Отчет по третьему домашнему заданию по курсу “Глубинное обучение в обработке звука”

Выполнил: Михаил Малафеев

В ходе третьего домашнего задания были реализованы модель на основе статьи FastSpeech и попытки подбора гиперпараметров для ускорения обучения. Во-первых, в силу малого количества времени для обучения было решено не использовать Dropout с предположением, что я даже переобучиться не успею за ~30 часов. Судя по экспериментам я не успел дойти до стадии переобучения, поскольку на мой личный вкус качество аудио со временем только становилось лучше. Наилучший результат был достигнут на гиперпараметрах из статьи и с расписанием CosineAnnealingLR step_size=2000 и min_lr=1e-7.[wandb1, wandb2, wandb3] Были попытки увеличения kernel_size для FFTBlock’a. Это привело к тому, что модель училась сильно быстрее, но была более шумной (и в конце улетела в космос) [wandb]. Скорее всего, модели проще было шуметь в такой конфигурации, поскольку явно собиралась информация о далеких буквах, если посмотреть на среднее предсказание aligner’a. Честно говоря, поставь модель на еще пару суток, я имел получить достаточно неплохие озвучивания, если судить по текущей динамике обучения. Также, обычное StepLR расписание затрудняло обучение и циклическое расписание было важным при длительном обучении модели (кажется для звука это правило). Недостатком модели получилось то, что модель генерировала достаточно тихий, распознаваемый звук, но достаточно зашумленный. За первое время модель обучилась произносить слова с очень большим шумом и дальше только уменьшала его.

Для воспроизведения можно воспользоваться ноутбуком в репозитории и запустить все в коллабе или следовать ReadMe.