

Отчет по TTS

Барановская Дарья

Основной целью задания было реализовать и обучить FastSpeech. Обучение оказалось достаточно сложно, модель сильно переобучается. Для лучшего обучения в конце я добавляла attention маски, но они обучились так хорошо как хотелось бы за 60к итераций, поэтому для финальной модели будет предоставлен код без масок (изменения откачены, если нужно, их можно найти в комментах гита).

Из экспериментов я пробовала менять число fft блоков, старательно подбирала lr, scheduler, dropout и batch_size. Остальное по максимуму старалась брать из статьи.

В финальной модели использовались следующие параметры:

batch_size вначале 3, потом 10

class ModelConfig:

 vocab_size: int = 57

 hidden_size: int = 384

 hidden_size_fft: int = 1536

 num_heads: int = 2

 kernel_size: int = 3

 n_fft_blocks: int = 6

 dropout = 0.1

Время обучения на датасфере: 14 часов.

Код: https://github.com/darya-baranovskaya/TTS_with_FastSpeech.git

Wandb репорт с зафиксированными аудио: https://wandb.ai/darya_baranovskaya/uncategorized/reports/HW3-Report--VmlldzoxMzA4ODc0

Wandb runs: https://wandb.ai/darya_baranovskaya/uncategorized?workspace=user-darya_baranovskaya