



ОБЗОР ИНСТРУМЕНТОВ ПОДГОТОВКИ ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ

Mike Smirnov

Этапы подготовки данных

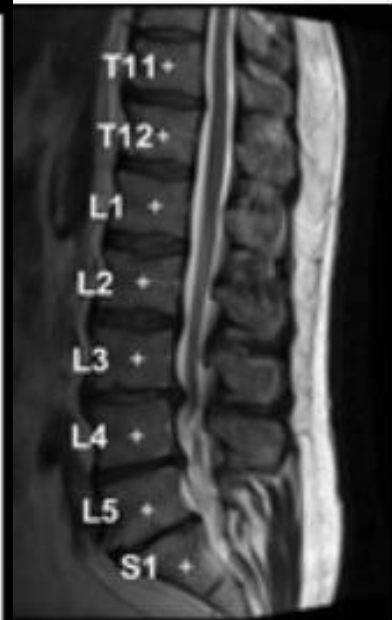
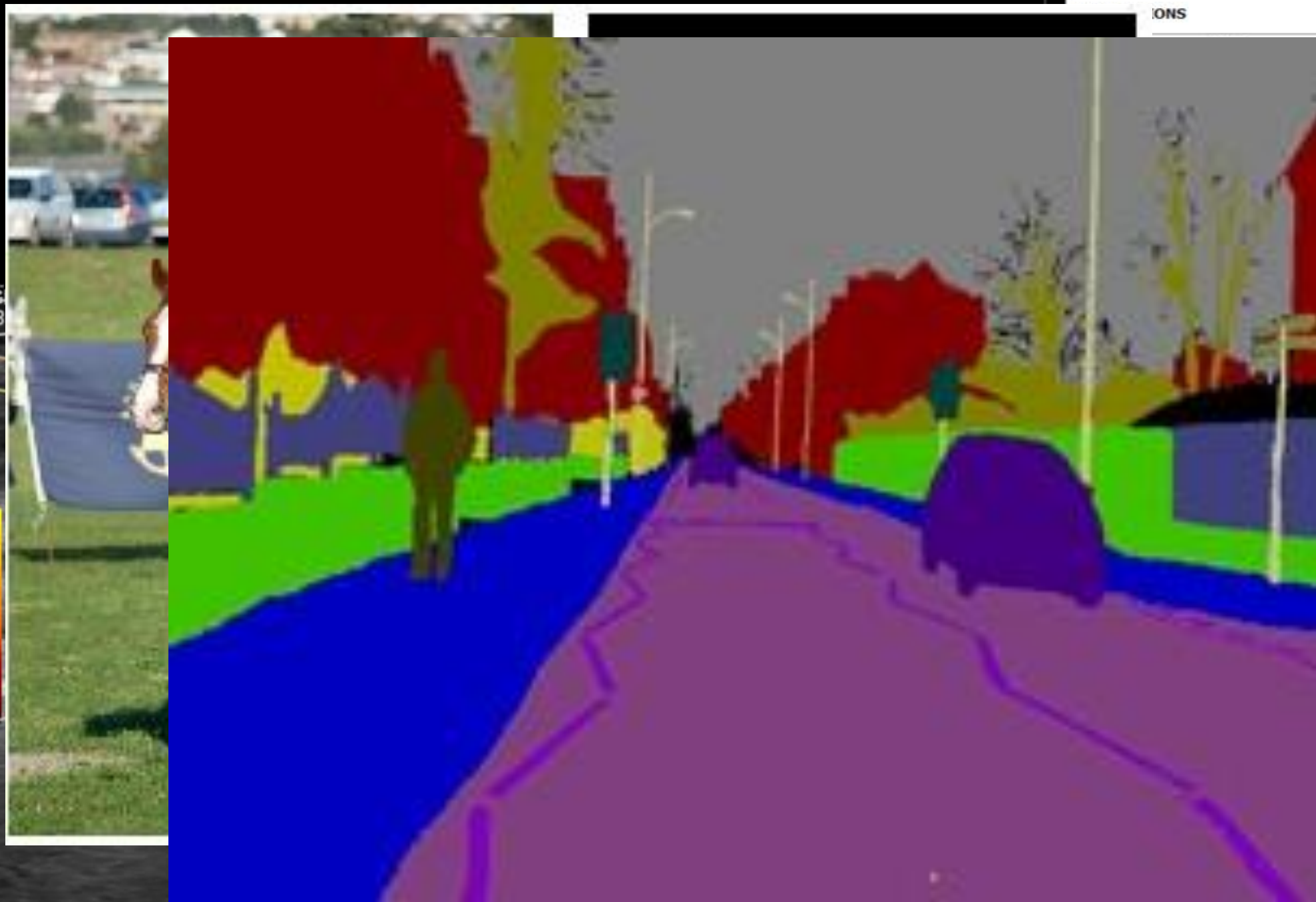
- Сбор данных
- Разметка данных
- Доступ к данным \ Хранение данных
- Подготовка данных для обучения конкретного алгоритма
- Визуализация данных

Сбор данных

- Готовые data-set
- Съёмка \ подготовка данных самому
- Сбор данных из интернета
- Симуляция данных

Разметка данных

- Выделение объектов
 - 2DBox
 - 3DBox
 - Другие фигуры
- Аннотация объектов
- Семантическая сегментация



LOG
S8

Разметка данных





- Cityscapes
- Vatic
- Crowd sourcing platforms
 - Yandex Toloka
 - Spare five

Annotate every object, even stationary and obstructed objects, for the entire video.

Instructions

+ New Object



- ▼ **Car 4** 
 - Outside of view frame
 - Occluded or obstructed
- ▼ **Person 3** 
 - Outside of view frame
 - Occluded or obstructed
- ▼ **Bicycle 2** 
 - Outside of view frame
 - Occluded or obstructed
- ▼ **Car 1** 
 - Outside of view frame
 - Occluded or obstructed

⏮ Rewind

▶ Play



⚙ Options

✓ Save Work

Доступ к данным \ Хранение данных

- Файловая система (изображение + json \ xml, h5 file)
- Базы данных
- Специализированные системы

Подготовка данных для обучения алгоритма

- OpenCV
- RapidJson

Задание на практику

- Установить
 - Cityscapes
 - <https://www.cityscapes-dataset.com>
 - <https://github.com/mcordts/cityscapesScripts>
 - Caffe
 - <https://github.com/BVLC/caffe>
- Найти\сделать картинки с дорог Нижнего Новгорода
 - Разметить машины
- Обучить сетку на Cityscapes data set
- Протестировать на сете из Нижнего Новгорода
- Добиться 70% распознавания машин на сете из нижнего Новгорода

