

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
**«Южно-Уральский государственный университет  
(национальный исследовательский университет)»**  
Высшая школа электроники и компьютерных наук  
Кафедра системного программирования

# **Разработка Android-приложения для распознавания жестового языка в режиме реального времени**

Рецензент:  
Технический директор  
ООО «Программные системы»  
Е.В. Водяницкий




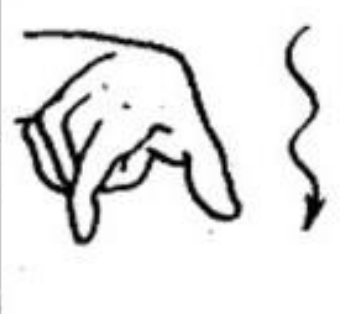
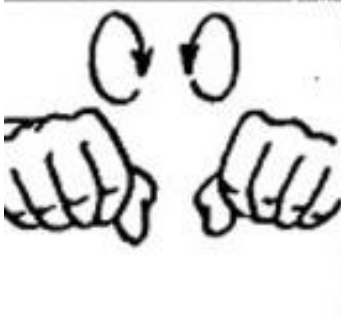
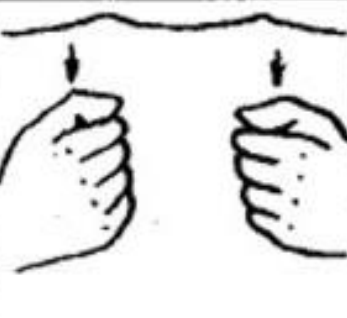




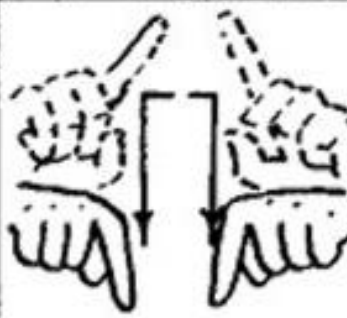

Автор:  
студент группы КЭ-229  
М.А. Свиридов  
Научный руководитель:  
доцент кафедры СП, к.ф.-м.н.  
А.Т. Латипова

Челябинск, 2024 г.

# Актуальность

- По данным ВОЗ около 470 миллионов человек страдают от потери слуха.
- К 2050 году количество таких людей превысит 700 миллионов.
- В России по различным источникам число таких людей около 9 миллионов.

# Русский жестовый язык

 <p>вертолет</p>	 <p>велосипед</p>	 <p>мотоцикл</p>	 <p>поход</p>
 <p>лодка</p>	 <p>рюкзак</p>	 <p>компас</p>	 <p>карта</p>
 <p>рыбалка</p>	 <p>отпуск</p>	 <p>путевка</p>	 <p>отдых</p>

# Цель и задачи

- **Цель:** разработка Android-приложения для распознавания жестового языка в режиме реального времени.
- **Задачи:**
  1. Провести обзор литературы и существующих приложений по предметной области.
  2. Подготовить набор данных, реализовать нейросетевую модель и провести ее обучение.
  3. Реализовать Android-приложение по распознаванию жестового языка в режиме реального времени.
  4. Провести тестирование приложения.

# Обзор решений

## Сурдофон

Приложение для общения глухих людей со слышащими.



## Особенности:

1. Показывают написанный текст в виде жестов с помощью анимированного «аватара».
2. Для распознавания жестов заявлена возможность загрузки и перевода видео.

## Недостатки:

1. Не предоставляют возможность по распознаванию жестов в режиме реального времени.

## Компьютерный переводчик на русский жестовый язык

Привет, меня зовут Саша.

Я — компьютерный переводчик на русский жестовый язык

Здесь я расскажу, что я умею делать

Давай начнем наше знакомство!



# Средства разработки нейросетевой МОДЕЛИ

- **Язык программирования:**  
Python
- **Библиотеки для работы с видеоизображениями:**  
opencv, numpy
- **Набор инструментов для обучения модели по видео:**  
MMAction2
- **Визуализация обучения:**  
TensorBoard
- **Библиотеки для обучения и экспорта модели:**  
pytorch, onnx

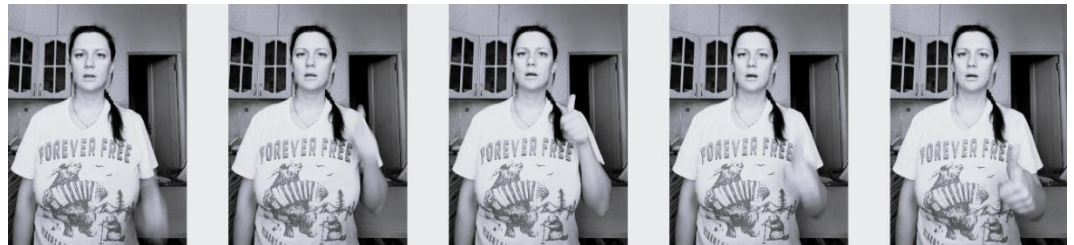
# Набор данных Slovo

1. Содержит 20 тыс. RGB-видео для 1000 жестов.
2. В каждом классе по 20 образцов.
3. Для обучения был разбит на обучающий и тестовый выборки, где 75% использовалось для обучения и 25% для тестов.
4. Средняя продолжительность видео с жестом – 50 кадров.

	attachment_id	user_id	width	height	length	text	train	begin	end
0	de81cc1c-...	1b...	1440	1920	14	привет	True	30	45
1	3c0cec5a-...	64...	1440	1920	32	утро	False	43	66
2	d17ca986-...	cf...	1920	1080	44	улица	False	12	31

where:

- `attachment_id` - video file name
- `user_id` - unique anonymized user ID
- `width` - video width
- `height` - video height
- `length` - video length
- `text` - gesture class in Russian Language
- `train` - train or test boolean flag
- `begin` - start of the gesture (for original dataset)
- `end` - end of the gesture (for original dataset)



# Обучение нейросетевой модели

- Архитектура модели – MViTv2.
- Варианты получения кадров<sup>1</sup>: 16x4, 32x3.
- Начало обучения с использованием контрольной точки обучения модели на наборе данных Kinetic-400<sup>2</sup>.

1. Wu Z., Xiong C., Ma C., Socher R., Davis L. Adaframe: Adaptive frame selection for fast video recognition.
2. Kay W., Carriera J., Simonyan K., Zhang B., Hiller C., Vijayanarasimhan S., Fabio V., Green T., Back T., Natsev P., Suleyman M., Zisserman A. The Kinetics Human Action Video Dataset.



# Подбор гиперпараметров

Скорость обучения	Точность	Dropout	Точность
0,001	0,64	0	0,78
<b>0,0001</b>	<b>0,81</b>	<b>0,1</b>	<b>0,81</b>
0,00001	0,59	0,2	0,77

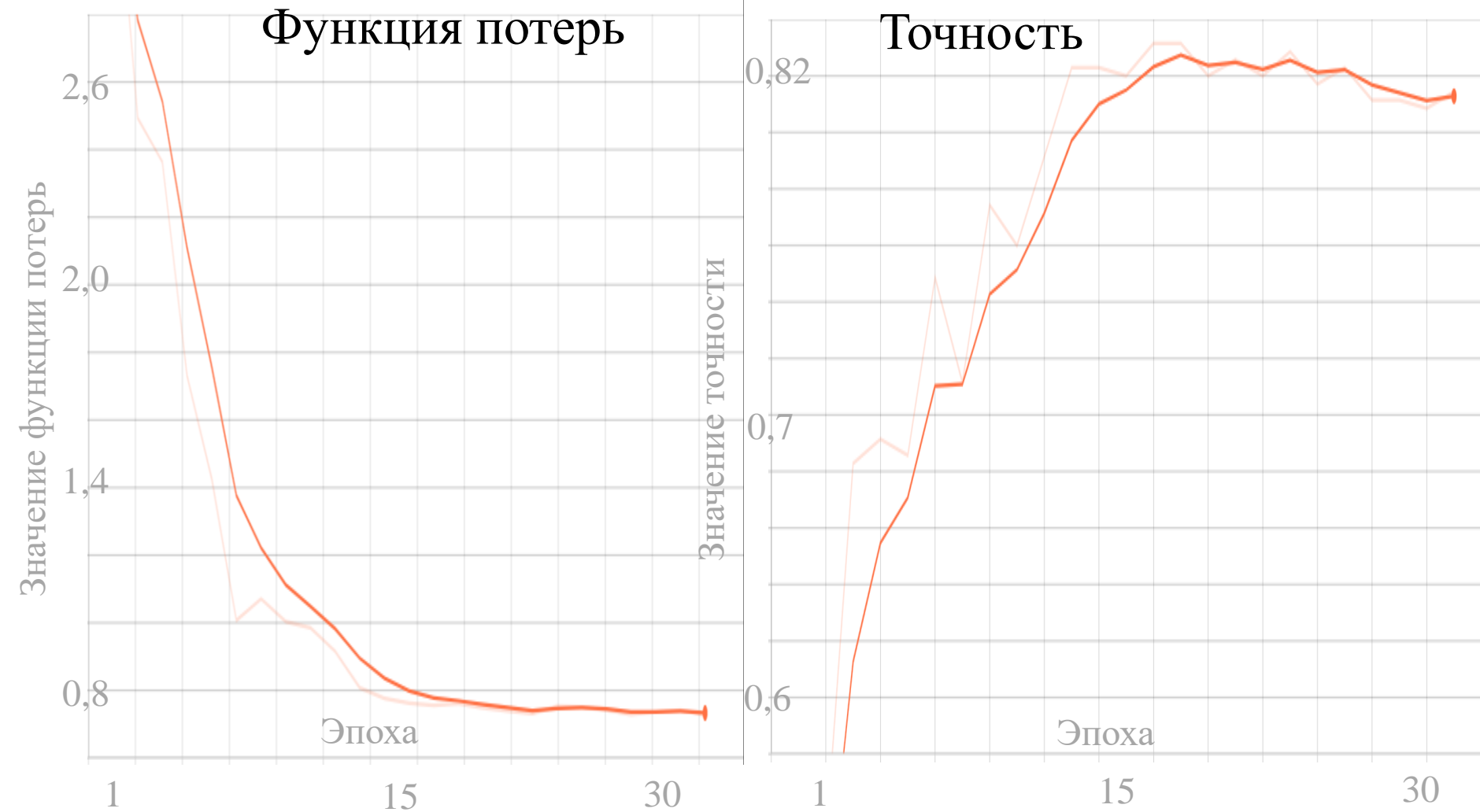
Weight decay	Точность
0,01	0,77
0,001	0,79
<b>0,0001</b>	<b>0,81</b>

# Подбор гиперпараметров

LRScheduler Step Size	Точность
1	0,65
5	0,74
<b>10</b>	<b>0,81</b>

LRScheduler gamma	Точность
<b>0,1</b>	<b>0,81</b>
0,5	0,64
0,9	0,58

# Подбор гиперпараметров



Наиболее подходящее кол-во эпох – **25**

# Результаты обучения

Модель	Accuracy	Precision	Recall
<b>MViTv2 16x4</b>	0,81	0,78	0,77
<b>MViTv2 32x3</b>	0,84	0,79	0,76

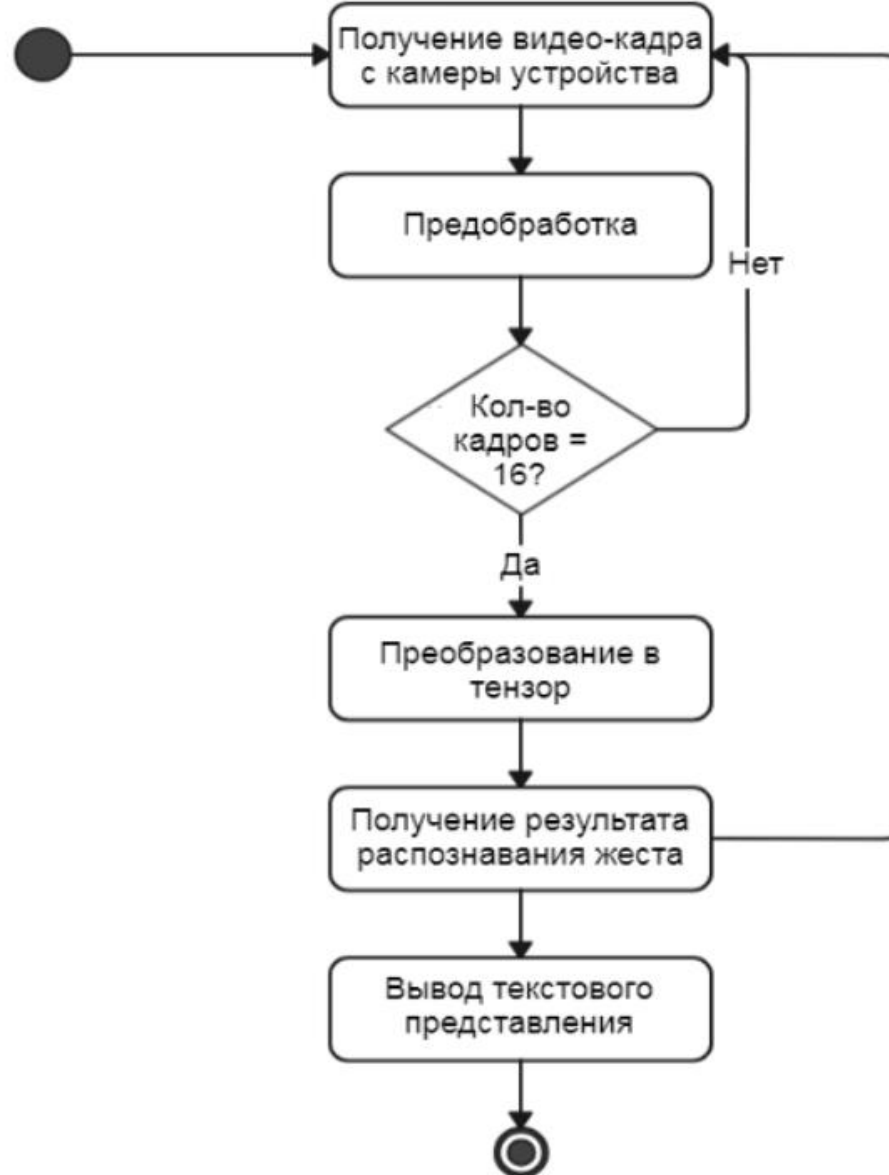
# Средства разработки Android-приложения

- **Платформа для создания Android-приложения:**  
.NET MAUI
- **Язык программирования:**  
C#
- **Язык разметки для создания интерфейса:**  
XAML
- **Библиотека для работы с видеоизображениями:**  
Emgu.CV
- **Библиотека для работы с моделью нейронной сети:**  
Microsoft.ML.OnnxRuntime

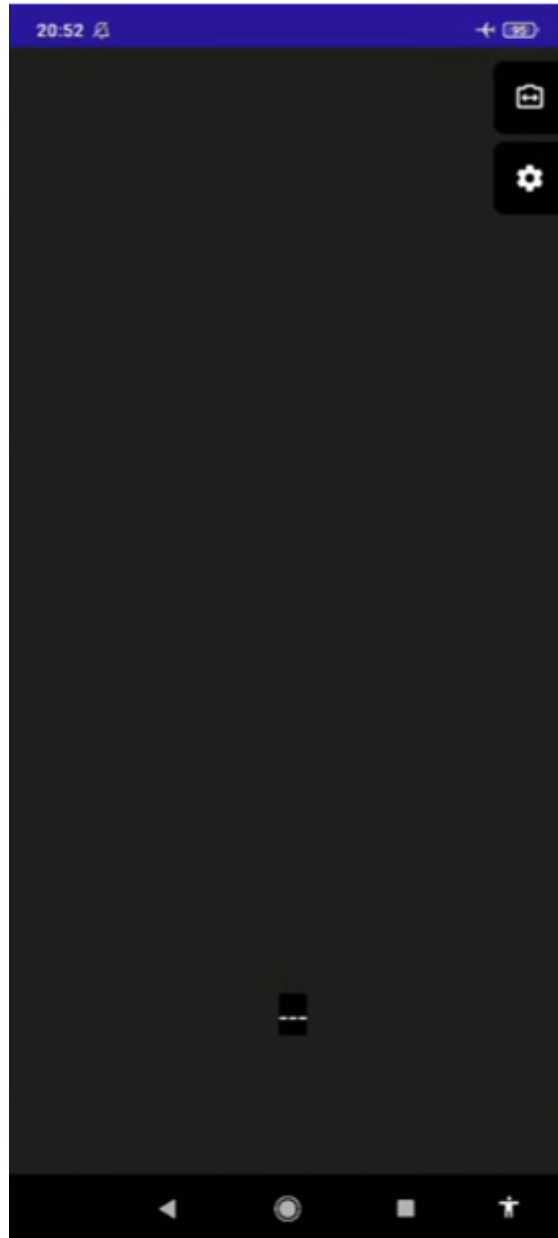
# Диаграмма вариантов использования



# Распознавание жеста

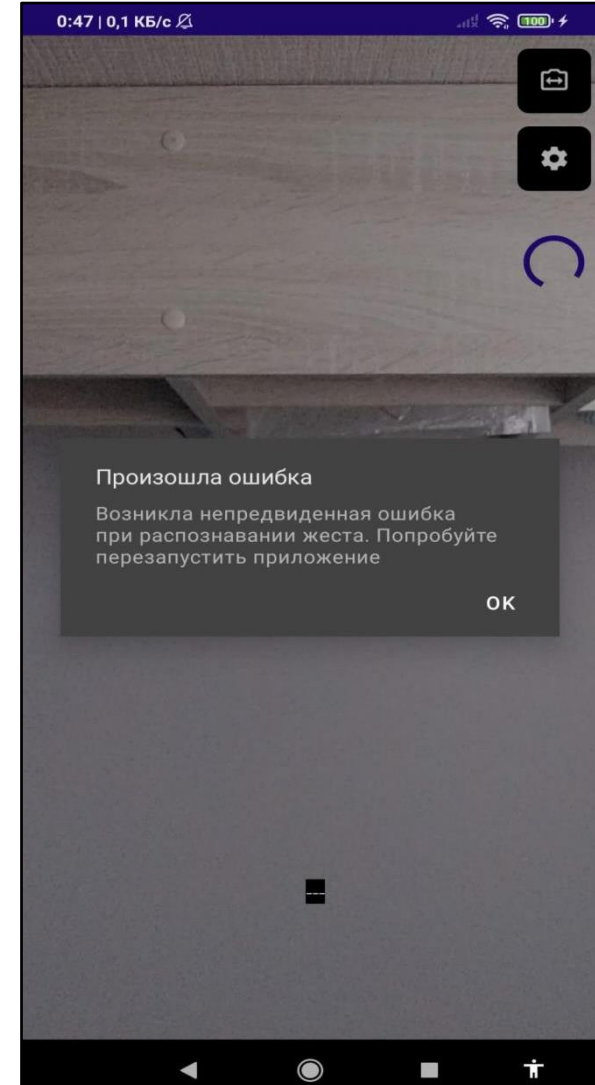
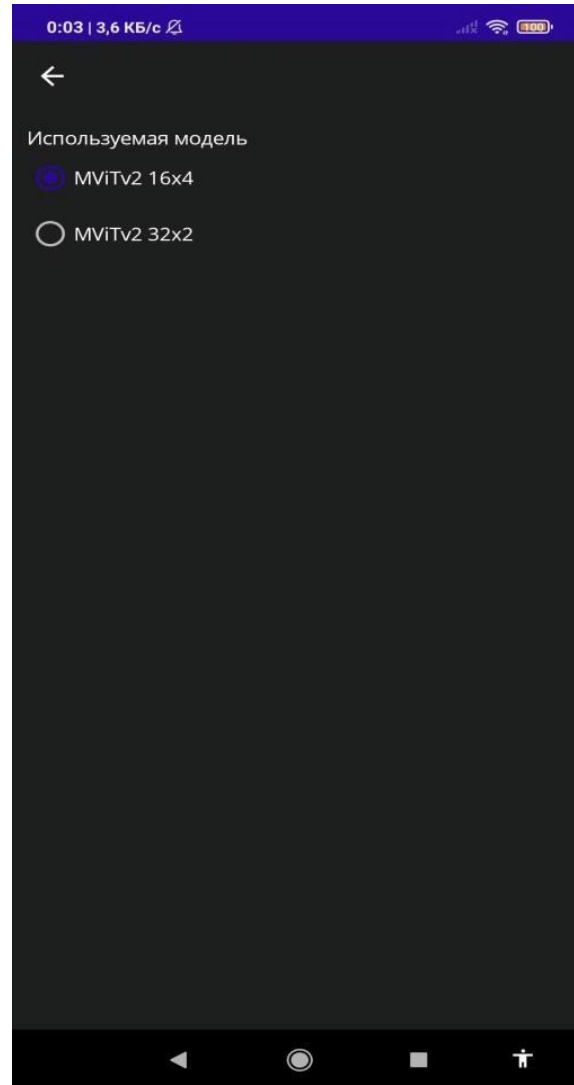
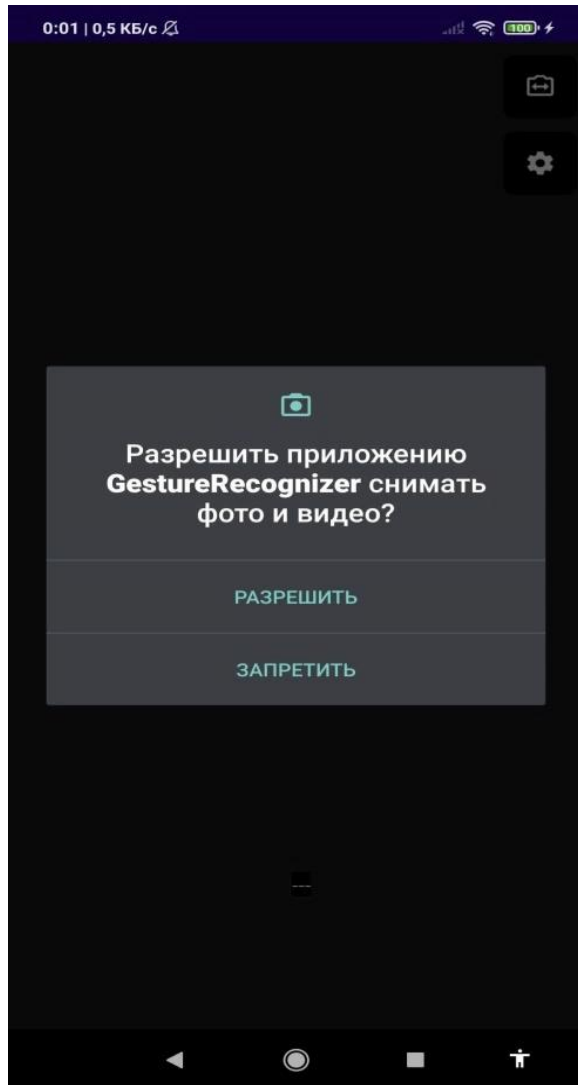


# Графический интерфейс





# Графический интерфейс



# Тестирование

Было произведено тестирование основного функционала приложения:

1. Отображение превью-изображения с передней и задней камер устройства.
2. Получение кадров для распознавания жестов.
3. Получение текстового представления показанного жеста на экране.
4. Переключение моделей.

Все результаты тестирования подтвердили корректность работы мобильного приложения.

# Основные результаты

1. Проведен обзор литературы и существующих приложений по предметной области.
2. Подготовлен набор данных, реализована нейросетевая модель и проведено ее обучение.
3. Реализовано Android-приложение по распознаванию жестового языка в режиме реального времени.
4. Проведено тестирование приложения.

Разработанное приложение полностью удовлетворяет необходимым требованиям и показывает корректный результат работы.

Дальнейшие улучшения: перенос на другие платформы, улучшение качества распознавания и скорости обработки данных.