

DLHLP HW3 Report

組長 Github ID: shuoenchang

組員(姓名+學號):

學號:R08922A02 系級: AI所碩一 姓名:張碩恩

學號:R08922130 系級: 資工所碩一 姓名:丁杰

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)

3-1.

```
N=128
L=30
B=128
H=256
P=3
X=7
R=2
norm_type=gLN
causal=0
mask_nonlinear='relu'
C=2
```

Average SDR improvement: 19.96
Average SiSNR improvement: 19.82

3-2.

```
N=512
L=40
B=256
H=256
P=3
X=7
R=3
norm_type=gLN
causal=1
mask_nonlinear='relu'
C=2
```

Average SDR improvement: 10.71
Average SiSNR improvement: 10.14

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)

以下實驗 norm_type=gLN, mask_nonlinear=relu, C=2

3-1.

在下面的表格中，部份參數是固定的(N=128, B =128, H=256, P=3, X=7, causal=0)，其餘參數如下表格中的L, R是有變動的。

	original	setting1	setting2 (best)
L	40	30	30
R	1	1	2
SISNR	17.69	17.86	19.82

L代表了 Length of the filters，變短之後結果有微幅度的提昇，我想這可能是因為訓練時模型可以更加注意資料短時間內的關聯性，若是有足夠的filter，反而可以讓模型學得更好。這與論文中提到的相符合 (“Shorter segment length consistently improves performance.”)

R 代表了 Number of repeats，調大後 performance 增加不少，因為參數比原本還多，能夠預測的能力也相對變強了。而增加 repeats，就像是在 deep learning 中增加層數一樣的意義。

3-2.

	best	setting1	setting2	setting3	setting4
N	512	512	512	512	128
L	40	40	40	40	40
B	256	128	256	256	128
H	256	256	256	256	256
P	3	3	3	3	3
X	7	7	7	7	7

R	3	3	2	2	2
Causal	0	0	0	1	0
SISNR	10.14	9.68	9.23	8.90	8.79

N 代表 Number of filters in autoencoder，增加他能讓模型 increases the overcompleteness of the basis signals，因此帶來比較好的成果。

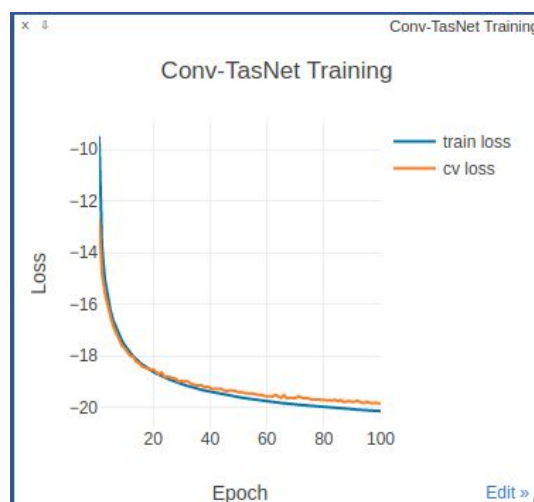
B 代表 Number of channels in bottleneck，增加他讓模型的參數量提高，也帶來比較好的成功，比較有趣的是論文中有提到 $H/B=5$ 時可以帶來最好的結果。

R 代表了 Number of repeats，調大後 performance 增加不少，因為參數比原本還多，能夠預測的能力也相對變強了。而增加 repeats，就像是在 deep learning 中增加層數一樣的意義。

Causal 代表使用程式使用 causal configuration，但結果比使用 Dilated Network 差，根據論文提到的，這或許是因為 causal convolution 和 layer normalization 這些運算造成的。

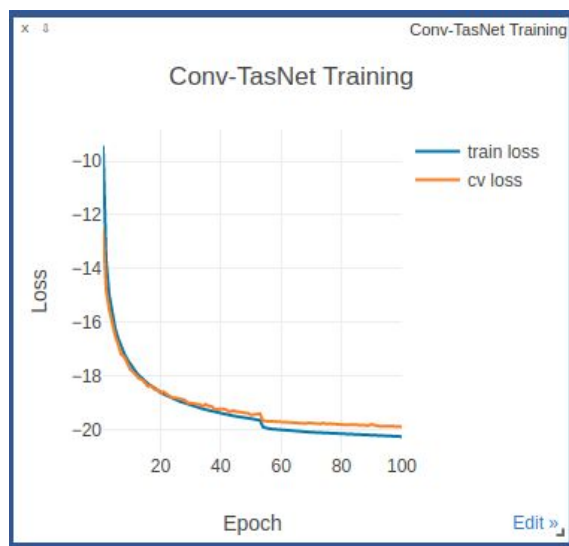
3. (3%)3-1, 3-2請分別試看看有無PIT的差異並記錄結果(loss learning curve, Si-SNR)

3-1 沒有PIT



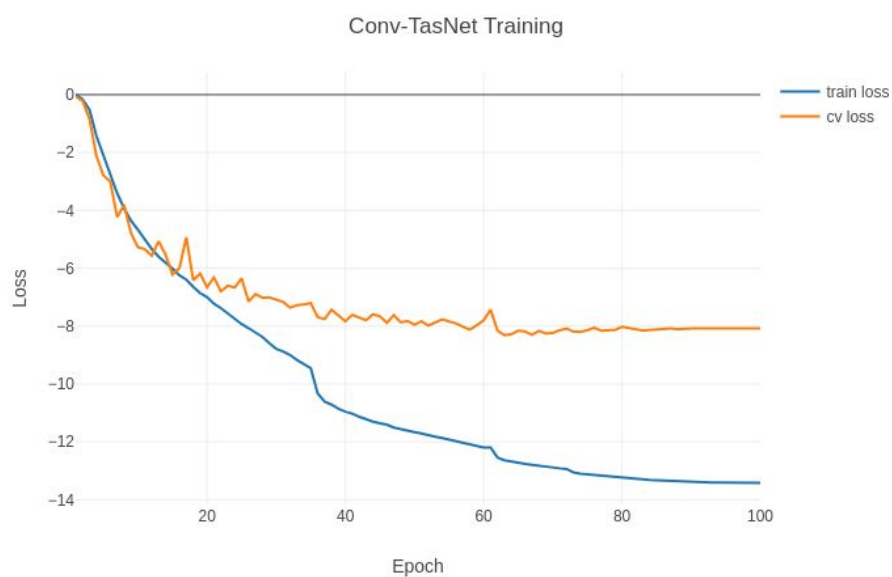
Average SDR improvement: 19.96
Average SISNR improvement: 19.82

3-1 有PIT



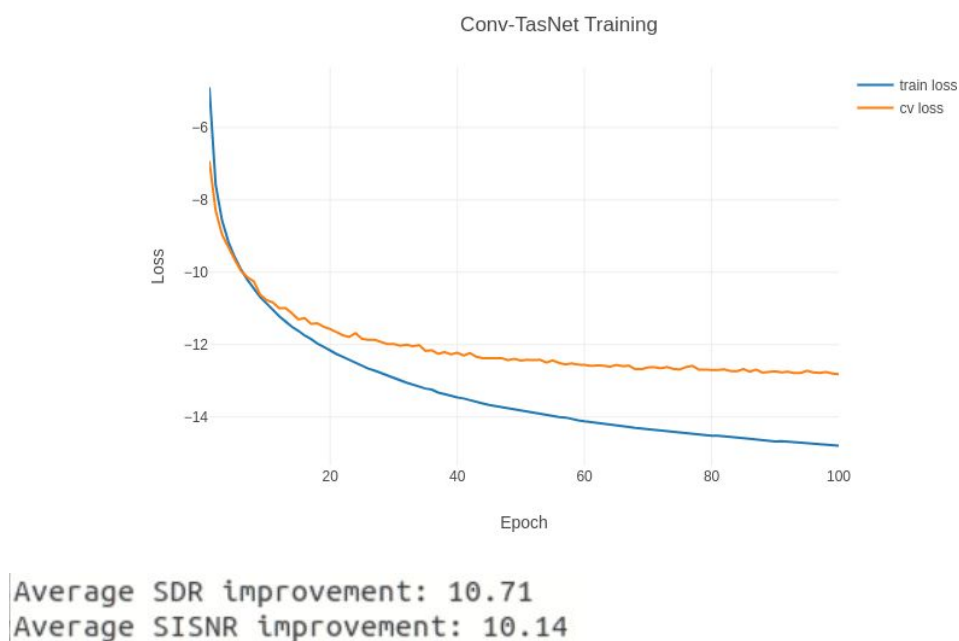
Average SDR improvement: 20.34
Average SISNR improvement: 20.21

3-2 沒有PIT



Average SDR improvement: 8.27
Average SISNR improvement: 4.52

3-2 有PIT



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

PIT 的運作是在 training 時，主要方式是透過 PIT 交換 output 去找到更好的 loss。在 3-1 中，只有兩個語者，且兩個語者皆是固定的人，因此有無 PIT 相較起來並沒有太大的差別。然而其中仍然存在着微小差異，而這些微小的差別可能來自於模型本身靠着交換語者，自己找到了更好的 mask，這狀況下的模型是 speaker independent 的（剛好在 epoch=50 附近時找到更好的，所以 loss 在此時突然降的更低了）。

在 3-2 中，我們就可以很明顯看出 PIT 的重要性，在多重語者的情況下，兩個語者各自代表的人不斷在改變。缺少了 PIT 時，不斷在改變模型的收斂變很慢，而且 cv loss 也與 train loss 越差越大。這就是因為模型變得只會判斷特定的狀況下的 Separation，而無法適用在任何情況。

總結來說，缺少了交換 output 去找到更好 loss 的過程，讓模型無法確實區分出 speaker 1 和 speaker 2，只要稍微對調一下，或是讓他變成不同人，就會讓模型結果變很差，而且此時的 speaker 必須固定不動。若要改善這個情況就要靠 PIT 的協助。