

ПОСТРОЕНИЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ ДЛЯ ОПРЕДЕЛЕНИЯ ЯЗЫКА ЖЕСТОВ

Рыжова Валерия Сергеевна (ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»)

Научный руководитель – доцент, кандидат технических наук Штенников Дмитрий Геннадьевич

(ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»)

Классификация языка жестов на основе глубокой нейросетевой модели. Классификация жестов в режиме реального времени.

Введение. Основой функционирования общества является взаимодействие между людьми, которое в свою очередь осуществляется с помощью различных видов коммуникаций. Однако существуют люди с ограниченными возможностями, которые не могут стать полноценными участниками общества из-за того, что не могут наладить общение с окружающими. К этой категории людей относятся глухонемые, глухие или плохо слышащие люди. Язык жестов, который они используют для общения, известен очень ограниченному кругу лиц. Поэтому люди с нарушением слуха испытывают трудности даже при выполнении таких повседневных задач как поход за покупками. Это является проблемой, но не такой неразрешимой, как может показаться на первый взгляд. Таким образом в рамках работы будет реализован классификатор языка жестов через web камеру ноутбука.

Основная часть. В рамках работы была реализована нейросетевая модель, переводящая язык жестов на английский. Работа велась с размеченным датасетом Sign Language Mnist Dataset английских жестов с информационного ресурса Kaggle, так как подходящего датасета с русскими жестами найдено не было. При подготовке к обучению мы изначально подготовили и разделили наш датасет на обучающую и тестовую выборку, а далее заполнили класс с данными метками.

Для построения глубокой нейросети была выбрана библиотека Pytorch. Была создана нейросетевая модель с 6 уровнями, содержащую в себе три сверточных слоя, за которыми следуют три полносвязных слоя. После первичного прохода по 5 эпохам, определена функция потерь и сделан вывод что для снижения функции потерь необходимо увеличить количество эпох до 10. Далее после обучения мы должны проверить нашу модель на контрольной выборке, которую модель не видела во время обучения. Так же экспортируем модель ONNX, проверяем экспортированную модель, а затем запускаем логический вывод для экспортированной модели. Таким образом мы можем сравнить результаты нашей оригинальной модели и экспортированной.

Ну и для большей наглядности была привязана камера при помощи библиотеки OpenCV, которая позволяет в режиме реального времени переводить жесты в буквы английского языка.

Вывод. В результате работы была построена нейросетевая модель, определяющая язык жестов с высокой точностью, а также попытка определения жестов режиме реального времени. Правда при этом точность сильно упала, так как буквы определяются правильно только при учете неподвижной руки в центре кадра и абсолютно однотонным, светлым фоном.

Рыжова В.С. (автор)

Подпись

Штенников Д.Г.(научный руководитель)

Подпись