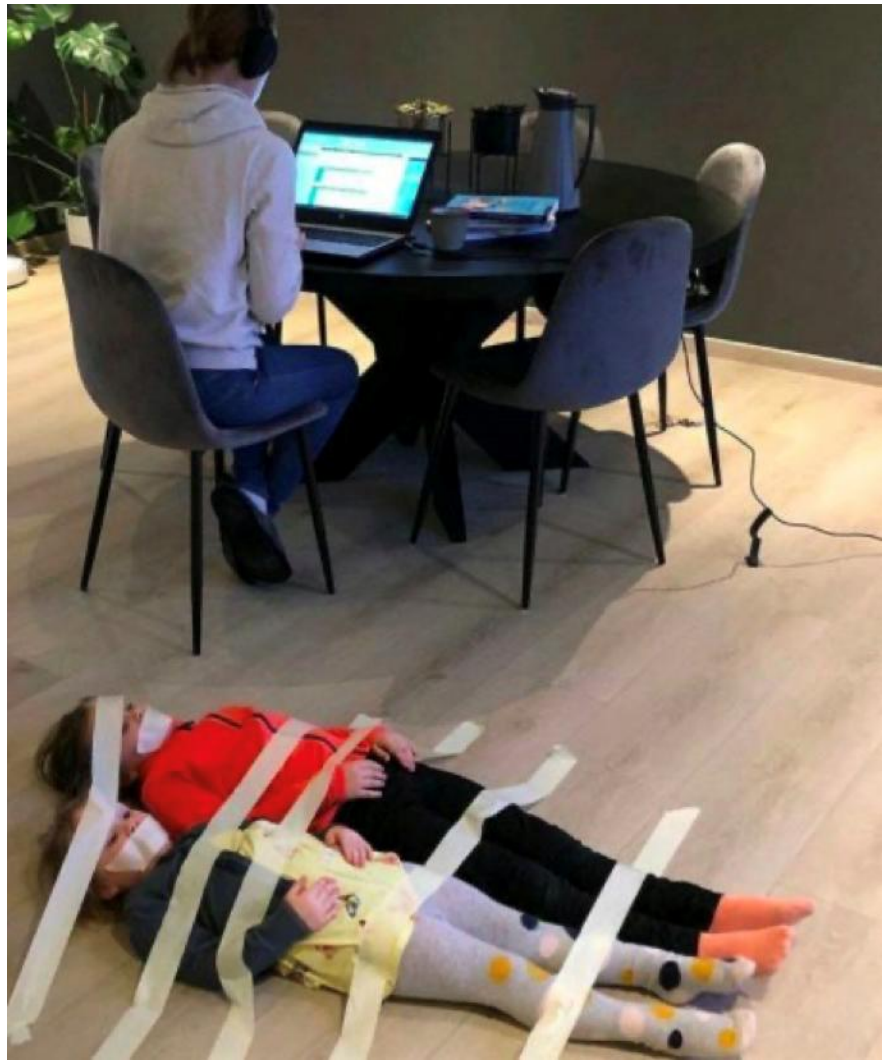
ПОДАВЛЕНИЕ ПОСТОРОННИХ ГОЛОСОВ И ШУМОВ В АУДИОКОНФЕРЕНЦИЯХ

Крамаренко Георгий

Проблема и актуальность

- Аудиоконференции стали частью повседневной жизни
- Качество связи портят фоновая речь и шум
- Это мешает сосредоточиться и утомляет



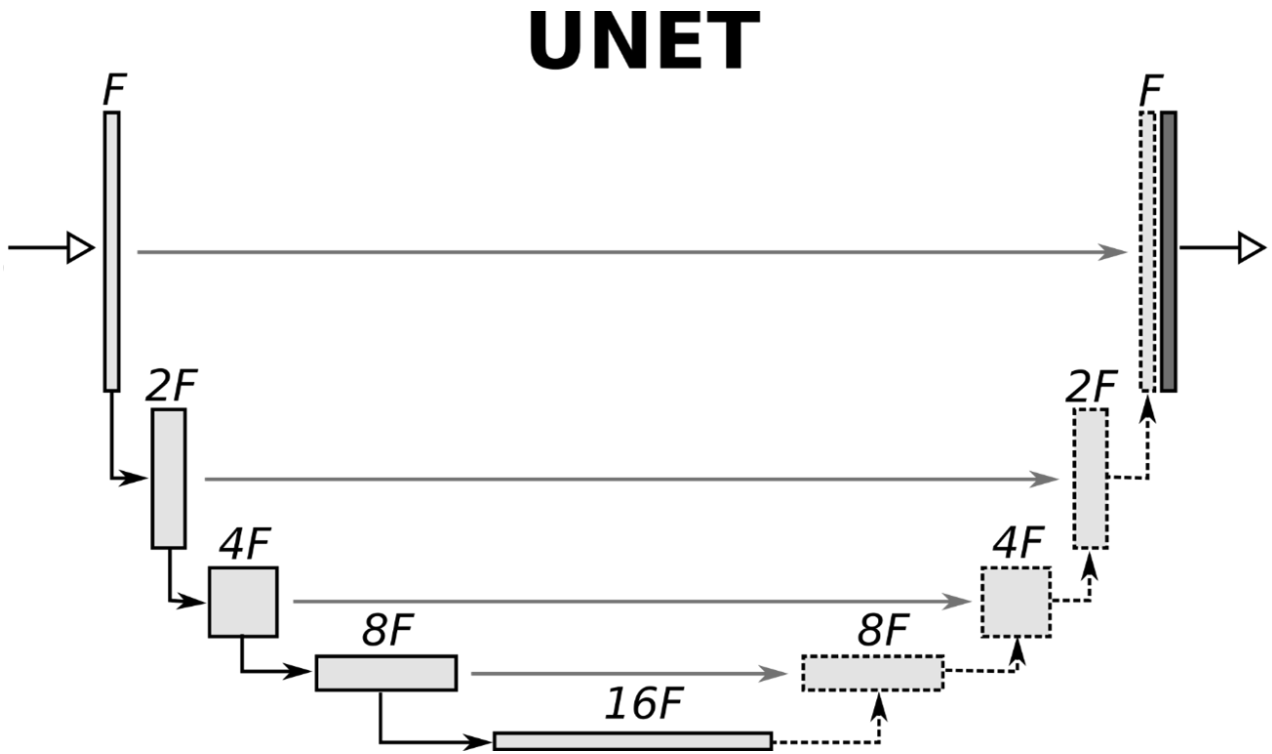
Постановка задачи

Реализовать алгоритм:

- удаляющий посторонние голоса и шумы.
- работающий в реальном времени.
- оценка качества по метрике SNR (signal to noise ratio).

$$(SNR)_{\mathrm{dB}} = 10\log_{10}\frac{\mathrm{signal\ power}}{\mathrm{noise\ power}}$$

Предлагаемое решение



DPCCN: Densely-Connected Pyramid Complex Convolutional Network

Предлагаемое решение

Я провел анализ архитектуры модели и выявил, что в есть модули, не позволяющие модели работать в реальном времени.

Они требуют информацию из будущего.

- Свёртки
- InstanceNorm
- Pyramid pooling

Предлагаемое решение

Я выбрал такие модификации модулей, реализовал их и интегрировал в модель.

- Свёртки → Свёртки с паддингом
- InstanceNorm → LayerNorm, Causal InstanceNorm
- Pyramid pooling → Causal Pyramid pooling

Предлагаемое решение

Данные для обучения:

Голос 1



Голос 2



Шум



Пример голоса 1



Смесь аудио

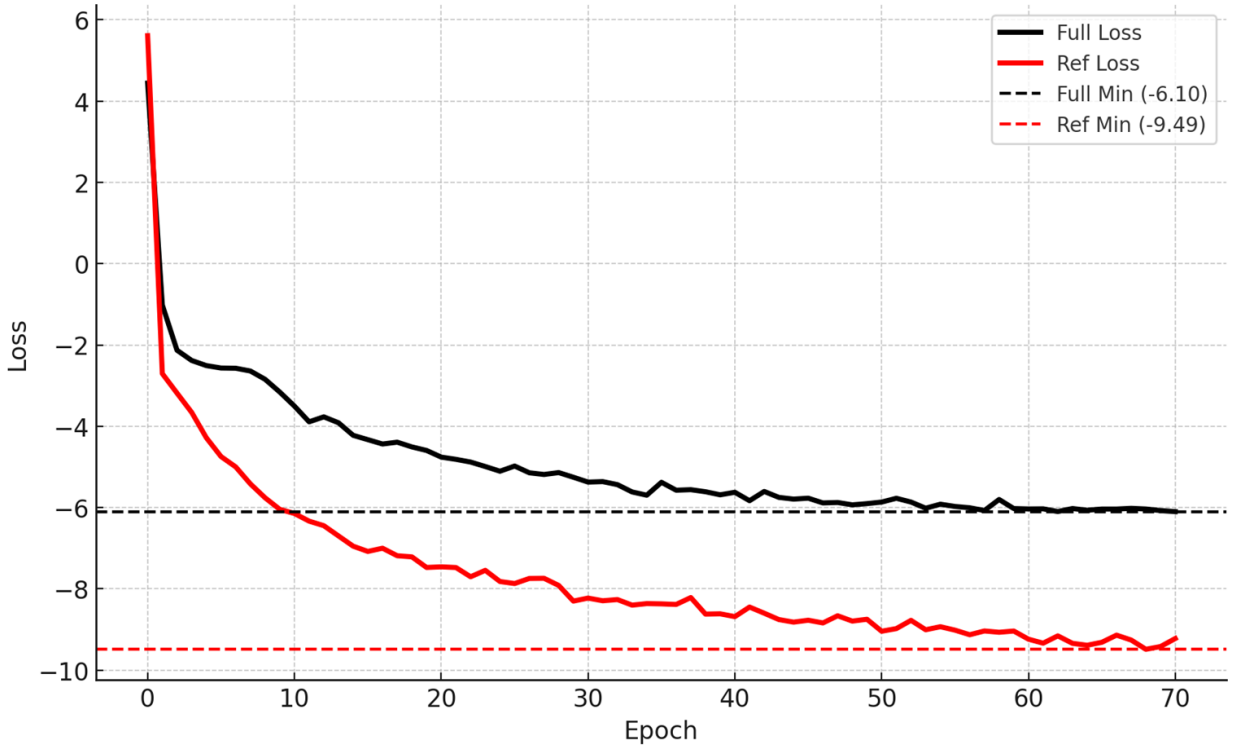


Модель DPCCN

Датасет Libri2Mix
43 часа аудио в обучающей выборке
5 часов в валидационной

Результаты

- Модель со всеми модификациями сходится хуже по Negative SNR Loss.
- Главная причина – изменения в нормализации.
- Измененный пулинг добавляет эхо.
- Однако модель теперь имеет алгоритмическую задержку в 1



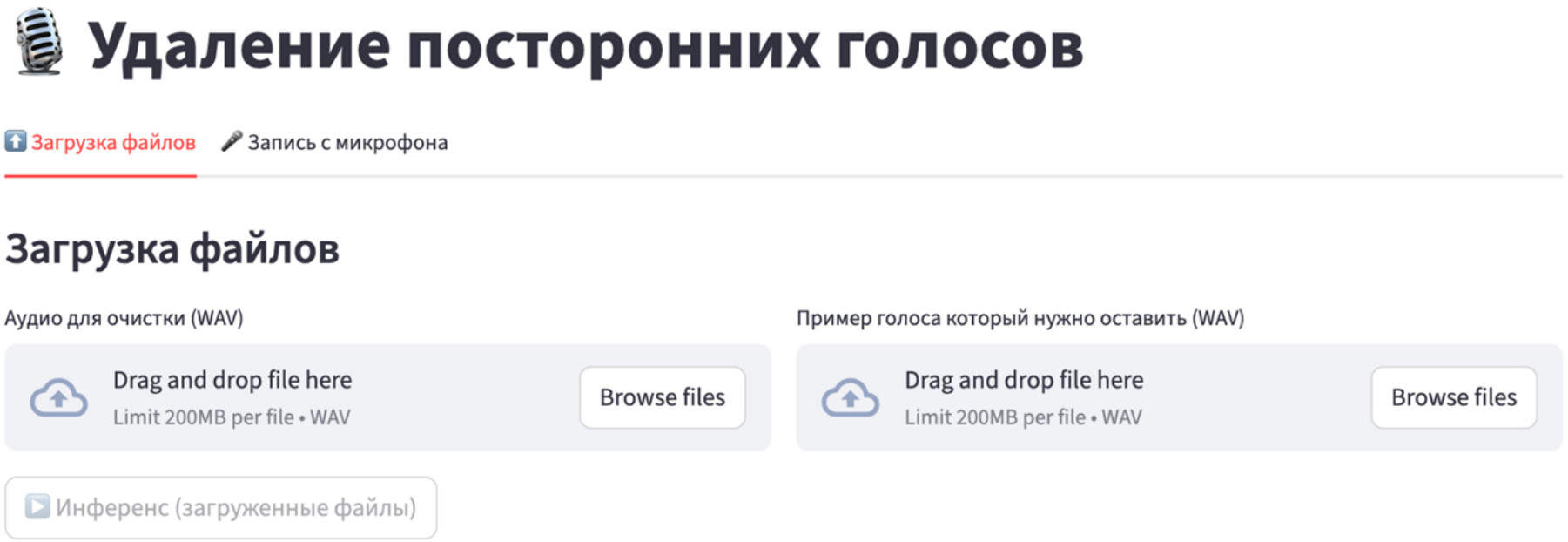
На графике изменение функции потерь по эпохам.
Чёрная линия —модель со всеми модификациями
Красная линия — оригинальная модель DPCCN.

Результаты

Таблица 8.1 — Сравнение моделей по SNR на 70-й эпохе валидации

Модель	**SNR (дБ)**
Без изменений (референс)	9.49
offline TCN conv layers → causal TCN conv layers	9.62
InstanceNorm → Стриминг InstanceNorm	6.64
Пирамидальный пулинг → Стриминг пулинг	8.53
Стриминг InstanceNorm + Стриминг пулинг	6.10

Демка


Удаление посторонних голосов
Загрузка файлов
Запись с микрофона
Загрузка файлов
Аудио для очистки (WAV)
Drag and drop file here
Limit 200MB per file • WAV
Browse files
Пример голоса который нужно оставить (WAV)
Drag and drop file here
Limit 200MB per file • WAV
Browse files
Инференс (загруженные файлы)

Заключение и выводы

Что сделано:

- Проведено 27 экспериментов
- Сожжено 1000+ GPU часов
- Получилось изменить часть модулей без большой потери качества

В дальнейшем

- Найди подходящую замену нормализации
- Провести замеры аппаратной задержки на целевых устройствах (ноутбуках)
- Сделать набор данных на русском языке

Спасибо за внимание