|  |
| --- |
| Библиотека распознавания речевых команд на пользовательском словаре с использованием аудиовизуальных данных диктора  OpenAV  Руководство программиста |
|  |
| Руководитель разработки |
| \_\_\_\_\_\_\_\_\_\_\_ Д. Иванько |
| \_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ 2024 г. |

Аннотация

Настоящий документ является руководством программиста 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

[Назначение программы 4](#_Toc167117632)

[Условия выполнения программы 5](#_Toc167117633)

[Выполнение программы 6](#_Toc167117634)

[Команда для запуска модуля распознавания аудиовизуальной речи: 28](#_Toc167117635)

[1.1. Сообщения пользователю 31](#_Toc167117636)

[Перечень сокращений 33](#_Toc167117637)

Назначение программы

Данная библиотека предназначена для создания пользовательских систем автоматического распознавания речевых команд на ограниченном наборе словаря с использованием аудио и видео модальностей.

Библиотека предназначена для использования под управлением операционных систем Microsoft Windows 8, 8.1, 10 или семейства GNU/Linux.

Программное решение имеет следующий функционал: запись речевых аудиовизуальных данных, предобработка, аугментация, обучение нейросетевой модели и ее тестирование. Результатом последовательного выполнения этих этапов является обученная, протестированная и готовая к работе нейросетевая модель по распознаванию речевых аудиовизуальных данных

Условия выполнения программы

Для использования библиотеки необходимы следующие условия: видеофайлы с оптическим разрешением не менее 1024х768 пикселей при 30 и более кадрах в секунду, частота аудиофайлов от 16кГц, устройство под управлением операционных систем Microsoft Windows 8, 8.1, 10 или семейства GNU/Linux.

Выполнение программы

Библиотека разделена на 9 модулей с точки зрения функционала: модуль записи речевых аудиовизуальных данных, модуль загрузки данных, модуль детектирования речевой активности, модуль предобработки речевых аудиоданных, модуль предобработки речевых видеоданных, модуль аугментации данных, модуль обучения нейросетевых моделей, модуль распознавания речи, модуль объединения модальностей. Далее приведено описание каждого из модулей.

Модуль записи речевых аудиовизуальных данных

Модуль предназначен для записи аудиовизуальных данных с их последующей сортировкой по заданному словарю.

Команда для запуска модуля записи данных:

openav\_recorder\_app --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

После выполнения команды запускается локальный сервер, на котором развертывается приложение для записи данных. Для того чтобы начать использование графического интерфейса модуля перейдите по адресу в браузере <http://127.0.0.1:5000>.

В конфигурационном файле определены следующие функции, с помощью которых выполняется поиск доступных устройств и определение их технических характеристик для записи аудиовизуальных данных.

Функция find\_my\_devices (camera = True, micro = True) определяет и возвращает название устройств, которые подключены в системе. Передаваемые параметры camera и micro по умолчанию имеют значение True. Если вам не требуется использование камеры или микрофона, то необходимо передать значение False.

Пример использования данной функции. Переменным video и audio присваиваются значения, являющиеся результатом выполнения функции.

video, audio = find\_my\_devices(True, True)

Значения, возвращаемые из функции find\_my\_devices() необходимы для определения параметров изображения и звука, которые поддерживает найденное в системе устройство.

Функция get\_available\_parameters (device) принимает название устройства <str> device и возвращает список параметров, которые поддерживает указанное устройство. В случае, если передается название вебкамеры, то возвращаемый список будет содержать все доступные параметры разрешения изображения и максимальное количество кадров, соответствующее этому разрешению.

Пример использования данной функции. Переменной available\_camera\_params присваивается список доступных параметров.

available\_camera\_params = get\_available\_parameters (video)

Функция get\_camera\_params (dict, prefer = 'max') возвращает выбранные параметры из полного списка доступных параметров вебкамеры. Обязательным принимаемым на вход аргументом является список dict, который был получен с помощью предыдущей функции. Аргумент prefer = 'max' по умолчанию установлен на получение максимально допустимых параметров устройства. Данное значение передается в формате 640x480, либо max или min. Результатом выполнение функции является возврат значений available\_size, fps.

Пример использования данной функции. Переменным available\_res и available\_fps присваиваются значения, полученные в ходе выполнения функции get\_camera\_params.

available\_res, available\_fps = get\_camera\_params(available\_params, '640x480')

В конфигурационном файле также необходимо указать словарь, в соответствии с которым будет выполняться запись данных. Словарь имеет вид key-phrase, где key выступает в качестве порядкового номера фразы в словаре.

Пример словаря

dict = [  
 {'key': 0, 'phrase': 'Левая'},  
 {'key': 1, 'phrase': 'Правая'},  
 {'key': 2, 'phrase': 'Нажать левую'},  
 {'key': 3, 'phrase': 'Отпустить левую'},  
 {'key': 4, 'phrase': 'Нажать правую'},  
 {....................................},  
 {'key': 26, 'phrase': 'Завершить'}  
 ]

Помимо функий в рамках модуля был реализован графический интерфейс модуля записи. Данный интерфейс состоит из двух частей (см. рисунок ).

* В левой части представлено изображение, получаемое с подключенной камеры, и кнопка начать запись (Rec).
* В правой части располагается блок управление словарем. В этом блоке отображаются следующие элементы:
  + Цифра в круге означает порядковый номер фразы в словаре;
  + Фраза из словаря;
  + Кнопки переключения элементов словаря prev и next, после нажатия на которые отображают предыдущую или следующую фразу соответственно.

Изображение выглядит как текст, снимок экрана, диаграмма, дизайн

Автоматически созданное описание

Рисунок 1 - Интерфейс модуля для записи данных

Нажатие на кнопку записи (Rec) начинает процесс записи данных. После нажатия на кнопку она изменяет свой цвет, что означает об успешном запуске записи (см. рисунок ).



Рисунок 2 - Активное состояние кнопки записи

Повторное нажатие на кнопку записи завершает процесс. Записанный файл сохраняется по пути, указанному в конфигурационном файле. Изменить словарь также можно в конфигурационном файле.

Модуль загрузки данных

Команда для запуска модуля загрузки данных:

openav\_download --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Данный модуль позволяет выполнять поиск, проверку и загрузку данных. Модуль предусматривает вложенность директорий. Конфигурационный файл модуля включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой.

Отображение процесса выполнения команды в терминале (таблица 1).

Таблица 1 – Параметры отображения метаданных

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр | Тип | Значение по умолчанию | Описание |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 2)

Таблица 2 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_исходным\_данным> | Директория, где находятся данные, которые необходимо загрузить или проверить |
| depth | int | 1 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["mov", "mp4", "webm", "wav"] | Список расширений файлов, которые будут обрабатываться. Указывать можно как для видео, так и для аудио |

Модуль детектирования речевой активности

Детектирование речевой активности производится двумя способами: на основе Silero VAD и Vosk. Дальше подробнее про каждый из них.

Команда для запуска детектирования речевой активности в аудиовизуальном сигнале на основе Silero VAD:

openav\_vad --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры детектора голосовой активности (Silero VAD);
* параметры кодирования выходного файла.

Отображение процесса выполнения команды в терминале (таблица 3)

Таблица 3 – Параметры команды

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 4)

Таблица 4 – Параметры файловой системы

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_save\_model | str | <./models> | Директория, где будут размещаться скачанные модели, в данном случае модель для работоспособности VAD |
| path\_to\_dataset | str | <путь\_к\_исходным\_данным> | Директория, где находятся данные, которые необходимо обработать VAD |
| path\_to\_dataset\_vad | str | <./dataset\_vad> | Директория, куда сохраняются фрагменты аудиовизуального сигнала после обработки VAD |
| dir\_va\_names | dict | {"video": "Video", "audio": "Audio"} | Директории для сохранения видео и аудио файлов. Названия директорий могут быть произвольными |
| force\_reload | bool | false | Включение принудительной загрузки модели VAD из сети |
| clear\_dirvad | bool | true | Очистка директории, в которую сохраняются фрагменты аудиовизуального сигнала |
| depth | int | 1 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["mov", "mp4", "webm", "wav"] | Список расширений файлов, которые будут обрабатываться. Указывать можно как для видео, так и для аудио |

Параметры детектора голосовой активности (Silero VAD) (таблица 5)

Таблица 5 – Параметры детектора голосовой активности

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| sampling\_rate | int | 16000 | Частота дискретизации. На текущий момент поддерживаются частоты: 8000 и 16000 |
| threshold | float | 0.5 | Порог вероятности речи (от 0.0 до 1.0). VAD выводит вероятности речи для каждого звукового фрагмента, вероятности выше установленного значения считаются **речью.** Параметр необходимо настраивать индивидуально в зависимости от набора данных. Например, для шумных условий параметр рекомендуется устанавливать в значение от 0.7 до 0.95. В условиях низкого уровня шума или его отсутствия, параметр лучше устанавливать на низкие значения 0.1 — 0.25 |
| min\_speech\_duration\_ms | float | 250 | Минимальная длительность речевого фрагмента. Рекомендуется устанавливать в пределах от 750 мс до 1000 мс. Настройка также является индивидуальной, которую необходимо подбирать в зависимости от набора данных |
| min\_silence\_duration\_ms | float | 100 | Минимальная длительность тишины в выборках между отдельными речевыми фрагментами, прежде чем разделить его. Рекомендуется устанавливать в пределах 500 мс для того, чтобы не разделялись предложения. Однако, можно настроить индивидуально, под свои условия |
| window\_size\_samples | int | 1536 | Количество выборок в каждом окне. Предупреждение! Модели VAD были обучены с использованием выборок 512, 1024, 1536 для частоты дискретизации 16000 и 256, 512, 768 для частоты дискретизации 8000. Настоятельно рекомендуется использовать эти значения, изменение значений может повлиять на производительность модели |
| speech\_pad\_ms | float | 30 | Внутренние отступы для итоговых речевых фрагментов. Рекомендуется использовать значение 250 мс — 400 мс чтобы избежать обрезания фрагментов речи. Настройка также является индивидуальная и устанавливается в соответствии с условиями пользователя |

Параметры кодирования выходного файла (таблица 6)

Таблица 6 – Параметры кодирования

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| type\_encode | str | crf | Типы кодирования. Доступные варианты: ['qscale', 'crf'] |
| crf\_value | int | 23 | Качество кодирования (от 0 до 51. Чем ниже значение, тем лучше качество и наоборот. Стоит учитывать, что изменения качества кодирования влияет на скорость обработки |
| presets\_crf\_encode | str | medium | Скорость кодирования и сжатия. Доступные варианты: ['ultrafast', 'superfast', 'veryfast', 'faster', 'fast', 'medium', 'slow', 'slower', 'veryslow']. Изменения параметра влияет на скорость кодирования и степень сжатия |
| sr\_input\_type | str | audio | Типы файлов для распознавания речи. Доступные варианты: ['audio', 'video'] |

Команда для запуска детектирования речевой активности в аудиовизуальном сигнале на основе Vosk:

openav\_vosk\_sr --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена.

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры детектора голосовой активности (Vosk);
* параметры кодирования выходного файла.

Отображение процесса выполнения команды в терминале (таблица 7)

Таблица 7 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (

таблица 8)

Таблица 8 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_save\_model | str | <./models> | Директория, где будут размещаться скачанные модели, в данном случае модель для работоспособности VAD |
| path\_to\_dataset | str | <путь\_к\_исходным\_данным> | Директория, где находятся данные, которые необходимо обработать |
| path\_to\_dataset\_vosk\_sr | str | <./dataset\_vosk> | Директория, куда сохраняются фрагменты аудиовизуального сигнала после обработки |
| dir\_va\_names | dict | {"video": "Video", "audio": "Audio"} | Директории для сохранения видео и аудио файлов. Названия директорий могут быть произвольными |
| force\_reload | bool | false | Включение принудительной загрузки модели Vosk из сети |
| folder\_name\_unzip | str | <название\_папки> | Название папки, в которую будет извлекаться модель Vosk |
| clear\_dirvosk\_sr | bool | true | Очистка директории, в которую сохраняются фрагменты аудиовизуального сигнала |
| depth | int | 1 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["mov", "mp4", "webm", "wav"] | Список расширений файлов, которые будут обрабатываться. Указывать можно как для видео, так и для аудио |

Параметры детектора голосовой активности (Vosk) (таблица 9)

Таблица 9 – Параметры детектора голосовой активности

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| sampling\_rate | int | 16000 | Частота дискретизации. На текущий момент поддерживаются частоты: 8000 и 16000 |
| speech\_left\_pad\_ms | float | 0 | Внутренний отступ до начала речевого фрагмента. Настройка поможет избавиться от лишней тишина в начале обработанного фрагмента |
| speech\_right\_pad\_ms | float | 300 | Внутренний отступ в конце речевого фрагмента. Настройка поможет избавиться от лишней тишина в конце обработанного фрагмента, либо наоборот увеличить длительность речевого фрагмента, если фраза незначительно обрезается после обработки |
| lang\_model | str | ru | Выбор языка, на котором необходимо обработать данные. Поддерживаемые языки: ru и en |
| dict\_size | str | big | Размер словаря, на котором была обучена модель для распознавания. big и small. Влияет на качество работы модели, однако стоит учитывать, что малая модель будет работать быстрее, чем большая. |

В текущей версии для русского языка используются модели vosk-model-ru-0.42 и vosk-model-small-ru-0.22, обученные на большом и малом словарях, соответственно.

Параметры кодирования выходного файла (таблица 10)

Таблица 10 – Параметры кодирования

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| type\_encode | str | crf | Типы кодирования. Доступные варианты: ['qscale', 'crf'] |
| crf\_value | int | 23 | Качество кодирования (от 0 до 51. Чем ниже значение, тем лучше качество и наоборот. Стоит учитывать, что изменения качества кодирования влияет на скорость обработки |
| presets\_crf\_encode | str | medium | Скорость кодирования и сжатия. Доступные варианты: ['ultrafast', 'superfast', 'veryfast', 'faster', 'fast', 'medium', 'slow', 'slower', 'veryslow']. Изменения параметра влияет на скорость кодирования и степень сжатия |
| sr\_input\_type | str | audio | Типы файлов для распознавания речи. Доступные варианты: ['audio', 'video'] |

Модуль предобработки речевых аудиоданных

Модуль выполняет предобработку речевых аудиоданных, в данном случае извлекается спектрограмма из исходной аудиодорожки.

Команда для запуска модуля предобработки речевых аудиоданных:

openav\_preprocess\_audio --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры предобработки речевых аудиоданных.

Отображение процесса выполнения команды в терминале (Таблица 11)

Таблица 11 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 12)

Таблица 12. Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_исходным\_данным> | Директория, где находятся данные, которые необходимо загрузить или проверить |
| path\_to\_dataset\_audio | str | <путь\_к\_конечным\_данным> | Директория, в которую будут сохраняться аудиоданные после предобработки |
| depth | int | 1 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["mov", "mp4", "webm", "wav"] | Список расширений файлов, которые будут обрабатываться. Указывать можно как для видео, так и для аудио |

Параметры предобработки речевых аудиоданных (таблица 13)

Таблица 13 – Параметры предобработки речевых аудиоданных

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| sampling\_rate | int | 16000 | Частота дискретизации аудиосигнала |
| n\_fft | int | 2048 | Размер параметра FFT, создает n\_fft // 2 + 1 бин |
| hop\_length | int | 512 | Длина перехода между окнами STFT |
| n\_mels | int | 128 | Количество фильтр блоков mel |
| power | float | 2.0 | Показатель степени магнитуды спектрограммы. Должно быть > 0 |
| center | bool | true | Включение установки отступов с обеих сторон относительно центра аудиодорожки |
| pad\_mode | str | reflect | Управление отступами, применяется, когда значение параметра center = True. Доступные значения constant, reflect, replicate, circular. По умолчанию reflect |
| norm | str | reflect | Управление отступами, применяется, когда значение параметра center = True. Доступные значения constant, reflect, replicate, circular. По умолчанию reflect |

Модуль предобработки речевых видеоданных

Модуль выполняет предобработку речевых видеоданных, в данном случае извлекаются области губ из исходных видеокадров.

Команда для запуска модуля предобработки видеоданных:

openav\_preprocess\_video --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры предобработки речевых видеоданных.

Отображение процесса выполнения команды в терминале (Таблица 14)

Таблица 14 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 15)

Таблица 15 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_исходным\_данным> | Директория, где находятся данные, которые необходимо загрузить или проверить |
| path\_to\_dataset\_video | str | <путь\_к\_конечным\_данным> | Директория, в которую будут сохраняться данные после предобработки |
| depth | int | 1 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["mov", "mp4", "webm", "wav"] | Список расширений файлов, которые будут обрабатываться. Указывать можно как для видео, так и для аудио |

Параметры предобработки речевых видеоданных (таблица 16)

Таблица 16 – Параметры предобработки речевых видеоданных

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| width | int | 112 | Ширина кадра с найденной областью губ |
| height | int | 112 | Высота кадра с найденной областью губ |
| color\_mode | str | gray | Цветовая гамма конечного изображения. Доступные значения: gray - изображение в градациях серого; rgb - изображение в цветном формате |

Модуль аугментации данных

Данный модуль позволяет генерировать дополнительные данные с помощью различных параметров аугментации. Команда для запуска аугментации данных:

openav\_augmentation --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена. Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры аугментации данных.

Отображение процесса выполнения команды в терминале (Таблица 17)

Таблица 17 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 18)

Таблица 18 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_input\_directory | str | <путь\_к\_исходным\_данным> | Директория, где находятся данные, которые необходимо аугментировать |
| path\_to\_output\_directory | str | <путь\_к\_аугментированным\_данным> | Директория, куда сохраняются аугментированные данные |
| clear\_diraug | bool | true | Очистка директории, в которую сохраняются аугментированные данные |
| depth | int | 1 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_input\_directory |
| ext\_search\_files | list | ["jpg", "png"] | Список расширений файлов, которые будут обрабатываться |

Параметры аугментации данных (таблица 19)

Таблица 19 – Параметры аугментации данных

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| crop\_px\_min | int | 0 | Минимальное количество пикселей для обрезки изображения с каждой стороны. Диапазон значений от 0 и 1000000 |
| crop\_px\_max | int | 16 | Максимальное количество пикселей для обрезки изображения с каждой стороны. Диапазон значений от 0 и 1000000 |
| crop\_percent\_min | float | 0 | Минимальный процент для обрезки изображения с каждой стороны. Диапазон значений от 0 и 1.0 |
| crop\_percent\_max | float | 0.5 | Максимальный процент для обрезки изображения с каждой стороны. Диапазон значений от 0 и 1.0 |
| flip\_lr\_probability | float | 0.5 | Значение коэффициента вероятности отражения по вертикальной оси. Диапазон значений от 0 и 1.0 |
| flip\_ud\_probability | float | 0.5 | Значение коэффициента вероятности отражения по горизонтальной оси. Диапазон значений от 0 и 1.0 |
| blur\_min | float | 0 | Минимальное значение коэффициента размытия изображения. Диапазон значений от 0 и 3.0 |
| blur\_max | float | 1 | Максимальное значение коэффициента размытия изображения. Диапазон значений от 0 и 3.0 |
| scale\_x\_min | float | 0.5 | Минимальное значение масштабирования по оси X. Диапазон значений от 0 и 10.0 |
| scale\_x\_max | float | 2 | Максимальное значение масштабирования по оси X. Диапазон значений от 0 и 10.0 |
| scale\_y\_min | float | 0.5 | Минимальное значение масштабирования по оси Y. Диапазон значений от 0 и 10.0 |
| scale\_y\_max | float | 2 | Максимальное значение масштабирования по оси X. Диапазон значений от 0 и 10.0 |
| rotate\_min | int | -45 | Минимальное значение угла поворота изображения. Диапазон значений от -90 и 90 |
| rotate\_max | int | 45 | Максимальное значение угла поворота изображения. Диапазон значений от -90 и 90 |
| contrast\_min | float | 0 | Минимальное значение коэффициента контрастности. Диапазон значений от -10.0 и 10.0 |
| contrast\_max | float | 1.0 | Максимальное значение коэффициента контрастности. Диапазон значений от -10.0 и 10.0 |
| alpha | float | 0.7 | Значения коэффициента MixUp. Диапазон значений от 0 и 1.0 |
| count | int | 5 | Количество применений процесса аугментации к изображению |

Модуль обучения нейросетевых моделей

Команда для запуска процесса обучения акустических нейросетевых моделей:

openav\_train\_audio --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена.

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры настройки процесса обучения нейросетевых моделей.

Отображение процесса выполнения команды в терминале (таблица 20)

Таблица 20 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 21)

Таблица 21 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_набору\_данных> | Директория, где размещается подготовленный набор данных для обучения нейросетевых моделей |

Параметры процесса обучения акустических нейросетевых моделей (таблица 22)

Таблица 22 – Параметры процесса обучения

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| len\_audio | int | ваше значение | Количество аудиофайлов |
| size\_spec | int | width: 224 height: 224 | Размер входного изображения спектрограммы в px |
| padding\_spec | bool | True | Добавление отступов на изображениях спектрограмм |
| seed | int | 42 | Параметр для инициализации случайных процессов, который обеспечивает воспроизводимость результатов и одинаковые начальные условия |
| batch\_size | int | 8 | Размер батча. Общее число тренировочных объектов, представленных в одном батче. Устанавливается в зависимости от возможности вашего оборудования |
| channels\_spec | int | 1 | Количество каналов изображения. 1- одноканальное изображение (в серых тонах), 3 - трёхканальное изображение (RGB) |
| lr | float | 0.0001 | Коэффициент скорости обучения. Чем меньше значение, тем дольше будет идти обучение модели. Однако, стоит помнить, что может наступить переобучение модели. Данный коэффициент подбирается эмпирическим путем |
| epoch | int | 100 | Количество эпох обучения модели. Проход одной эпохи значит, что весь датасет прошел через нейронную сеть в прямом и обратном направлении только один раз. Параметр epoch\_stop позволяет избежать переобучения модели |
| epoch\_stop | int | 5 | Количество эпох, в течение которых модель не прогрессирует в обучении, т.е. если по прошествии заданного числа эпох модель не показывала прирост accuracy (либо ваша метрика), то в данном случае процесс обучения останавливается и сохраняется модель на эпохе с наибольшей точностью. Данный параметр позволяет предотвратить переобучение модели, а также снизить длительность обучения |

Команда для запуска процесса обучения визуальных нейросетевых моделей:

openav\_train\_video --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена.

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры настройки процесса обучения нейросетевых моделей.

Отображение процесса выполнения команды в терминале (таблица 23)

Таблица 23 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 24)

Таблица 24 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_набору\_данных> | Директория, где размещается подготовленный набор данных для обучения нейросетевых моделей |

Параметры процесса обучения визуальных нейросетевых моделей (таблица 25)

Таблица 25 – Параметры процесса обучения

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| len\_video | int | ваше значение | Количество видеофайлов |
| size\_lips | int | width: 112 height: 112 | Размер входного изображения области губ в px |
| padding\_lips | bool | True | Добавление отступов на изображениях губ |
| seed | int | 42 | Параметр для инициализации случайных процессов, который обеспечивает воспроизводимость результатов и одинаковые начальные условия |
| batch\_size | int | 8 | Размер батча. Общее число тренировочных объектов, представленных в одном батче. Устанавливается в зависимости от возможности вашего оборудования |
| channels\_lips | int | 1 | Количество каналов изображения. 1- одноканальное изображение (в серых тонах), 3 - трёхканальное изображение (RGB) |
| lr | float | 0.0001 | Коэффициент скорости обучения. Чем меньше значение, тем дольше будет идти обучение модели. Однако, стоит помнить, что может наступить переобучение модели. Данный коэффициент подбирается эмпирическим путем |
| epoch | int | 100 | Количество эпох обучения модели. Проход одной эпохи значит, что весь датасет прошел через нейронную сеть в прямом и обратном направлении только один раз. Параметр epoch\_stop позволяет избежать переобучения модели |
| epoch\_stop | int | 5 | Количество эпох, в течение которых модель не прогрессирует в обучении. Т. е. если по прошествии, например, 5 эпох модель не показывала прирост accuracy (либо ваша метрика), то в данном случае процесс обучения останавливается и сохраняется модель на эпохе с наибольшей точностью. Данный параметр позволяет предотварить переобучение модели, а также снизить длительность обучения |

Модуль распознавания речи

Команда для запуска процесса распознавания речи:

openav\_test\_audio --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена. Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры настройки процесса тестирования обученных нейросетевых моделей.

Отображение процесса выполнения команды в терминале (таблица 26)

Таблица 26 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 27)

Таблица 27 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_набору\_данных> | Директория, где размещается подготовленный набор данных для тестирования обученных нейросетевых моделей |
| path\_to\_model | str | <путь\_к\_моделям> | Директория, где размещаются обученные нейросетевые модели |
| depth | int | 3 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["wav", "aac"] | Список расширений файлов, которые будут обрабатываться |

Параметры процесса тестирования обученных акустических нейросетевых моделей (таблица 28)

Таблица 28 – Параметры процесса распознавания речи

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| size\_spec | int | width: 224 height: 224 | Размер входного изображения спектрограммы в px |
| channels\_spec | int | 1 | Количество каналов изображения. 1- одноканальное изображение (в серых тонах), 3 - трёхканальное изображение (RGB) |
| metric | str | accuracy | Метрика, в соответствии с которой будет вывод результатов тестирования обученных нейросетевых моделей |

Команда для запуска процесса распознавания речи визуальных нейросетевых моделей:

openav\_test\_video --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена.

Конфигурационный файл включает в себя следующие настройки:

* отображение процесса выполнения программы в терминале (командной строке);
* работа с файловой системой;
* параметры настройки процесса тестирования обученных нейросетевых моделей.

Отображение процесса выполнения команды в терминале (таблица 29)

Таблица 29 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 30)

Таблица 30 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_набору\_данных> | Директория, где размещается подготовленный набор данных для тестирования обученных нейросетевых моделей |
| path\_to\_model | str | <путь\_к\_моделям> | Директория, где размещаются обученные нейросетевые модели |
| depth | int | 3 | Глубина иерархии для получения данных. Указывается количество подкаталогов в директории path\_to\_dataset |
| ext\_search\_files | list | ["mov", "mp4", "webm"] | Список расширений файлов, которые будут обрабатываться |

Параметры процесса тестирования обученных визуальных нейросетевых моделей (таблица 31)

Таблица 31 – Параметры процесса распознавания речи

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| size\_lips | int | width: 112 height: 112 | Размер входного изображения области губ в px |
| channels\_lips | int | 1 | Количество каналов изображения. 1- одноканальное изображение (в серых тонах), 3 - трёхканальное изображение (RGB) |
| metric | str | accuracy | Метрика, в соответствии с которой будет вывод результатов тестирования обученных нейросетевых моделей |

Модуль объединения модальностей

Команда для запуска модуля объединения аудио- и видеомодальностей:

openav\_train\_audiovisual --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Для запуска команды необходимо обязательно указать путь к конфигурационному файлу. Запускать программу необходимо из директории, где она расположена.

Конфигурационный файл включает в себя следующие настройки:

* Отображение процесса выполнения программы в терминале (командной строке)
* Работа с файловой системой
* Параметры настройки процесса обучения нейросетевых моделей

Отображение процесса выполнения команды в терминале (таблица 32)

Таблица 32 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 33)

Таблица 33 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_набору\_данных> | Директория, где размещается подготовленный набор данных для обучения нейросетевых моделей |
| subfolders | str | train: "train" val: "val" test: "test" | Директории с обучающей, тестовой и валидационной выборками |
| path\_to\_model\_fa | str | <путь\_к\_весам\_аудио\_модели> | Путь к предобученной модели, которая используются для извлечения акустических признаков. |
| path\_to\_model\_fv | str | <путь\_к\_весам\_видео\_модели> | Путь к предобученной модели, которая используются для извлечения визуальных признаков. |
| path\_to\_save\_models | str | <путь\_сохранения\_обученных\_моделей> | Путь, по которому будут храниться обученные нейросетевые модели |

Параметры процесса обучения визуальных нейросетевых моделей (таблица 34)

Таблица 34. Параметры процесса обучения

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| n\_classes | int | 26 | Количество классов для задачи классификации. Соответствуют количеству фраз из базы данных |
| classes | list | ["1\_Левая", "2\_Правая", ... "26\_Калибровка"] | Список названий классов, которые представлены в базе данных |
| seed | int | 42 | Параметр, задающий начальное значение генератора псевдослучайных чисел в PyTorch. Установка фиксированного seed обеспечивает воспроизводимость результатов между разными запусками программы. При одинаковом seed все случайные операции, такие как инициализация весов нейронной сети, перемешивание данных и др., будут давать одинаковые результаты. |
| max\_segment | int | 2 | Гиперпараметр, определяющий максимальное количество перекрывающихся сегментов, на которые разбиваются длинные последовательности аудио и видео данных перед подачей их на вход нейронной сети для обучения или вывода |
| epochs | int | 150 | Количество эпох обучения модели. Проход одной эпохи значит, что весь датасет прошел через нейронную сеть в прямом и обратном направлении только один раз. Параметр patience позволяет избежать переобучения модели |
| patience | int | 15 | Количество эпох, в течение, которых модель не прогрессирует в обучении, т.е. если по прошествии заданного числа эпох модель не показывала прирост accuracy (либо ваша метрика), то в данном случае процесс обучения останавливается и сохраняется модель на эпохе с наибольшей точностью. Данный параметр позволяет предотвратить переобучение модели, а также снизить длительность обучения |
| batch\_size | int | 2 | Размер батча. Общее число тренировочных объектов, представленных в одном батче. Устанавливается в зависимости от возможности вашего оборудования |
| leaning\_rate | float | 0.0001 | Коэффициент скорости обучения. Чем меньше значение, тем дольше будет идти обучение модели. Однако, стоит помнить, что может наступить переобучение модели. Данный коэффициент подбирается эмпирическим путем |
| weight\_decay | float | 0.0 | Параметр, используемый для регуляризации весов нейронной сети путем добавления L2-регуляризации к функции потерь во время обучения. Позволяет предотвратить переобучение и подбирается экспериментальным путем |
| hidden\_units | int | 256 | Количество скрытых единиц (hidden units) в декодере нейронной сети. Этот параметр определяет размерность внутреннего представления данных в декодере, что влияет на емкость (expressive capacity) и способность декодера извлекать и обобщать сложные зависимости в данных. Подбирается эмипирическим путем |
| hidden\_features | int | 128 | Количество скрытых признаков (hidden features) или временных шагов, используемых в нейросетевые модели. Этот параметр связан с тем, как модель обрабатывает последовательные данные, такие как аудио и видео. Увеличение этого параметра позволяет модели обрабатывать более длинные входные последовательности, но также увеличивает вычислительную сложность и требования к памяти, уменьшение может ускорить обучение, но при этом модель будет видеть только более короткие временные зависимости. Подбирается эмпирически |
| input\_dim | int | 512 | Размерность входных векторов признаков для аудио и видео данных, подаваемых в модель. Например, параметр input\_dim=512 указывает, что на вход нейронной сети будут подаваться векторы признаков размера 512 для каждого временного шага последовательности. Размер входного вектора обычно является результатом предварительной обработки и извлечения низкоуровневых признаков из исходных аудио/видео данных с помощью отдельных моделей или методов обработки сигналов. |
| shape\_audio | int | channels: "1" n\_mels: "64" samples: "306" | Здесь определяется размерность входных тензоров аудиоданных, ожидаемые моделью. Это форма задается тремя числами channels - количество каналов в аудиоданных, чаще всего аудио является моно, поэтому здесь канал равен 1. При стерео он равен 2. n\_mels - количество мел-частотных кепстральных коэффициентов (MFCC), используемых для представления аудиоданных. samples - количество временных выборок или окон, входящих в одно аудио представление. Являются фиксированными параметрами |
| shape\_video | int | frames: "29" channels: "3" width: "88" height: "88" | Здесь определяется размерность входных тензоров видеоданных, ожидаемые моделью. frames - это количество видеокадров в одном сегменте. channels - это количество цветовых каналов (R, G, B) в видеокадре. Для изображения в серых тонах количество каналов равно 1. width - ширина видеокадра в пикселях. height - высота видеокадра в пикселях. Являются фиксированными параметрами |
| encoder\_decoder | int | 5 | Количество блоков энкодера и декодера в архитектуре трансформера, который используется в этой модели. Доступные значения от 1 до 50. Правильный выбор количества энкодер/декодер блоков является важным параметром и часто подбирается экспериментально в зависимости от размера данных, доступных вычислительных ресурсов и требуемого качества модели |
| optimizer | str | lion | Выбор оптимизатора обучения нейросетевой модели. Может существенно оказывать влияние на скорость сходимости, стабильность и окончательную точность модели. Разные оптимизаторы имеют свои преимущества и недостатки, подходящие для определенных задач и архитектур моделей. Доступные варианты adam, adamw, sgd, lion |
| requires\_grad | str | av | Этот параметр предоставляет гибкий способ настройки режима обучения сложных моделей, позволяя либо полностью зафиксировать предобученные компоненты, либо обучать их совместно с основной частью модели. none - обучаются все компоненты модели: аудио, видео и трансформер. a - заморозка весов для аудио компонента. v - заморозка весов для видео компонента. av - заморозка весов для аудио и видео компонентов, обучается только трансформер |

Команда для запуска модуля распознавания аудиовизуальной речи:

openav\_test\_audiovisual --config <путь\_к\_вашему\_конфигурационному\_файлу>.yaml

Конфигурационный файл включает в себя следующие настройки:

* Отображение процесса выполнения программы в терминале (командной строке)
* Работа с файловой системой
* Параметры процесса распознавания аудиовизуальных речи

Отображение процесса выполнения команды в терминале (таблица 35)

Таблица 35 – Параметры отображения процесса

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| hide\_metadata | bool | false | Включение отображения метаданных |
| hide\_libs\_vers | bool | false | Включение отображения версий установленных библиотек в командной строке |

Работа с файловой системой (таблица 36)

Таблица 36 – Параметры работы с файловой системой

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| path\_to\_dataset | str | <путь\_к\_набору\_данных> | Директория, где размещается подготовленный набор данных для обучения нейросетевых моделей |
| subfolders | str | test: "test" | Директория с тестовой выборкой |
| path\_to\_model | str | <путь\_к\_обученной\_модели> | Путь, по которому находится обученная нейросетевая модель |
| path\_to\_save\_confusion\_matrix | str | <путь\_сохранения\_матриц\_спутывания> | Путь, по которому будут храниться матрицы спутывания |

Параметры процесса обучения визуальных нейросетевых моделей (таблица 37)

Таблица 37 – Параметры процесса распознавания аудиовизуальной речи

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Тип** | **Значение по умолчанию** | **Описание** |
| n\_classes | int | 26 | Количество классов для задачи классификации. Соответствуют количеству фраз из базы данных |
| classes | list | ["1\_Левая", "2\_Правая", ... "26\_Калибровка"] | Список названий классов, которые представлены в базе данных |
| max\_segment | int | 2 | Параметр, определяющий максимальное количество перекрывающихся сегментов, на которые разбиваются длинные последовательности аудио и видео данных перед подачей их на вход нейронной сети для обучения или вывода |
| hidden\_units | int | 256 | Количество скрытых единиц (hidden units) в декодере нейронной сети. Этот параметр определяет размерность внутреннего представления данных в декодере, что влияет на емкость (expressive capacity) и способность декодера извлекать и обобщать сложные зависимости в данных |
| hidden\_features | int | 128 | Количество скрытых признаков (hidden features) или временных шагов, используемых в нейросетевые модели. Этот параметр связан с тем, как модель обрабатывает последовательные данные, такие как аудио и видео. Увеличение этого параметра позволяет модели обрабатывать более длинные входные последовательности, но также увеличивает вычислительную сложность и требования к памяти, уменьшение может ускорить обучение, но при этом модель будет видеть только более короткие временные зависимости |
| input\_dim | int | 512 | Размерность входных векторов признаков для аудио и видео данных, подаваемых в модель. Например, параметр input\_dim=512 указывает, что на вход нейронной сети будут подаваться векторы признаков размера 512 для каждого временного шага последовательности. Размер входного вектора обычно является результатом предварительной обработки и извлечения низкоуровневых признаков из исходных аудио/видео данных с помощью отдельных моделей или методов обработки сигналов. |
| shape\_audio | int | channels: "1" n\_mels: "64" samples: "306" | Здесь определяется размерность входных тензоров аудиоданных, ожидаемые моделью. Это форма задается тремя числами channels - количество каналов в аудиоданных, чаще всего аудио является моно, поэтому здесь канал равен 1. При стерео он равен 2. n\_mels - количество мел-частотных кепстральных коэффициентов (MFCC), используемых для представления аудиоданных. samples - количество временных выборок или окон, входящих в одно аудио представление. Являются фиксированными параметрами |
| shape\_video | int | frames: "29" channels: "3" width: "88" height: "88" | Здесь определяется размерность входных тензоров видеоданных, ожидаемые моделью. frames - это количество видеокадров в одном сегменте. channels - это количество цветовых каналов (R, G, B) в видеокадре. Для изображения в серых тонах количество каналов равно 1. width - ширина видеокадра в пикселях. height - высота видеокадра в пикселях. Являются фиксированными параметрами |
| encoder\_decoder | int | 5 | Количество блоков энкодера и декодера в архитектуре трансформера, который используется в этой модели. Доступные значения от 1 до 50. Правильный выбор количества энкодер/декодер блоков является важным параметром и часто подбирается экспериментально в зависимости от размера данных, доступных вычислительных ресурсов и требуемого качества модели |
| save\_confusion\_matrix | bool | true | Сохранение построенных матриц спутывания |
| figsize\_confusion\_matrix | int | width: "2600" height: "2600" font\_size: "14" dpi: "600" pad\_inches: "0" | Здесь определяются параметры отображения матрицы спутывания. width и height - размер изображения матрицы спутывания в пикселях. font\_size - размер шрифта на изображении матрицы. dpi - влияет на качество изображения. pad\_inches - величина отступов от построенной матрицы спутывания |
| classes | list | ["1\_Левая", "2\_Правая", ... "26\_Калибровка"] | Список названий классов, которые представлены в базе данных |

## Сообщения пользователю

В ходе работы библиотеки могут выдаваться следующие сообщения об ошибках:

1. BlurError. Указан неверный диапазон значений размытия.
2. ContrastError. Указан неверный диапазон значений контрастности.
3. CropPXError. Указан неверный диапазон обрезки в пикселях.
4. CropPercentsError. Указан неверный диапазон обрезки в процентах.
5. CustomException. Класс для всех пользовательских исключений.
6. FlipLRProbabilityError. Указано неверное значение вероятности отражения по вертикальной оси.
7. FlipUDProbabilityError. Указано неверное значение вероятности отражения по горизонтальной оси.
8. InvalidContentLength. Не определен размер файла для загрузки.
9. IsNestedCatalogsNotFoundError. Вложенные директории, где хранятся данные не найдены.
10. IsNestedDirectoryANotFoundError. Вложенная директория, для аудиофрагментов не найдена.
11. IsNestedDirectoryVNotFoundError. Вложенная директория, для видеофрагментов не найдена.
12. MixUpAlphaError. Указан неверный коэффициент для MixUp-аугментации.
13. PresetCFREncodeVideoError. Указан неподдерживаемый параметр, обеспечивающий определенную скорость кодирования и сжатия видео.
14. RotateError. Указан неверный диапазон значений угла наклона.
15. SRInputTypeError. Указан неподдерживаемый тип файла для распознавания речи.
16. SamplingRateError. Указана неподдерживаемая частота дискретизации речевого сигнала.
17. ScaleError. Указан неверный диапазон значений масштабирования.
18. TypeEncodeVideoError. Указан неподдерживаемый тип кодирования видео.
19. TypeMessagesError. Указан неподдерживаемый тип сообщения.
20. WindowSizeSamplesError. Указано неподдерживаемое количество выборок в каждом окне.

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 |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |