

Модель

Генератор

Тут я сделал почти всё как в статье. В конфиге можно менять количество всяких слоёв, hidden size и всё такое. Обучал я максимально облегченную модель, чтобы она влезла в память Colab Pro, там в основном дают P100 на сутки непрерывно. Я заметил, что в моём случае выгоднее увеличивать hidden size, а не количество слоев. Модель даже заработала гораздо лучше, чем я ожидал.

Дискриминатор и функции потерь

1. Mel loss: тут я перегонял ground truth wav и fake wav в мелы и считал l1 loss с коэффициентом 45, как в статье.
2. Adversarial loss: эту часть я довольно сильно упростил. Я использовал только один дискриминатор. Это свёрточная сеть, в конце я беру средний результат от каждого семпла – это и есть вердикт. Также дискриминаторе шаг я делаю не всегда, если ошибка и так очень низкая, я не трогаю веса. В той модели, что я сдаю, это не очень пригождается, но помогает локально, где я беру очень слабый генератор, чтобы он влезал в 4 ГБ видеопамати.

Графики и аудио по ходу обучения

С ними можно ознакомиться здесь: https://wandb.ai/s_isaev/hfg/runs/221g7p3w?workspace=user-s_isaev

Воспроизводимость (обучение и inference)

Можно ознакомиться в README.txt

Аудио для MOS

Отдельно приложил в anytask.