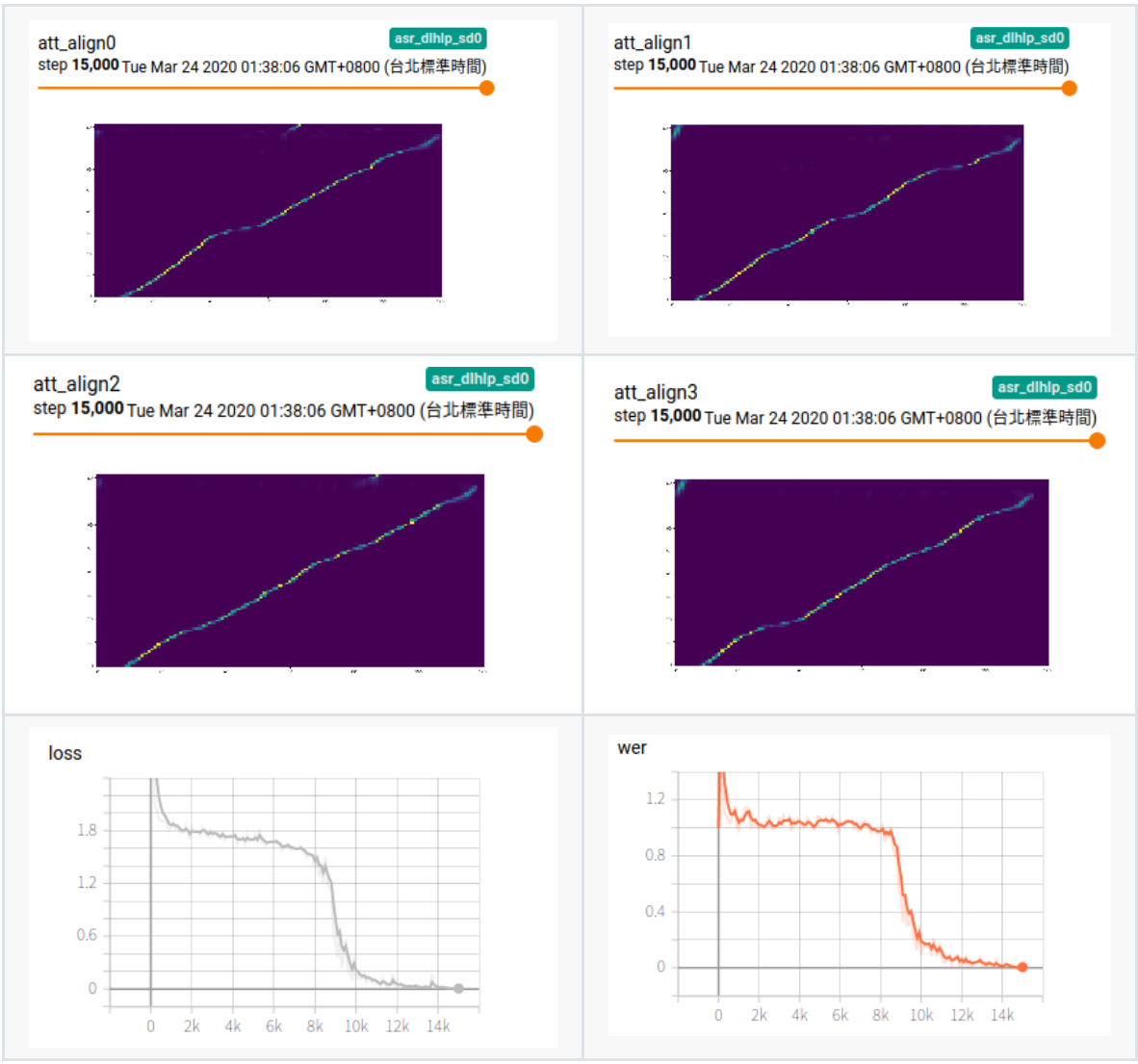
Homework 1 - End-to-end Speech Recognition

學號：b05902111 系級： 資工四 姓名：婁敦傑

學號：b05902010 系級： 資工四 姓名：張頌平

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.

- 參數設定： max step = 15001, valid step = 500



CER	WER	Kaggle public / private
2.5708	8.4902	1.638 / 1.644

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

- Joint CTC-attention ASR model 收斂的比較快
- seq2seq attention-based ASR model有時會因為句子太長或句子含有noise，造成alignment錯誤，而錯誤的alignment，就會使model 學錯。加入CTC，根據forward-backward algorithm 計算CTC loss 改善alignment錯誤的問題，改善model的預測，因此訓練上收斂的也會較快。

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.

- Joint CTC-attention ASR model 表現最好
- Joint CTC-attention ASR model (decoding with seq2seq decoder) 會比seq2seq attention-based ASR model好，是因為改善alignment的問題，而Joint CTC-attention ASR model (decoding only in CTC)表現不好，第一是訓練時ctc_weight = 0.25，預測時的ctc_weight = 1.0，兩者並不一致，第二是沒有加入beam search 而是選擇 greedy decode。

CER	WER	Kaggle public / private
6.4412	22.2437	4.034 / 3.810

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.

- 參數設定： lm_max step = 20000, lm_valid step = 250, lm_weight = 0.5

CER	WER	Kaggle public / private
2.2112	7.1366	1.358 / 1.392

5. (2%) Try decoding the model in 4. with different beam size (e.g. 2, 5, 10, 20, 50). Which beam size is the best?

- beam size = 20 表現最好

beam size	CER	WER	Kaggle public / private
2	2.2112	7.1366	1.358 / 1.392
5	2.1153	6.8449	1.304 / 1.296
10	2.1015	6.7979	1.298 / 1.290
20	2.0915	6.7738	1.292 / 1.282

Bonus: (1%)

Nothing