

Отчет по HiFiGAN

Логи Wandb -

https://wandb.ai/pos/hifi_project/runs/1tj2gt69?workspace=user-pos

Реализована модель HiFiGAN V3.

Изначальной проблемой была нехватка параметров в самой статье, пришлось заимствовать (воровать) из репозитория <https://github.com/jik876/hifi-gan> следующие вещи:

- slope у LeakyReLU = 0.1
- Для MSD параметры в репозитории немного отличались от параметров в исходной статье про [MelGAN](#), поэтому для потенциально лучшего качества взял модифицированные параметры из репозитория

Все остальное реализовано в соответствии со статьей.

Завелось почти сразу, но была проблема с фичурайзером - в статье этого не описано, но в репозитории обучаются не на полных звуковых дорожках, а на кусках длиной 8192, что довольно неплохо с точки зрения баланса качества обучения и экономии ресурсов. Чтобы размеры изначальной вавки и выходной из генератора были равны (и не надо было костыльно обрезать входные и выходные данные), пришлось добавить пару строчек в MelSpectrogram (относительно прошлой дз), чтобы поддерживать нормальный размер при конвертации входной вавки в спектрограмму.

Из интересных фактов, изначально хотел завести конфигурацию V1 (делал ее для чекпоинта), но решил что V3 должна учиться быстрее (разницы как оказалось почти никакой) и выбрал ее. Итоговая модель - 165957 итерация обучения.

Запуск на тестовых данных и на валидационных в ноутбуке Hifigan_test. (а вообще в wandb логировались тестовые данные каждые 200 итераций)

Итог

В отчет больше добавить нечего, курс отличный, состояние после последних 2 месяцев фулл-тайм работы и учебы на МОПе примерно такое:

