	256	3	7	2	gLN	0	relu	8.45
3-2	1	256	60	128	256	3	7	2	gLN	0	relu	8.40
3-2	1	256	20	128	256	3	7	2	gLN	0	relu	8.84
3-2	1	512	16	128	512	3	8	3	gLN	0	relu	10.41

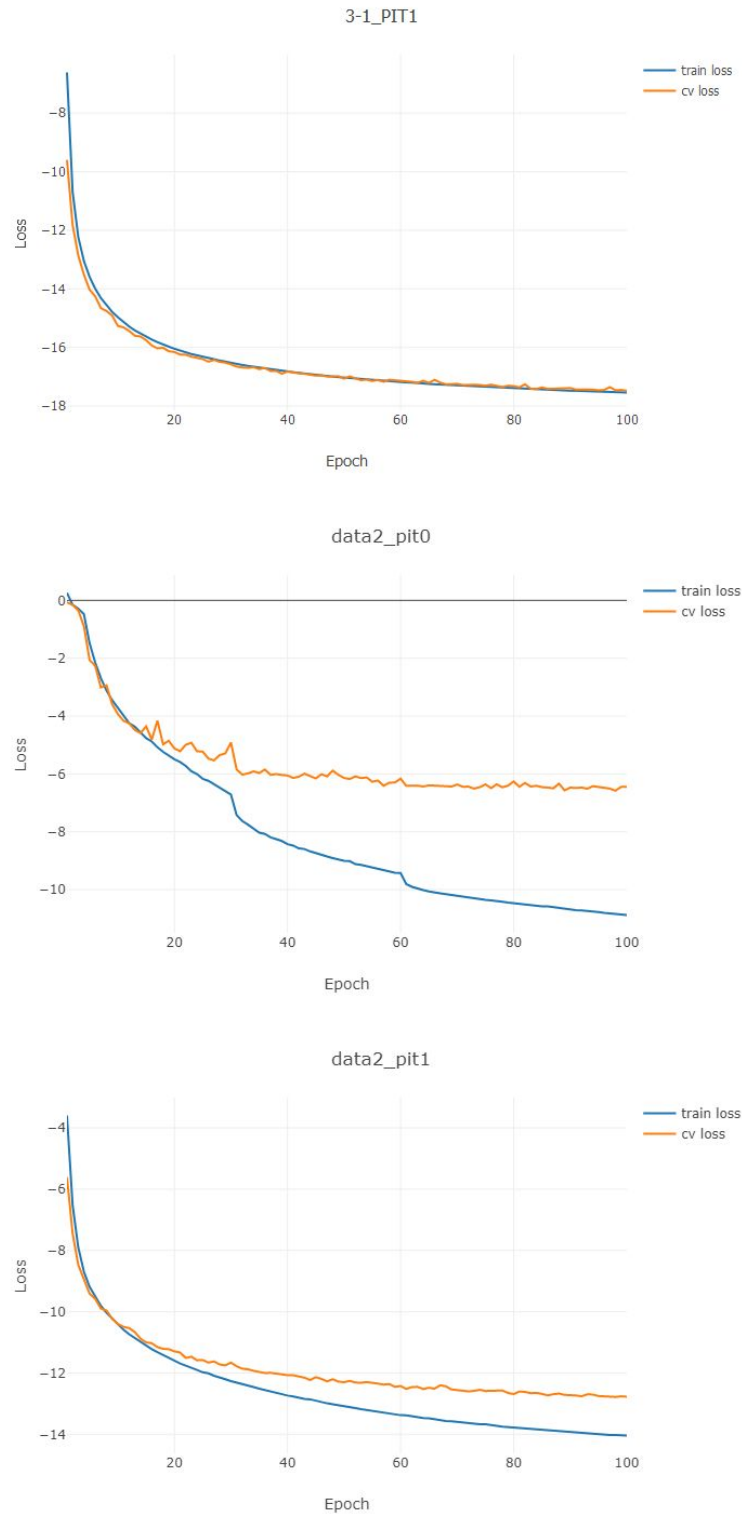
- a. 3-1: 兩個語者的情況下，model只需要少量的參數就能有不錯的結果。

- b. 3-2: 根據實驗的結果發現，encoder 1D convolution的kernel大小(L)越小，對於model分離的效果越好，可能是因為convolution block的dilation設計，就是為了讓model可以接收更大範圍的資訊，因此，encoder使用較小的kernel可以捕捉到較小的變化量。encoder就是想要取代STFT，而一般STFT的window size一般也不會太大，使feature在時間上的sensitive比較高。

3. (3%)3-1, 3-2請分別試看看有無PIT的差異並記錄結果(loss learning curve, Si-SNR)

data	pit	N	L	B	H	P	X	R	Norm	Causal	Mask nonlinear	SI-SNRi
3-1	0	128	40	128	256	3	7	2	gLN	0	relu	17.59
3-1	1	128	40	128	256	3	7	2	gLN	0	relu	17.52
3-2	0	128	40	128	256	3	7	2	gLN	0	relu	1.62
3-2	1	128	40	128	256	3	7	2	gLN	0	relu	8.45





4. (2%) 思考一下為何有無PIT會影響3-1, 3-2的結果並寫下你的看法
- a. 3-1: 因為只有兩個語者, model只需要固定第一個分割的結果是speaker A的聲音、第二個分割的結果是speaker B的聲音, 然

後硬去fit，就能有好的結果。因此有沒有使用pit，結果相差不大。

使用pit的情況下，model要不停地交換兩個語者的label，可能是使結果比較差一點點的原因。

- b. 3-2: 在3-2的音檔有多個語者中選兩個語者來混音，而這個會有一個問題就是在於擺ground truth的順序會不固定，例如：speaker A, speaker B，而另一種排序可能是speaker B, speaker C，在這個例子中可以發現到speaker B的擺放位置不固定，所以透過PIT計算模型輸出與各種ground truth組合的loss，並以最低的loss來做backward，讓模型自己學習。

5. bonus(2%)：

請自己找兩段音訊合起來(請不要使用作業給的data)測看看是否能成功分離，上傳音訊(含原音檔、合成後音檔及經過model分離的音檔)，紀錄Si-SNR於report中，並給出至少一種improve Si-SNR的方法(調參數除外)。

改使用Dual-path RNN,

ref: <https://github.com/JusperLee/Dual-Path-RNN-Pytorch>

Test data of 3-2	
Method	SI-SNRi
Conv-Tasnet	10.41
DPRNN-Tasnet	12.06

自製音檔 (bert + ernie)	
Method	SI-SNRi
Conv-Tasnet	-1.23
DPRNN-Tasnet	-0.25

