|  |
| --- |
| Библиотека распознавания речевых команд на пользовательском словаре с использованием аудиовизуальных данных диктора  OpenAV  Руководство пользователя |
|  |
| Руководитель разработки |
| \_\_\_\_\_\_\_\_\_\_\_ Д. Иванько |
| \_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ 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 4](#_Toc167117383)

[Перечень сокращений 10](#_Toc167117384)

Руководство пользователя

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

Изображение выглядит как текст, Человеческое лицо, снимок экрана, человек

Автоматически созданное описание

Рисунок 1 – Интерфейс модуля записи

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



Рисунок 2 – Запуск модуля загрузки

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

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

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 ).



Рисунок 4 – Запуск модуля детектирования речевой активности (Vosk)

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

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

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.

Изображение выглядит как снимок экрана, Прямоугольник

Автоматически созданное описание

Рисунок 6 – Данные до обработки

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

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

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



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

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

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

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.

Изображение выглядит как снимок экрана, Прямоугольник

Автоматически созданное описание

Рисунок 10 – Данные до обработки

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

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

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



Рисунок 12 – Запуск модуля предобработки речевых аудиоданных

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

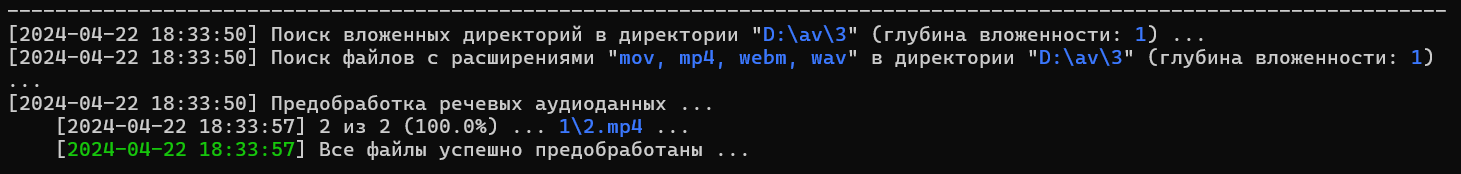


Рисунок 13 – Вывод модуля предобработки речевых аудиоданных

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



Рисунок 14 – Запуск модуля предобработки речевых видеоданных

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

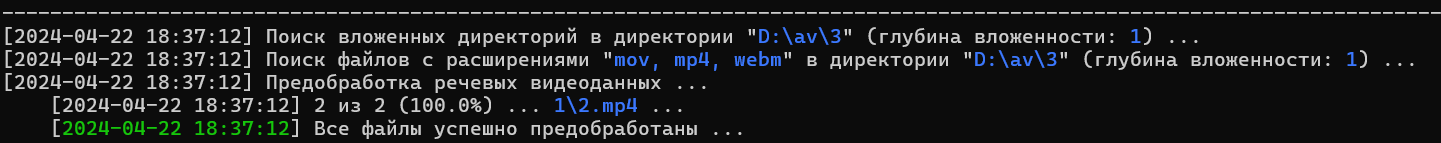


Рисунок 15 – Вывод модуля предобработки речевых видеоданных

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



Рисунок 16 – Запуск модуля аугментации

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

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 17 – Вывод модуля аугментации

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

Изображение выглядит как снимок экрана, текст, Прямоугольник, Красочность

Автоматически созданное описание

Рисунок 18 – Аудиоданные до аугментации

Изображение выглядит как снимок экрана, Красочность, Фиолетовый, Пурпурный цвет

Автоматически созданное описание

Рисунок 19 – Аудиоданные после аугментации

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



Рисунок 20 – Запуск модуля обучения нейросетевых моделей (аудио)

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

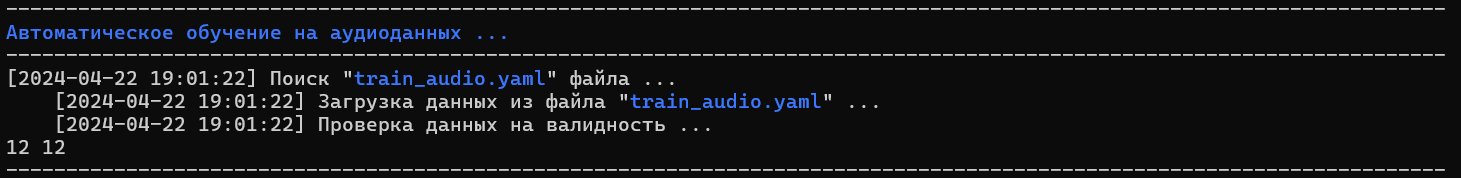


Рисунок 21 – Вывод модуля обучения нейросетевых моделей (аудио)

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



Рисунок 22 – Запуск модуля обучения нейросетевых моделей (видео)

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

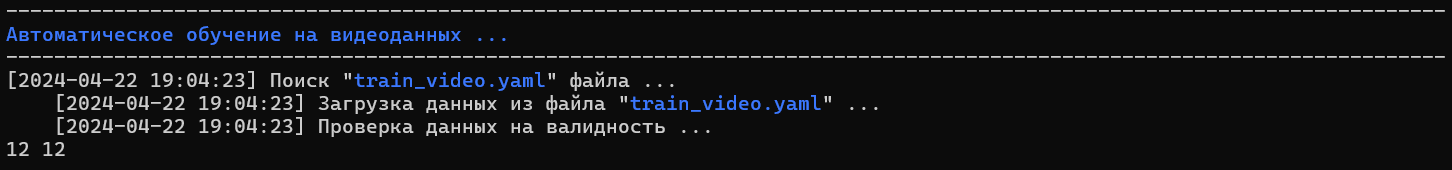


Рисунок 23 – Вывод модуля обучения нейросетевых моделей (видео)

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



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

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

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

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

Перечень сокращений

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Лист регистрации изменений | | | | | | | | | |
| Номера листов страниц | | | | | Всего  листов  страниц  в докум. | №  Документа | Входящий  № сопрово-  дительного  документа  и дата. | Подп. | Дата |
| Изм. | Изменен-ных | Заменен-  ных | Новых | Аннули-  рован-  ных |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |