

БИБЛИОТЕКА РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД
НА ПОЛЬЗОВАТЕЛЬСКОМ СЛОВАРЕ С ИСПОЛЬЗОВАНИЕМ
АУДИОВИЗУАЛЬНЫХ ДАННЫХ ДИКТОРА

OpenAV

Руководство пользователя

Руководитель разработки



Д. Иванько

20 мая 2024 г.

АННОТАЦИЯ

Настоящий документ является руководством пользователя к библиотеке OpenAV для распознавания речевых команд на пользовательском словаре с использованием аудиовизуальных данных диктора. Библиотека предназначена для решения задач автоматического распознавания речевых команд на основе интеллектуального анализа аудиовизуальных данных. Таким образом, на основе акустической информации (с микрофона) и визуальной информации (с видеокамеры) выполняется комплекс вычислений по распознаванию речи в режиме близком к реальному времени. Аудиовизуальная информация анализируется гибридным способом с использованием современных технологий искусственного интеллекта.

СОДЕРЖАНИЕ

Руководство пользователя	3
Перечень сокращений	8

РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ

Интерфейс модуля записи (см. рисунок) позволяет осуществить запись пользовательского набора данных.

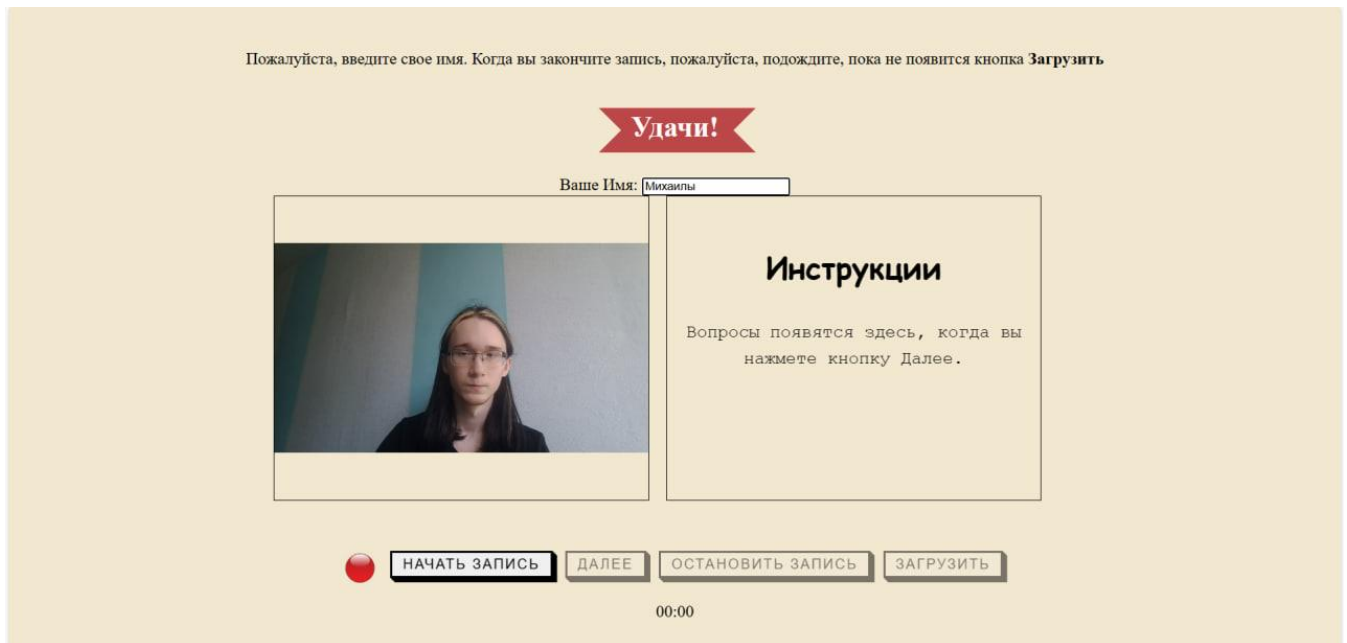


Рисунок 1 – Интерфейс модуля записи

Загрузка данных выполняется модулем загрузки данных с помощью выполнения консольной команды (рисунок).

```
(venv) PS C:\Users\USER\Downloads\OpenAV-develop\OpenAV-develop> openav_download --config C:\Users\USER\Downloads\OpenAV-develop\OpenAV-develop\openav\rsrs\download.yaml
```

Рисунок 2 – Запуск модуля загрузки

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (рисунок).

```
.....
Поиск, проверка и загрузка данных ...
.....
[2024-04-22 17:33:32] Поиск "download.yaml" файла ...
[2024-04-22 17:33:32] Загрузка данных из файла "download.yaml" ...
[2024-04-22 17:33:32] Проверка данных на валидность ...
.....
[2024-04-22 17:33:32] OpenAV - библиотека распознавания речевых команд на пользовательском словаре с использованием аудиовизуальных данных диктора:
Авторы:
  Рюмин Дмитрий [dl_03.03.1991@mail.ru]
  Иванько Денис [denis.ivankov@gmail.com]
  Шилов Николай [nick@ias.spb.su]
  Маркитантов Максим [m.markitantov@yandex.ru]
  Карпов Алексей [karpov@ias.spb.su]
Сопровождающие:
  Рюмин Дмитрий [dl_03.03.1991@mail.ru]
Версия: 1.0.0-a2
Лицензия: MIT
.....
[2024-04-22 17:33:33] Версии установленных библиотек ...
```

Рисунок 3 – Вывод модуля загрузки

Детектирование речевой активности выполняется одним из двух модулей (Vosk или Silero VAD)

Детектирование речевой активности (Vosk) производится с помощью выполнения консольной команды (см. рисунок).

```
(venv) PS C:\Users\USER\Downloads\OpenAV-develop\OpenAV-develop> openav_vosk_sr --config .\openav\rsrs\vosk_sr.yaml
```

Рисунок 4 – Запуск модуля детектирования речевой активности (Vosk)

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
[2024-04-12 15:16:58] Поиск вложенных директорий в директории "D:\av\3" (глубина вложенности: 1) ...
[2024-04-12 15:16:58] Поиск файлов с расширениями "mov, mp4, webm, wav" в директории "D:\av\3" (глубина вложенности: 1) ...
[2024-04-12 15:16:59] Загрузка Vosk модели "vosk-model-ru-0.42.zip" ...
[2024-04-12 15:16:59] Разархивирование архива "vosk-model-ru-0.42.zip" ...
[2024-04-12 15:16:59] Активация Vosk модели "vosk-model-ru-0.42" ...
[2024-04-12 15:18:48] Анализ файлов ...
[2024-04-12 15:19:50] 2 из 2 (100.0%) ... 1\2.mp4 ...
```

Рисунок 5 – Вывод модуля детектирования речевой активности (Vosk)

Ниже представлены данные до обработки (см. рисунок) и после обработки (см. рисунок). Каждая видеозапись была разделена на фрагменты, каждый из фрагментов соответствует одной речевой команде на исходной видеозаписи.

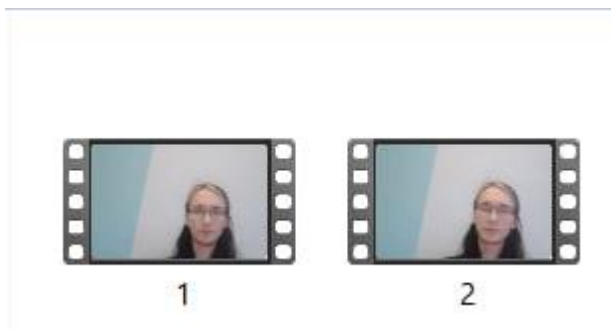


Рисунок 6 – Данные до обработки

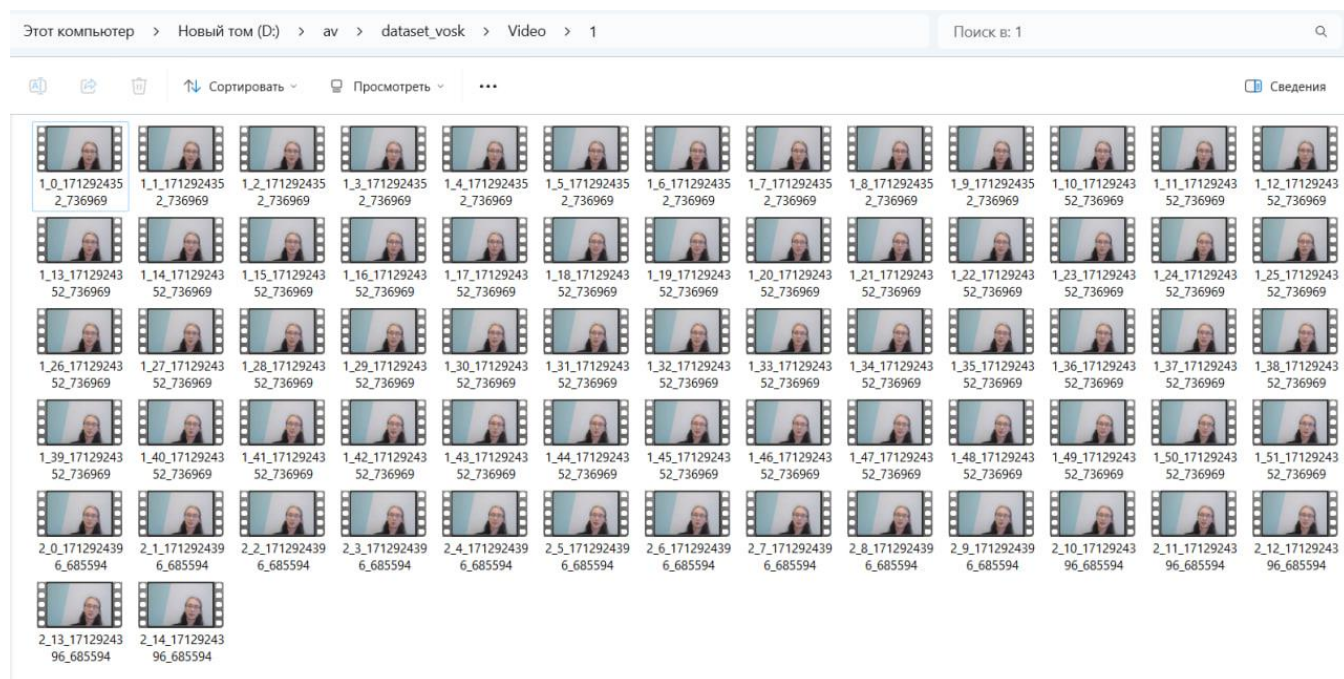


Рисунок 7 – Данные после обработки модулем детектирования речевой активности (Vosk)

Детектирование речевой активности (Vosk) производится с помощью выполнения консольной команды (см. рисунок).

```
(venv) D:\openAV>openav_vad --config C:\Users\USER\Downloads\OpenAV-develop\OpenAV-develop\openav\rsrs\vad.yaml
```

Рисунок 8 – Запуск модуля детектирования речевой активности (Silero VAD)

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
-----
[2024-04-22 17:54:01] Загрузка VAD модели "silero_vad" из репозитория https://github.com/snakers4/silero-vad ...
[2024-04-22 17:54:04] Поиск вложенных директорий в директории "D:\av\3" (глубина вложенности: 1) ...
[2024-04-22 17:54:04] Поиск файлов с расширениями "mov, mp4, webm, wav" в директории "D:\av\3" (глубина вложенности: 1)
...
[2024-04-22 17:54:04] Анализ файлов ...
[2024-04-22 17:54:04] 1 из 2 (50.0%) ... 1\1.mp4 ...
[2024-04-22 17:55:21] 2 из 2 (100.0%) ... 1\2.mp4 ...
-----
```

Рисунок 9 – Вывод модуля детектирования речевой активности (Silero VAD)

Ниже представлены данные до обработки (см. рисунок) и после обработки (см. рисунок). Каждая видеозапись была разделена на фрагменты, каждый из фрагментов соответствует одной речевой команде на исходной видеозаписи.

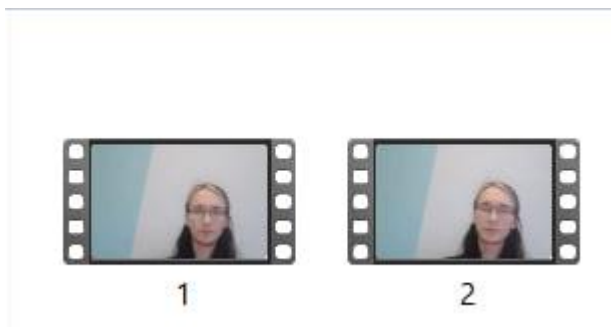


Рисунок 10 – Данные до обработки

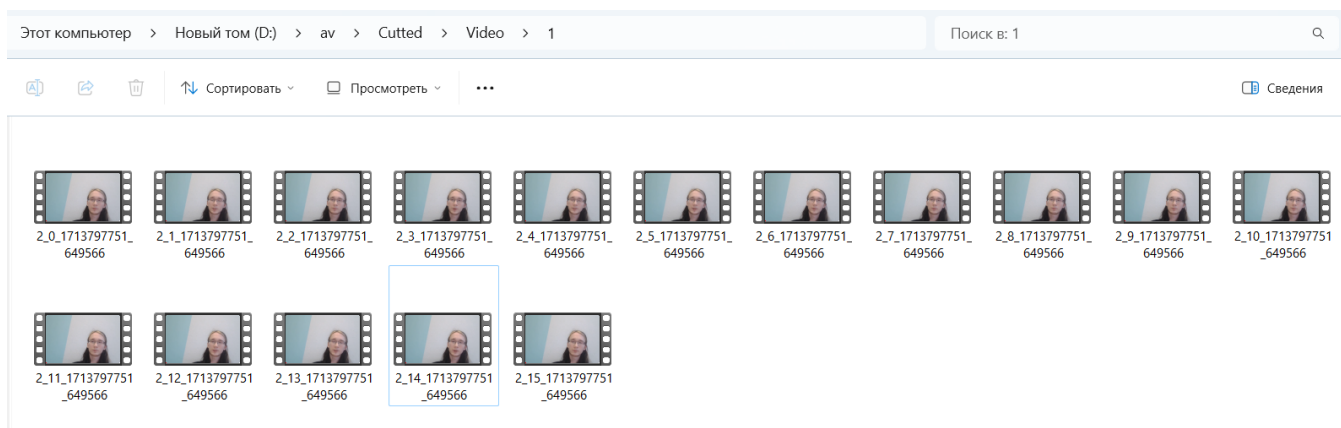


Рисунок 11 – Данные после обработки модулем детектирования речевой активности (Silero VAD)

Предобработка аудиоданных выполняется модулем предобработки данных с помощью выполнения консольной команды (см. рисунок).

```
(venv) D:\openAV>openav_preprocess_audio --config C:\Users\USER\Documents\OpenAV\openav\rsrs\audio_preprocessing.yaml
```

Рисунок 12 – Запуск модуля предобработки речевых аудиоданных

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
-----
[2024-04-22 18:33:50] Поиск вложенных директорий в директории "D:\av\3" (глубина вложенности: 1) ...
[2024-04-22 18:33:50] Поиск файлов с расширениями "mov, mp4, webm, wav" в директории "D:\av\3" (глубина вложенности: 1)
...
[2024-04-22 18:33:50] Предобработка речевых аудиоданных ...
[2024-04-22 18:33:57] 2 из 2 (100.0%) ... 1\2.mp4 ...
[2024-04-22 18:33:57] Все файлы успешно предобработаны ...
-----
```

Рисунок 13 – Вывод модуля предобработки речевых аудиоданных

Предобработка видеоданных выполняется модулем предобработки данных с помощью выполнения консольной команды (см. рисунок).

```
(venv) D:\openAV>openav_preprocess_video --config C:\Users\USER\Documents\OpenAV\openav\rsrs\video_preprocessing.yaml
```

Рисунок 14 – Запуск модуля предобработки речевых видеоданных

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
[2024-04-22 18:37:12] Поиск вложенных директорий в директории "D:\av\3" (глубина вложенности: 1) ...
[2024-04-22 18:37:12] Поиск файлов с расширениями "mov, mp4, webm" в директории "D:\av\3" (глубина вложенности: 1) ...
[2024-04-22 18:37:12] Предобработка речевых видеоданных ...
[2024-04-22 18:37:12] 2 из 2 (100.0%) ... 1\2.mp4 ...
[2024-04-22 18:37:12] Все файлы успешно предобработаны ...
```

Рисунок 15 – Вывод модуля предобработки речевых видеоданных

Аугментация данных выполняется модулем аугментации данных с помощью выполнения консольной команды (см. рисунок).

```
(venv) D:\openAV>openav_augmentation --config C:\Users\USER\Documents\OpenAV\openav\rsrs\augmentation.yaml
```

Рисунок 16 – Запуск модуля аугментации

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
[2024-04-22 18:57:26] Поиск вложенных директорий в директории "D:\openAV\dataset_audio" (глубина вложенности: 1) ...
[2024-04-22 18:57:26] Поиск файлов с расширениями "jpg, png" в директории "D:\openAV\dataset_audio" (глубина вложенности: 1) ...
[2024-04-22 18:57:26] Анализ файлов ...
[2024-04-22 18:57:26] 2 из 2 (100.0%) ... 1\2_1713800037_31743.png ...
[2024-04-22 18:57:26] Все файлы успешно обработаны ...
```

Рисунок 17 – Вывод модуля аугментации

Ниже представлены аудиоданные в форме спектрограммы до аугментации (см. рисунок) и после применения аугментации (см. рисунок).



Рисунок 18 – Аудиоданные до аугментации

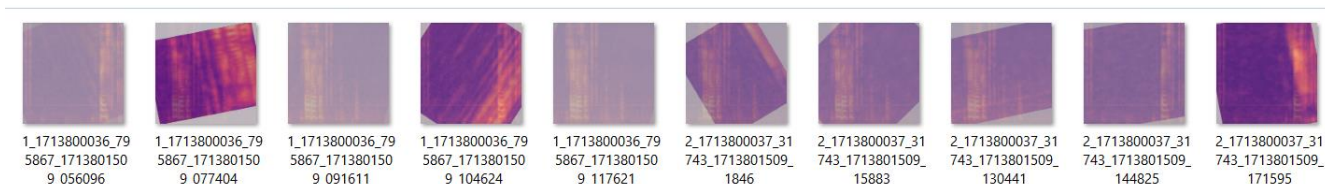


Рисунок 19 – Аудиоданные после аугментации

Обучение нейросетевых моделей на аудиоданных выполняется модулем обучения нейросетевых моделей с помощью выполнения консольной команды (см. рисунок).

```
(venv) D:\openAV>openav_train_audio --config C:\Users\USER\Documents\OpenAV\openav\rsrs\train_audio.yaml
```

Рисунок 20 – Запуск модуля обучения нейросетевых моделей (аудио)

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
-----
Автоматическое обучение на аудиоданных ...
-----
[2024-04-22 19:01:22] Поиск "train_audio.yaml" файла ...
[2024-04-22 19:01:22] Загрузка данных из файла "train_audio.yaml" ...
[2024-04-22 19:01:22] Проверка данных на валидность ...
12 12
-----
```

Рисунок 21 – Вывод модуля обучения нейросетевых моделей (аудио)

Обучение нейросетевых моделей на видеоданных выполняется модулем обучения нейросетевых моделей с помощью выполнения консольной команды (см. рисунок).

```
(venv) D:\openAV>openav_train_video --config C:\Users\USER\Documents\OpenAV\openav\rsrs\train_video.yaml
```

Рисунок 22 – Запуск модуля обучения нейросетевых моделей (видео)

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
-----
Автоматическое обучение на видеоданных ...
-----
[2024-04-22 19:04:23] Поиск "train_video.yaml" файла ...
[2024-04-22 19:04:23] Загрузка данных из файла "train_video.yaml" ...
[2024-04-22 19:04:23] Проверка данных на валидность ...
12 12
-----
```

Рисунок 23 – Вывод модуля обучения нейросетевых моделей (видео)

Обучение нейросетевых моделей на аудиовизуальных данных выполняется модулем обучения нейросетевых моделей с помощью выполнения консольной команды (см. рисунок)

```
(venv) D:\openAV>openav_train_audiovisual --config C:\Users\USER\Documents\OpenAV\openav\rsrs\train_audiovisual.yaml
```

Рисунок 24 – Запуск модуля обучения нейросетевых модулей (аудиовизуальные данные)

Вывод результатов работы модуля и информационные сообщения так же находятся в консоли (см. рисунок).

```
-----
[2024-05-03 16:54:13] Поиск вложенных директорий в директории "C:\Users\USER\Documents\rer\openAV\data" (глубина вложенности: 3) ...
[2024-05-03 16:54:13] Поиск файлов с расширениями "mov, mp4, webm" в директории "C:\Users\USER\Documents\rer\openAV\data" (глубина вложенности: 3) ...
[2024-05-03 16:54:13] Разбиение найденных файлов на выборки (обучающая, валидационная, тестовая) ...
[2024-05-03 16:54:13] Обучающая - 19 (76.0%), валидационная - 4 (16.0%), тестовая - 2 (8.0%) ...
[2024-05-03 16:54:14] Запуск процесса обучения ...
[2024-05-03 16:54:14] Эпоха: 1 из 2 ...
[2024-05-03 16:54:31] Значения ошибки: обучение - 3.152742, валидация - 2.98662, тест - 3.046073
[2024-05-03 16:54:31] Валидация: точность 1.0 | Тест: точность 1.0
[2024-05-03 16:54:31] Точность на валидационной выборке увеличилась (0 ----> 1.0). Сохранение модели ...
[2024-05-03 16:54:31] Точность для тестовой выборки: 1.0
[2024-05-03 16:54:31] Точность на тестовой выборке увеличилась (0 ----> 1.0)
-----
```

Рисунок 25 – Вывод модуля обучения нейросетевых моделей (аудиовизуальные данные)

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

Лист регистрации изменений

[illegible]