

組長 GitHub ID : ming024

組員

- r08922080 資工所碩一 簡仲明
- r08921062 電機所碩一 黃健祐
- b04501127 土木四 凌于凱

1. 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)

part	SiSNR	N	L	B	H	P	X	R	norm type	causal	mask	C
3 - 1	19.36	128	40	128	256	3	7	2	gLN	0	ReLU	2
3 - 2	10.10	128	40	128	144	3	8	4	gLN	0	ReLU	2

2. 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)

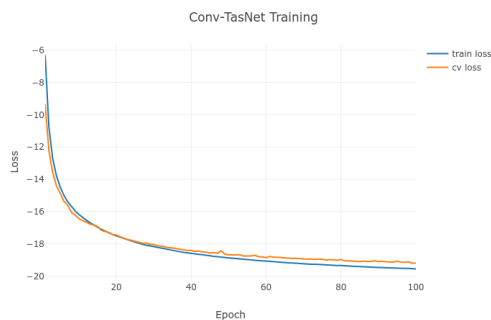
以上題的參數為基礎，並全部使用 PIT，對不同的 hyperparameter 進行調整：

	3 - 1			3 - 2		
causal	0	0	1	0	0	1
R	2	1	2	4	2	4
SiSNR	19.02	17.42	17.32	10.10	8.40	8.01

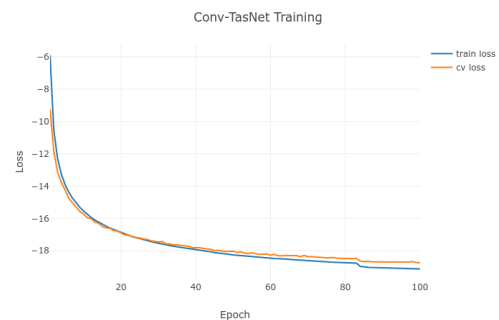
由表格可知，causal 的 model 表現會比 non-causal 的差一些，因為 non-causal 的 model 在進行 separation 時能夠使用整段訊號的資訊，但若達到 real-time 的話，causal 的 model 是比較符合實際應用的。

R 值的大小與 model 的表現有顯著的關係：repeat 的次數越多，分離出來的音訊品質就越高。但 R 越大時，訓練所需的資源就會快速增長，也會讓 inference 的時間變長，不利於 real-time 的應用。

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)



(a) 3-1 without PIT (SiSNR = 19.36)



(b) 3-1 with PIT (SiSNR = 19.02)

由 (a)、(b) 兩圖可知，有無 PIT 對 3 - 1 的影響並沒有很大。

4. 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。

在 3 - 1 中，因為只有兩個 speakers，所以就算不用 PIT，model 也可以使兩個 outputs 分別固定分離出某個 speaker 的 utterances；而在 3 - 2 中，因為有多個 speakers，所以沒有辦法強迫 model 讓某個 output 固定輸出某個特定 speaker。