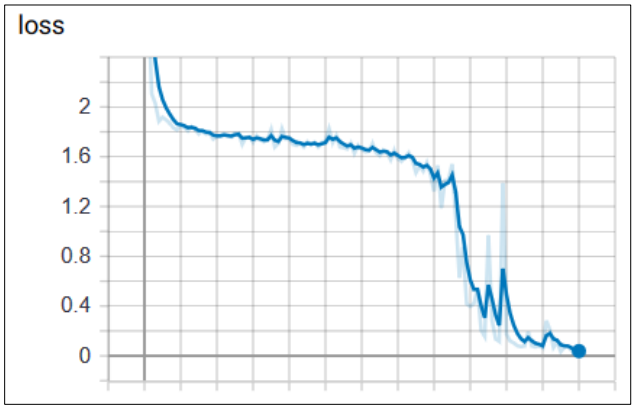
Homework 1 - End-to-end Speech Recognition

學號：r08922067 系級：資工所碩二 姓名：鄭淵仁

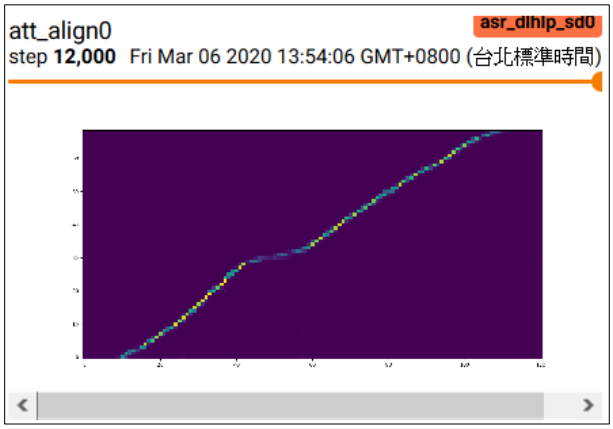
（我忘記 kaggle deadline 比 report deadline 早，所以來不及傳 kaggle QQ，所以就沒有 reproduce 了）

1. (2%) Train a seq2seq attention-based ASR model. Paste the learning curve and alignment plot from tensorboard. Report the CER/WER of dev set and kaggle score of testing set.

- learning curve:



- alignment plot:



- CER/WER of dev set

Error Rate (%)	Mean
Character	3.1984
Word	10.5168

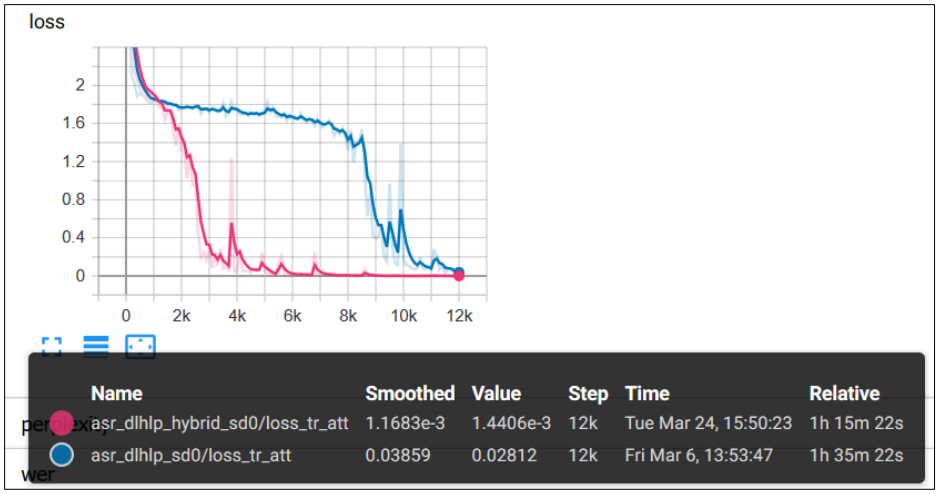
- Kaggle score

Submission and Description	Private Score	Public Score
answer.csv 26 minutes ago by YJC	2.17600	2.09600

2. (2%) Repeat 1. by training a joint CTC-attention ASR model (decoding with seq2seq decoder). Which model converges faster? Explain why.

- Which converges faster?

As the following loss curve shows, joint CTC-attention ASR converges faster.



(Red line: CTC-attention ASR / blue line: only attention ASR model)

- Why?

相較於 encoder 要先接到 decoder 再 output 的第 1 題，第 2 題的 encoder 必須要同時能直接 output 出來做為 CTC 的訓練目標，因此 encoder 會收斂得比較穩定。而當 encoder 比較穩定，decoder 也會因此收斂得比較好、快。

3. (2%) Use the model in 2. to decode only in CTC (ctc_weight=1.0). Report the CER/WER of dev set and kaggle score of testing set. Which model performs better in 1. 2. 3.? Explain why.

- Which performs better?

As the following table shows, **Model 2** performs the best in Model 1., 2., 3.

	Dev		Kaggle	
Model	**char error rate**	**word error rate**	**public**	**private**
1	3.1984	10.5168	2.17600	2.09600
2	**2.1770**	**7.3222**	**1.29800**	**1.28200**
3	2.8411	9.7304	1.58400	1.53400

- Why?

首先，Model 2、3 裡面的 encoder 必須同時能直接 output 出來做為 CTC 的訓練目標，所以 encoder 收斂得比較穩定，也因此 decoder 也會收斂得比較穩定，所以 performance 很容易會比 Model 1 好。

其次，Model 2 比 Model 3 多了一個 decoder，模型複雜度比較高，所以比較可以去 fit 一些比較複雜的 pattern，所以 Model 2 會表現得比 Model 3 好。

4. (2%) Train an external language model. Use it to help the model in 1. to decode. Report the CER/WER of dev set and kaggle score of testing set.

- CER/WER of dev set

Error Rate (%)	Mean
Character	2.8485
Word	9.2125

- Kaggle score

Submission and Description	Private Score	Public Score
lm_2.csv a minute ago by YJC	2.04199	2.02999

5. (2%) Try decoding the model in 4. with different beam size (e.g. 2, 5, 10, 20). Which beam size is the best?

As the following table shows:

- In development sets, beam size = 10 is the best.
- In Kaggle, beam size = 10 and 20 is the best in public leaderboard;
- In Kaggle, beam size = 20 is the best in private leaderboard.

Beam	Dev		Kaggle	
Size	char error rate	word error rate	public	private
2	2.8485	9.2125	2.04199	2.02999
5	2.7660	9.0170	1.92800	1.84200
10	**2.7631**	**8.9973**	**1.90800**	1.84200
20	2.7642	9.0019	**1.90800**	**1.84000**

Bonus: (1%)

Nothing.