

DL hw4 part2

alexander.telepov

May 2021

1 Bugs

Some parts quite tricky for efficient implementation like *src_mask*, *src_padding_mask* construction or Gaussian upsampling. Adding *< bos >*, *< eos >* symbols into account and properly handle padding (using it in attention masking and loss calculation) also force me to go throw some bugs.

2 Story of training runs

Longest run was on Collab for 3 hours but it stacked not logging. Due to time limits and poor usability of Collab I failed to provide sufficient amount of experiments. Also I test my code for training on one sample: spectrogram looks more or less similar to groundtruth and audio too. Plots below provided for such case. Other runs on Collab were launched with bugs/ or freeze not logging anything. In total I have launched 5 runs with working code. I haven't find out which tricks are important but implement all tips from notebook.



Figure 1: Train spectrogram

train_loss

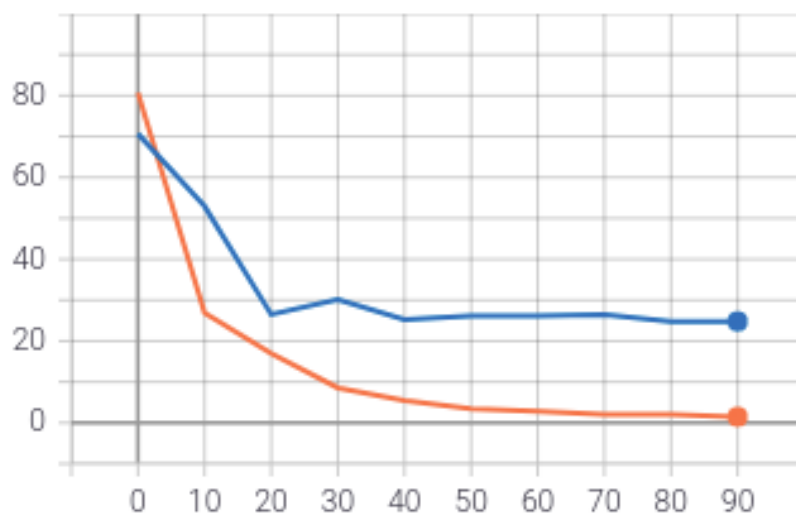


Figure 2: Train loss



Figure 3: Val spectrogram

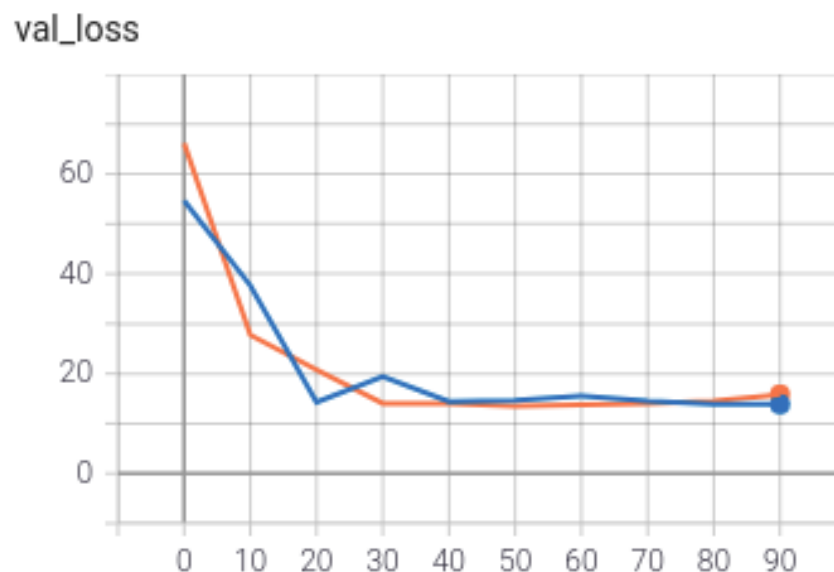


Figure 4: Val loss

3 Audio quality

I can't obtain any meaningful quality. I'm not sure that best model could be improved if trained for longer, because obtained quality is poor and these may be because of bug in implementation or bad hyperparameters.

4 Hardest things

The hardest thing for me in this assignment was to get familiar with text problems (with melgrams, special tokens, architectures to process this type of data). Also I faces with a lot of Collab problems. Many times Collab just freeze dumping nothing to log files for a hour during training. There were other bugs with data: loaded and unpacked dataset for some reason were corrupted by next launch, several times. Unpacking dataset also freeze several times and fall down after 40 minutes. Other challenge which ai faces with is quite time consuming training (10 epoch - 1.5h).

5 What I learned

These task contain quite a lot of coding exercises, these definitely improve my coding skills. Also these task force me to get familiar with text processing and attention mechanism.

6 Questions this task has provoked to me

I'm wondering if real papers really don't describe masking mechanism in attention. For which reason only *old_sequence_len*-axis is masked? Also I'm interested why pointwise losses are used for training and duration model trained separately.

7 How would I improve this assignment

Dataset size decreasing may facilitate training on Collab, also more detailed instructions can be useful with small time limits.