

Assignment 1. Report.

Tishin Vladislav

Task 1. Имплементация `LogMelFilterBanks`

В ходе выполнения 1 задачи я имплементировал `LogMelFilterBanks` воспользовавшись материалами с лекции и документации `torchaudio`. После я сгенерировал файл с частотой 16kHz и успешно прошел проверку на ассертах.

Task 2. Имплементация пайплайна

Dataloader

Для реализации даталoadеров я унаследовался от `SPEECHCOMMANDS` для создания кастомного датасета и переписал `._walker` для фильтрации данных с лейблами `yes/no`. Далее я инициализировал даталoadеры с полученными наборами данных и передал в них реализованные функции для паддинга и препроцессинга батчей.

Model Architecture

Архитектуру модели я реализовал с помощью четырех сверточных слоев, батчнормов, пулингов и одного линейного слоя, а также реализованного в 1 пункте `LogMelFilterBanks` через который данные проходят в первую очередь.

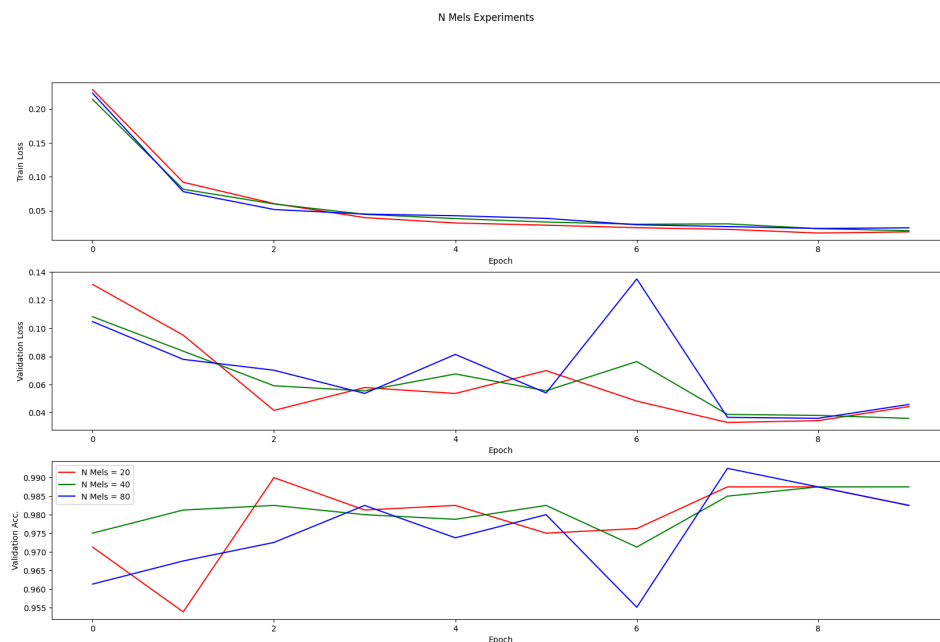
Num params and FLOPS

Для подсчета количества параметров и FLOPS я использовал библиотеку `ptflops`.

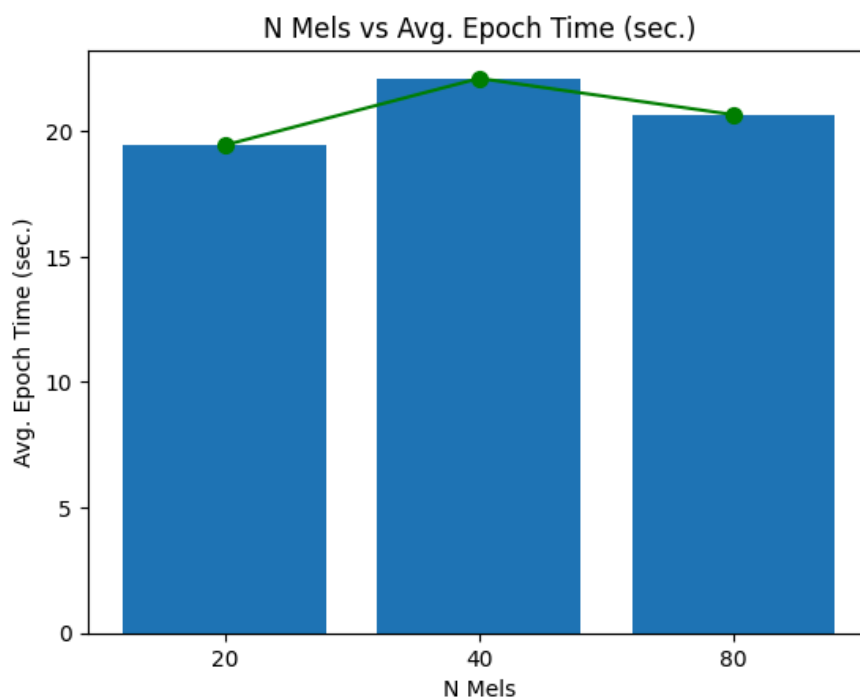
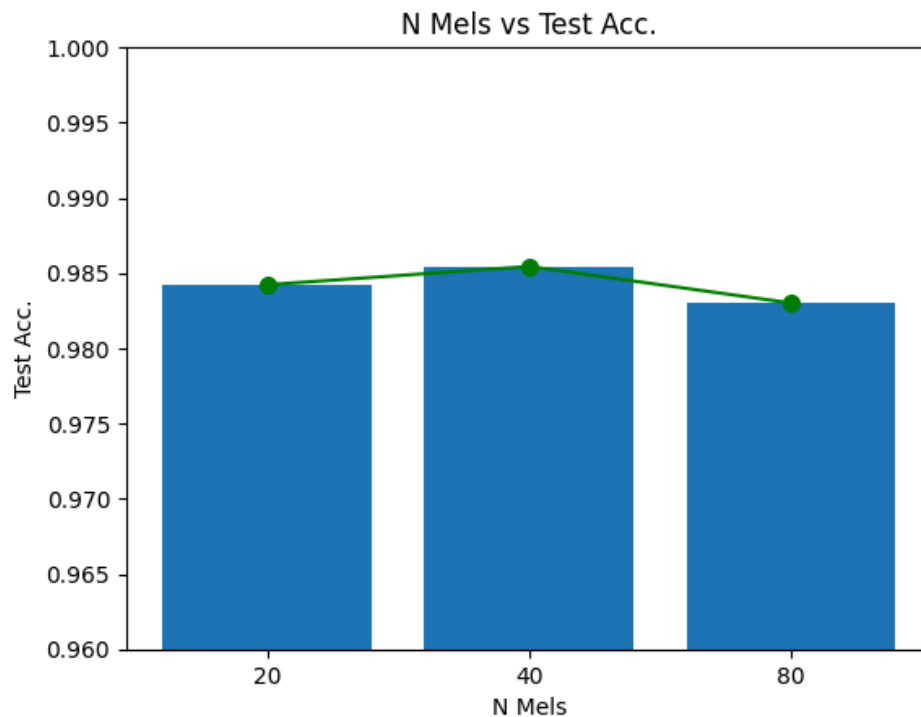
Training setup

Для обучения использовались `adam` optimizer с `lr = 0.003` и `StepLR` scheduler. Так как мы решаем задачу бинарной классификации, то в качестве лосса выбран `BCEWithLogitsLoss`.

Task 3. N Mels Experiments

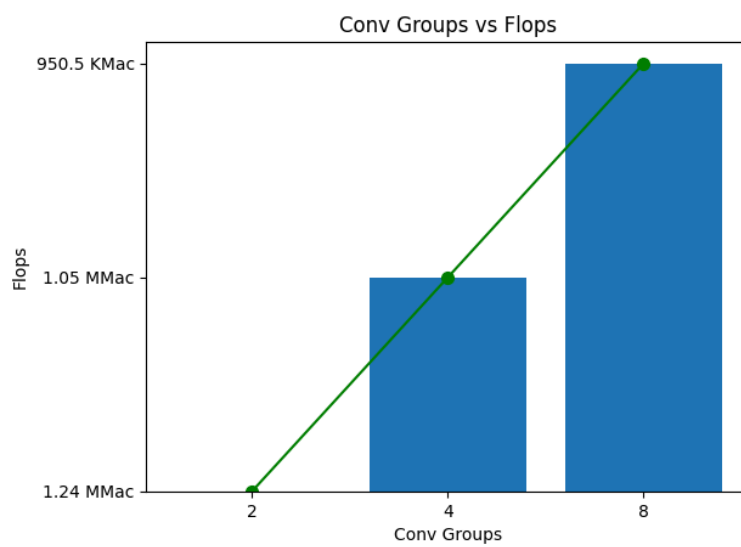
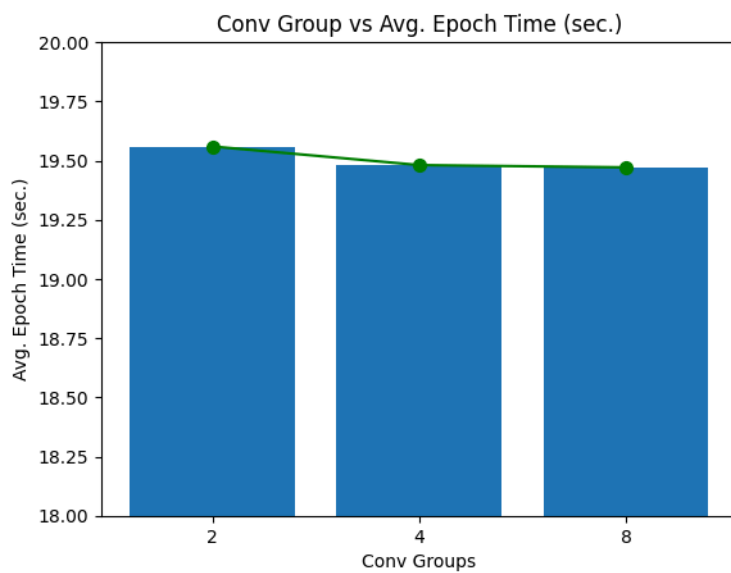
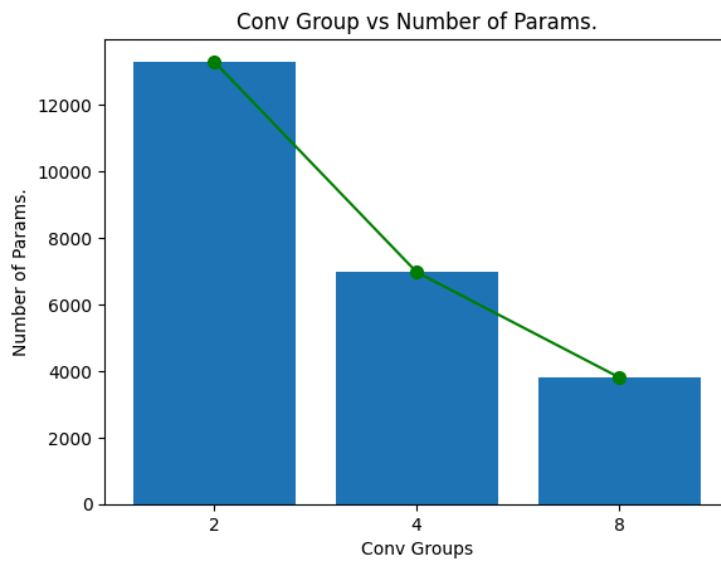


По результатам обучения все модели показали приблизительно одинаковые результаты, но видно, что модель с большим `n_mels` обучалась менее стабильно, чем варианты с 20 и 40.



При этом по метрикам с валидации и теста видно, что модель с `n_mels=20` показывает наилучший результат. Ее мы и возьмем в качестве бейзлайна для следующего задания. Также по графикам видно, что в среднем ее обучение заняло больше времени.

Task 4. Conv Groups Experiments



В среднем, обучение на всех conv groups заняло одинаковое время. При этом количество параметров вдвое уменьшалось при каждом увеличении числа groups, а FLOPS наоборот, вдвое увеличивалось.