

# Проект по предсказанию лиц и жестов с веб-камеры.

## Подтема

Предсказание жестов с веб-камеры и программная реакция на жесты.





## Дмитрий Яковлев

Закончил курсы повышения квалификации в МГТУ им. Баумана.

Закончил факультет искусственного интеллекта GeekBrains..

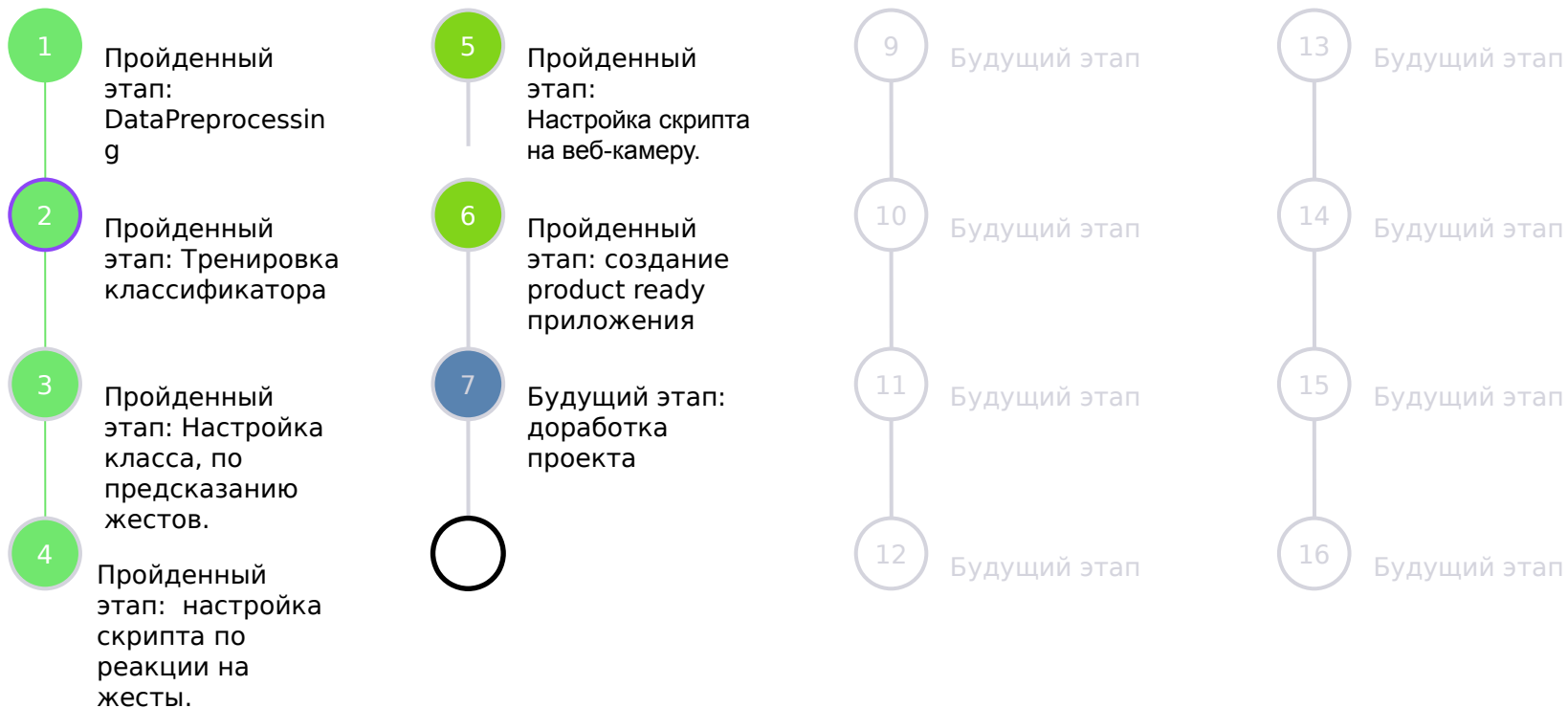
Закончил Сибирскую аэрокосмическую академию по специальности «Системы управления ракетно-космическими объектами и комплексами летательных аппаратов»

Немного о себе.

- Инженер по работе с клиентами и проектировщиками на предприятии, производящем высоковольтное оборудование.



# План проекта

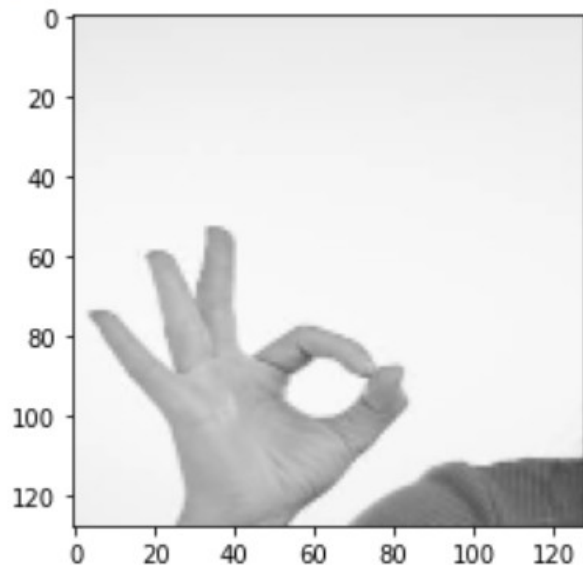


## Достигнутые цели

```
59]: fcd.run('/kaggle/input/gestures/8.jpg')
```

Загруженное изображение:

Предсказанный жест: 09\_c



Все поставленные цели выполнены:

- Освоен механизм получения тренировочных и валидационных данных из многоструктурного хранилища изображений и их аугментации с помощью метода Dataset (Pytorch).
- Освоено дообучение и применение предобученных моделей на языке Pytorch.
- Освоена работа с методами детекции лиц и их применения для распознавания жестов.
- Освоена работа с веб-камерой в методах компьютерного зрения.



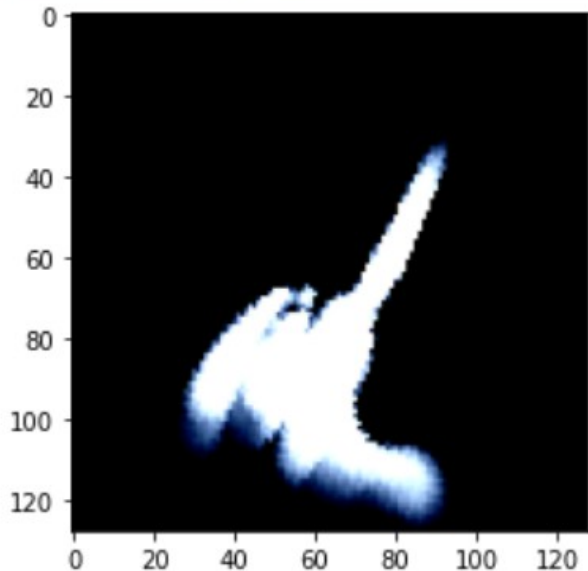
## Решение задачи / План работы

- Исследование данных на платформе Kaggle.
- Исследование рабочего датасета.
- Разработка классификатора.
- Разработка класса по распознаванию жестов.
- Разработка рабочего скрипта на основе видео с веб-камеры.



## Трудности

```
torch.Size([64, 3, 128, 128]) torch.Size([64])  
Ground Truth 04_fist_moved
```

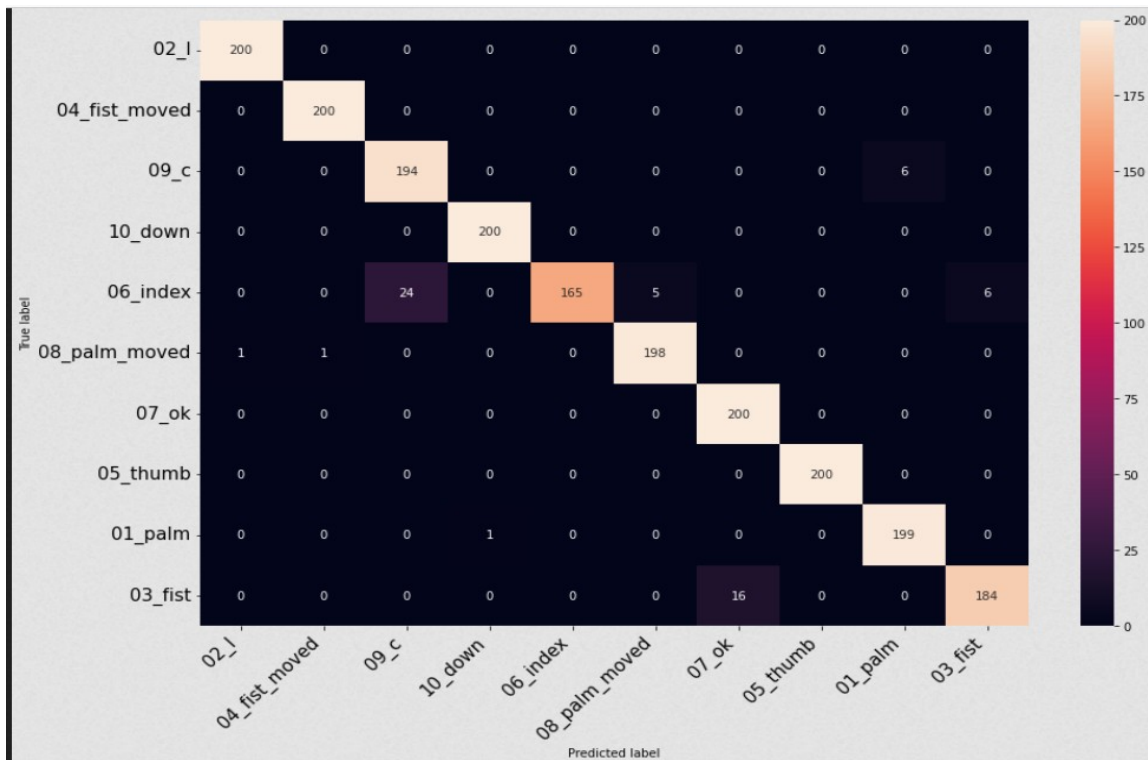


Сложная многоуровневая структура обучающего датасета. Сложно было все скомпоновать и упаковать методом Dataset . Трудные для распознавания картинки даже для человека. Есть близкие по начертаниям картинки из разных классов. Но мне было интересно генерировать датасеты для нейросети из такой сложной структуры.

Да уж! Мне самому трудно разобрать что это за жест.



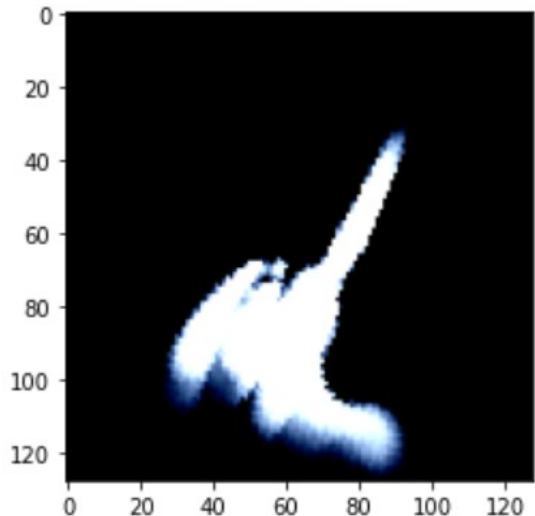
## Выбор классификатора



Учитывая сложность задачи, в качестве модели классификатора решил использовать предобученную модель. Учитывая, что выбран огромный обучающий датасет, выбрал относительно «легкую» модель ResNet18. Дообучил сначала классификационные слои а затем последние 9 слоев. Ассигу на валидационной выборке получил выше 96%. Из матрицы ошибок видно что не очень хорошо отличает шестерку от девятки и семерку с тройкой. Думаю что в том числе это связано с ротацией до 30 градусов изображений в аугментации. Видимо угол ротации надо снижать. Ну и конечно с тем, что модель выбрана слишком простая.

## Предложения по проекту

```
torch.Size([64, 3, 128, 128]) torch.Size([64])
Ground Truth 04_fist_moved
```



В первую очередь надо подобрать более качественный датасет с жестами, например из MNIST. Как видно из картинки, даже человеку бывает трудно разобрать что за в нем жесты. Выбрать из зоопарка моделей более подходящий вариант чем устаревшая ResNet18. Более тонко настроить аугментацию.

Да уж! Мне самому трудно разобрать что это за жест.