

# HW1. Digital Signal Processing.

Тихонов Антон

## 1 Введение

В данной работе исследуется влияние параметра группировки (groups) в сверточных слоях нейронной сети и количества мел-частотных полос на эффективность и точность модели распознавания речевых команд на примере набора данных Binary Speech Commands.

## 2 Методология

Для проведения экспериментов была использована модель SpeechCNN с тремя сверточными слоями и полносвязным классификатором. Входные данные представлены в виде мел-спектрограмм, полученных из аудиосигналов. Варьировались следующие параметры:

- Количество мел-полос: 20, 40, 80
- Значение параметра groups в сверточных слоях: 2, 4, 8, 16

	Name	Type
0	conv1	Conv1d
1	conv2	Conv1d
2	conv3	Conv1d
3	pool	AdaptiveAvgPool1d
4	fc	Linear
5	criterion	BCEWithLogitsLoss

Рис. 1: Зависимость количества параметров от групп

### 3 Результаты экспериментов

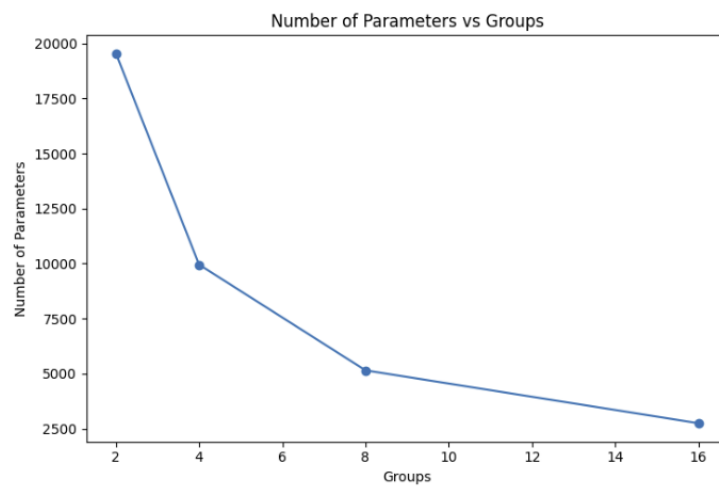


Рис. 2: Зависимость количества параметров от групп

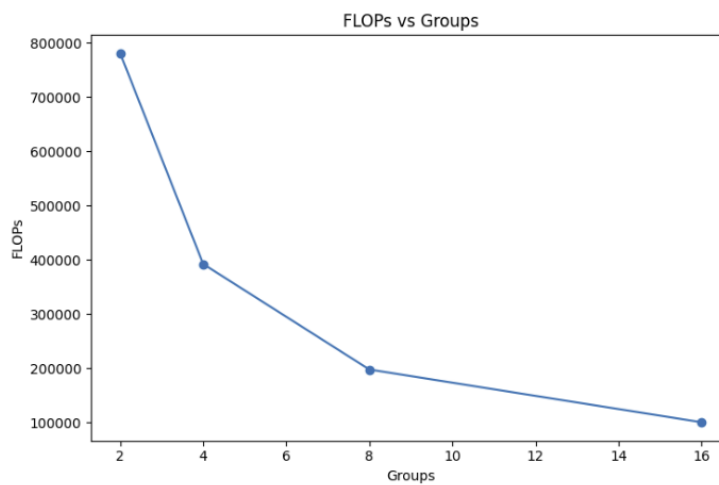


Рис. 3: Зависимость FLOPs от групп

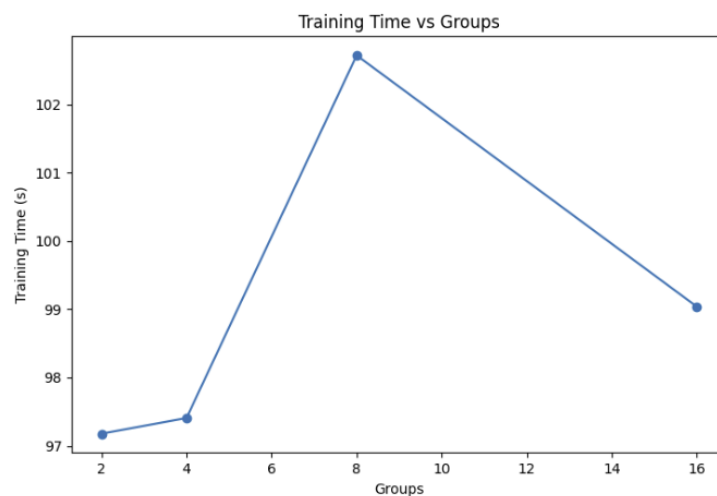


Рис. 4: Зависимость времени обучения от групп

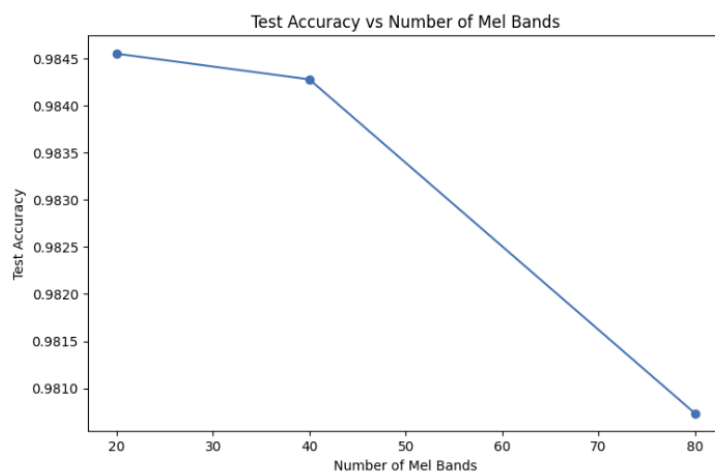


Рис. 5: Зависимость ассурасу от количества mel bands

## 4 Анализ и выводы

Проведенные эксперименты показали, что:

### 4.1 Влияние параметра groups

Увеличение параметра groups приводит к существенному сокращению количества параметров модели и вычислительной сложности. При этом на-

блюдается нелинейная зависимость времени обучения от количества групп, что может быть связано с особенностями реализации.

## **4.2 Влияние количества мел-полос**

Эксперименты показали, что использование меньшего количества мел-полос (20) обеспечивает более высокую точность распознавания по сравнению с традиционно используемыми 80 полосами. Это может быть объяснено тем, что для простой задачи бинарной классификации команд "yes" и "no" избыточное количество признаков может привести к переобучению модели.

## **5 Заключение**

Результаты исследования демонстрируют эффективность применения групповых сверток для создания легковесных моделей распознавания речи. Оптимальный баланс между вычислительными затратами, размером модели и точностью распознавания достигается при использовании 20 мел-полос и параметра groups равного 4 или 16.