Демидовский университет

Алгоритм распознавания жестов на видео

Подготовила: студент группы ИТ-21Мо Куликова Э. В.

Научный руководитель: к. ф.-м. н., доцент Лагутина Н. С.

Ярославль 2023

Цель исследования

Разработать алгоритм распознавания жестов рук на видео в реальном времени с использованием методов компьютерного зрения и машинного обучения.

Задача: разработать и реализовать алгоритм распознавания цифр и букв русского и американского жестового языка на видео в реальном времени.

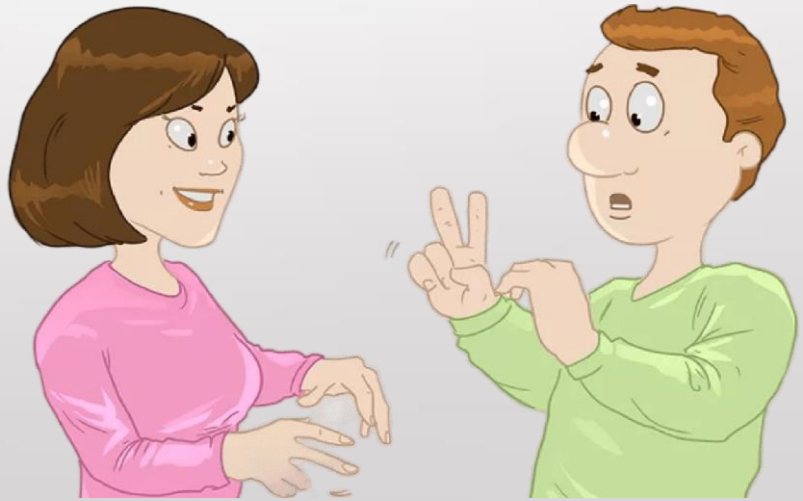
Подзадачи:

- Самостоятельное создание корпуса данных с цифрами и буквами жестового русского языка;
- Подбор корпуса данных с цифрами и буквами жестового американского языка;
- Обработка наборов данных для получения дополнительных характеристик, которые могут быть использованы для классификации цифр и букв;
- Проведение экспериментов по классификации цифр и букв русского и американского жестового языка;
- Определение качества работы разработанного алгоритма.

Описание предметной области

Демидовский университет

- Язык жестов является основным средством общения для глухих или слабослышащих людей;
- Язык жестов основан на жестах рук, движениях тела и выражении лица;
- Существуют значительные различия в жестовом языке даже внутри одной языковой группы;
- Не хватает универсального набора русского жестового языка для исследований в области компьютерного зрения и машинного обучения.



Описание предметной области

Демидовский университет

- Распознавание жестового языка (SLR) – сложная задача, особенно для динамических знаков;
- Существует два типа распознавания жестов рук: на основе ношения специальных устройств и на основе машинного зрения;
- Методы SLR на основе компьютерного зрения можно разделить на статические и динамические;
- Статистические знаки рассматривают как одно изображение, в то время как динамические знаки – видеопоследовательность кадров;
- Есть несколько ключевых проблем, связанных с распознаванием жестов рук в видео, таких как постоянное движение рук, изменяющиеся углы и перекрытия, изменчивость формы и размера ладони.

Обзор аналогов

Демидовский университет

№	Суть	Использование специальных устройств	Модель	Набор данных	Кол-во жестов	Кол-во экземпляров для обучения	Средняя F-мера
1	классификатор жестов для управления протезами кисти	да	RNN с использованием блоков долговременной кратковременной памяти (LSTM) и плотных слоев	Записан самостоятельно	5	100 000	0.8509
2	определения положения рук по картам изображений, полученных на основе данных глубины	нет	Модель множественного параллельного потока 2D CNN	1) Kaggle 2) First Person 3) Dexter	1) 10 2) 9 3) 7	1) 13 375 2) 98 842 3) 19 519	1) 1 2) 1 3) 0.92
3	распознавание жестов русского жестового языка	нет	рекуррентная сеть (LSTM)	Записан самостоятельно	35 000	10 000 видеофайлов	0.95

Сбор корпусов данных

Демидовский университет

Русский жестовый язык (RSL):

- Сбор добровольцев для записи видео русского жестового языка;
- Запись видео с жестами чисел от 1 до 10 с участием 19 добровольцев (10 женщин и 9 мужчин в возрасте 20-55 лет) ;
- Запись видео с жестами букв русского алфавита с участием 11 добровольцев (7 женщин и 4 мужчин в возрасте 20-55 лет) ;
- Разбивка видео на кадры и выбор лучших кадров;
- Классификация фреймов по классам для цифр (1-10) и букв (25 классов);
- Запись 2 дополнительных видео для проверки распознавания знаков в реальных условиях.

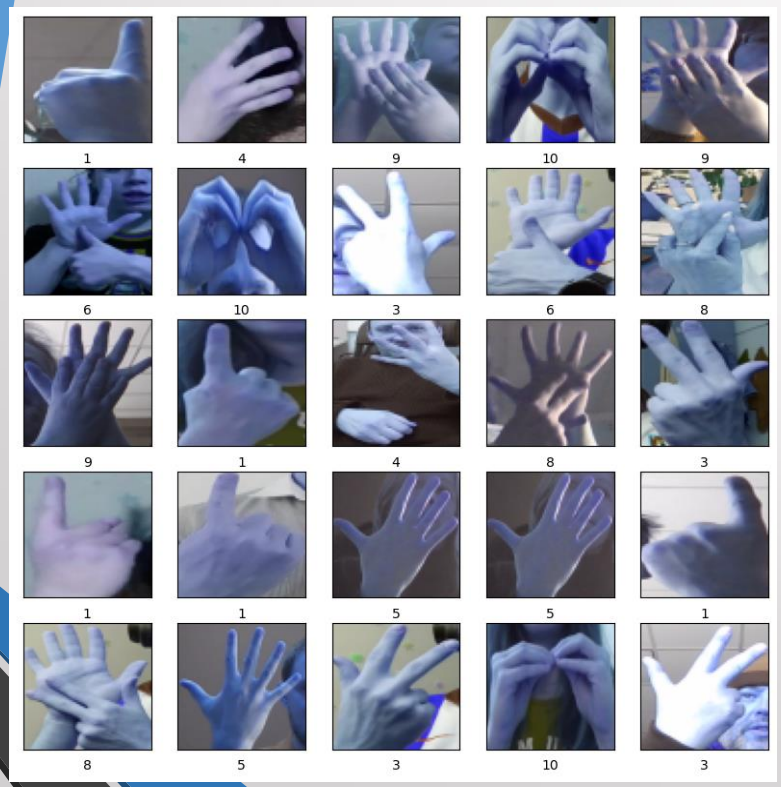
Американский жестовый язык (ASL):

- Цифры: содержит 10 классов по 500 цветных изображений рук на темном фоне с разрешением 400 x 400;
- Буквы: содержит 3000 цветных изображений жестов с разрешением 400 x 400 для каждой буквы.

Обработка корпусов данных

Использовалось решение **«MediaPipe Hand Landmarker»** позволяет обнаружить ориентиры рук на изображении. Его можно использовать для локализации ключевых точек рук и визуализации ориентиров.

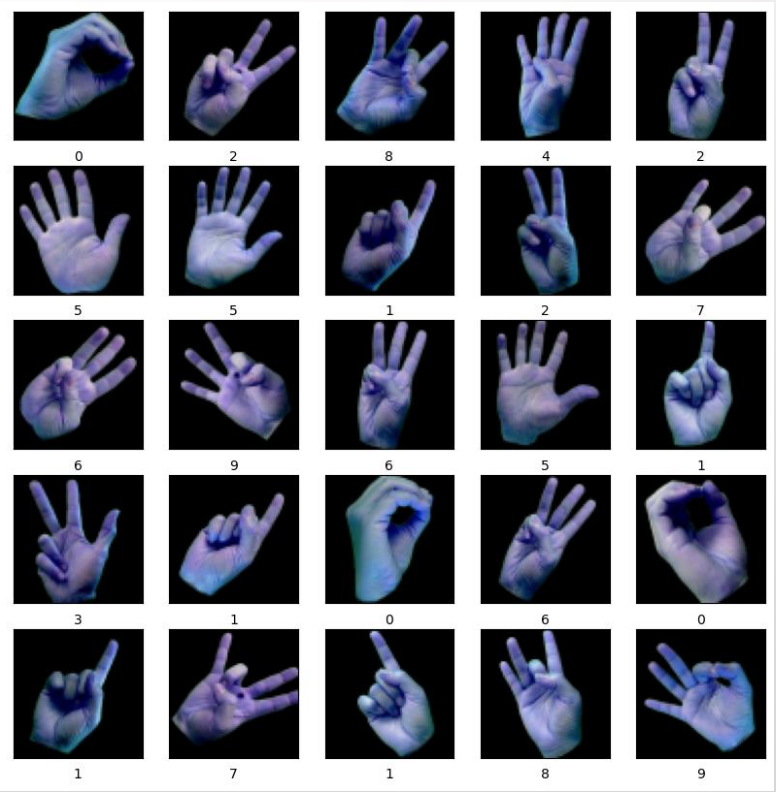
Пример жестов цифр RSL



Пример с ключевыми точками


```
{
  "num_hands": 2,
  "landmarks": [
    [
      1,
      0.9376944347895629,
      -1.6562921700824518e-7
    ],
    [
      0.8552957420378341,
      0.8466151352315685,
      -0.0012534725246950984
    ],
    [
      0.6261349278546047,
      0.7566050155486314,
      -0.0054916092194616795
    ],
    [
      0.4559098215631299,
      0.68231147092206,
      -0.009846813045442104
    ],
    [
      0.3920729425002743,
      0.5881747352426417,
      -0.013850336894392967
    ],
    ...
    [
      -0.29899818204940826,
      0.3388255268253013,
      -0.06701805721968412
    ]
  ]
}
```


Пример жестов цифр ASL



Алгоритм

Демидовский университет

Для классификации жестов рук по ключевым точкам рук и фото были выбраны четыре модели: MLP, 1D CNN, LSTM и 2D CNN. Применялся инструмент аугментации из-за небольшого размера корпуса данных.

Алгоритм:

1. Подключение решения «MediaPipe Hand Landmarker» в режиме одиночного обнаружения;
2. Захват видеопотока в реальном времени с веб-камеры (с применением методов OpenCV) ;
3. Первичная обработка каждого видеокадра для распознавания руки: отражение кадра вокруг оси Y, преобразование изображения BGR в RGB и передача его в конвейер «MediaPipe Hand Landmarker» для обнаружения ладони и получение координат ключевых точек;
4. Вторичная обработка каждого видеокадра для классификации жестов: кадрирование кадра в соответствии с точками ориентации руки, нормализация кадра и передача его в одну из обученных моделей для предсказания ;
5. Визуализация ориентиров обнаруженных рук и отображения предсказания на соответствующий жест;
6. Повторение шагов 3-5, пока пользователь не остановит режим распознавания;
7. Отпустить захват видеопотока и освободить ресурсы.

Результаты экспериментов

Демидовский университет

	Модель	RSL			ASL		
		Точность	Полнота	F-мера	Точность	Полнота	F-мера
Цифры	MLP	0.84	0.86	0.84	0.39	0.34	0.30
	1D CNN	0.80	0.82	0.80	0.10	0.18	0.11
	LSTM	0.77	0.72	0.70	0.14	0.15	0.12
	2D CNN	0.65	0.65	0.63	0.16	0.24	0.17
Буквы	MLP	0.41	0.50	0.42	0.33	0.41	0.30
	1D CNN	0.52	0.61	0.53	0.32	0.38	0.30
	LSTM	0.09	0.19	0.09	0.17	0.21	0.15
	2D CNN	0.17	0.24	0.17	0.00	0.04	0.00
Цифры и буквы	MLP	0.29	0.36	0.31	0.22	0.16	0.15
	1D CNN	0.22	0.28	0.23	0.12	0.09	0.09
	LSTM	0.15	0.14	0.13	0.08	0.07	0.06
	2D CNN	0.26	0.20	0.17	0.03	0.06	0.03

Заключение

Демидовский университет

- Создан корпус данных русского жестового языка;
- Разработан алгоритм с использованием четырех созданных нейросетевых моделей: MLP, 1D CNN, LSTM и 2D CNN, способный распознавать 10 цифр и 25 букв русского жестового языка;
- Применена аугментация данных для увеличения количества доступных выборок данных и повышения качества моделей машинного обучения;
- Оценена производительность алгоритма на реальных видео;
- Дальнейшее развитие алгоритма может включать повышение точности распознавания, создание более подходящих моделей классификации жестов и расширение его функциональности для других типов жестов;
- Имеет потенциал для применения в различных областях, связанных с компьютерным зрением, компьютерным управлением и социальной интеграцией, для снижения социальной изоляции и улучшения качества жизни людей с нарушениями слуха и речи.

Список использованных источников

Демидовский университет

- Miah A.S.M., Hasan M.A.M., Shin, J., Okuyama, Y., Tomioka Multistage Spatial Attention-Based Neural Network for Hand Gesture Recognition // Computers 2023. Vol. 12, no. 13;
- Abdallah M.S., Samaan, G.H., Wadie A.R., Makhmudov F., Cho Y. Light-Weight Deep Learning Techniques with Advanced Processing for Real-Time Hand Gesture Recognition // Sensors 2023. Vol. 23, no. 2;
- Toro-Ossaba A., Jaramillo-Tigreros J., Tejada J.C., Peña A., López-González A., Castanho R.A. LSTM Recurrent Neural Network for Hand Gesture Recognition Using EMG Signals // Applied Sciences 2022. Vol. 12, no. 19;
- Noreen I., Hamid M., Akram U., Malik S., Saleem M. Hand Pose Recognition Using Parallel Multi Stream CNN // Sensors 2021. Vol. 21, no. 24;
- М. Г. Гриф, Р. Элаккия, А. Л. Приходько, М. А. Бакаев, Е. Раджалакшми Распознавание русского и индийского жестовых языков на основе машинного обучения // Системы анализа и обработки данных 2021. Том 83, № 3. – С. 53-74.