

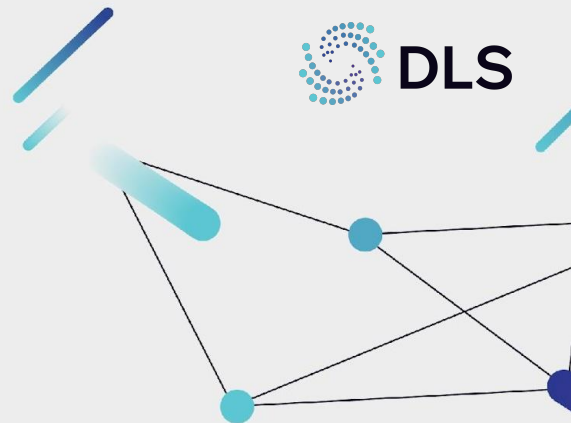


Deep  
Learning  
School

Задача детекции.  
Введение.

# Общий план лекций

1. Введение в задачу детекции:
  - 1.1. Постановка задачи
  - 1.2. Применение
  - 1.3. Детекторы до нейросетей
  - 1.4. Общий обзор нейросетевых детекторов
  - 1.5. Популярные датасеты
2. Базовое устройство детекторов и теоретический минимум:
  - 2.1. Как выглядит общая структура детектора
  - 2.2. Как детектор делает предсказания
  - 2.3. Метрики
  - 2.4. Как обучается детектор
  - 2.5. Аугментации
3. Обзор самых популярных архитектур начиная от R-CNN до YOLOv12

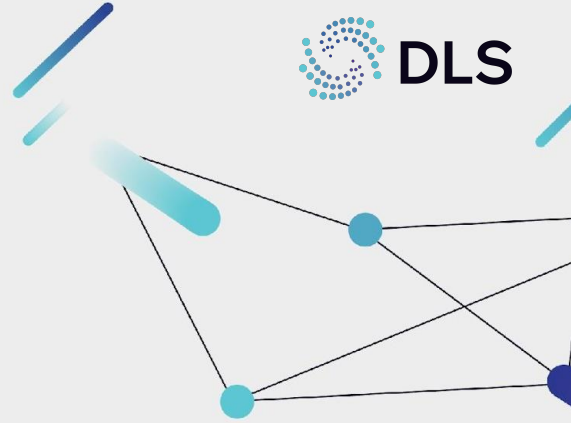


# План на сегодня

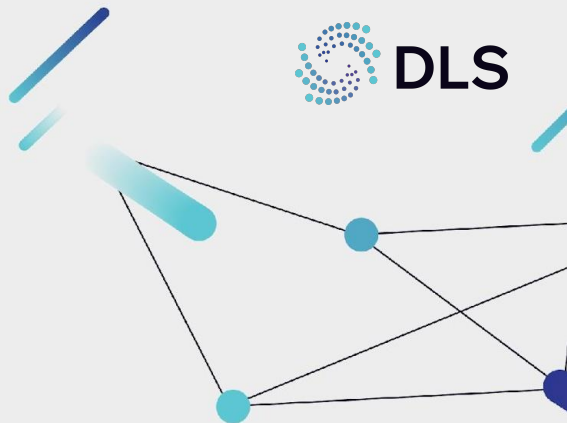
1. Узнаем отличия детекции от других задач машинного обучения.
2. Посмотрим на примеры использования детекторов в реальной жизни.
3. Ответим на вопрос: “Как задачу детекции решали до изобретения нейросетей?”
4. Поговорим о нейросетевых детекторах.
5. Перечислим самые популярные датасеты для детекции.



# Recap



# Recap. Classification.



Представим, мы обучили модель  
многоклассовой классификации.

Спросим у модели, что  
изображено на картинке?

# Recap. Classification.



Представим, мы обучили модель  
многоклассовой классификации.

Спросим у модели, что  
изображено на картинке?

**Собака**



# Recap. Classification.



Представим, мы обучили модель  
многоклассовой классификации.

Спросим у модели, что  
изображено на картинке?

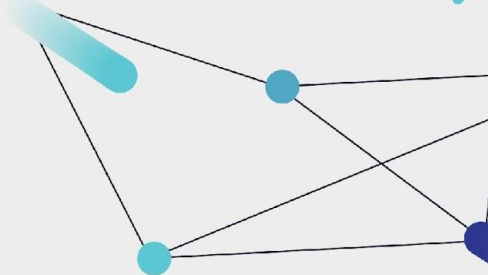
**Собака... Или фрисби?**

Но мы не можем найти и собаку и  
фризби.

# Recap. Multi-label classification.



А если взять multi-label  
классификацию?



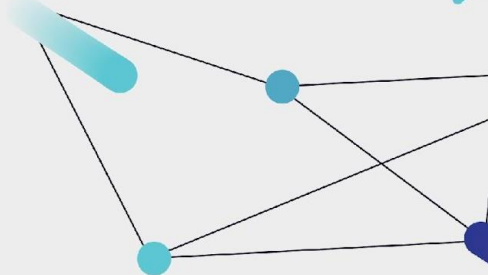


# Recap. Multi-label classification.

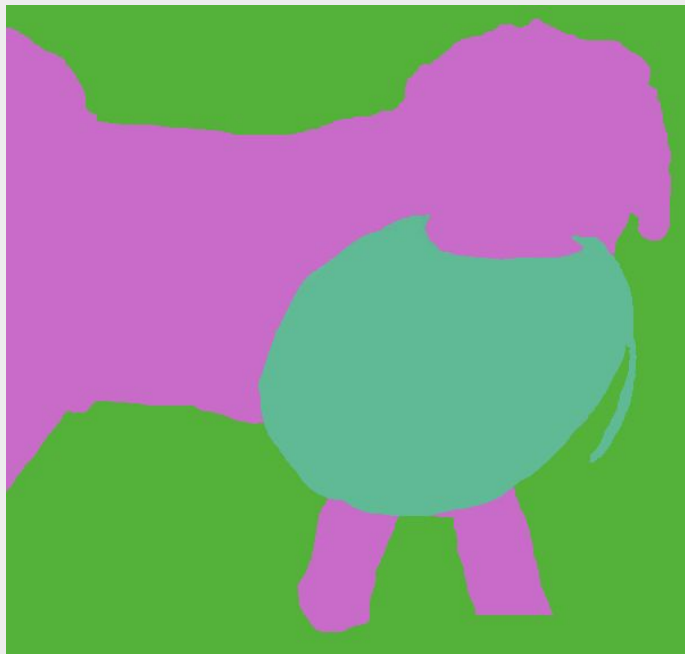


А если взять multi-label  
классификацию?

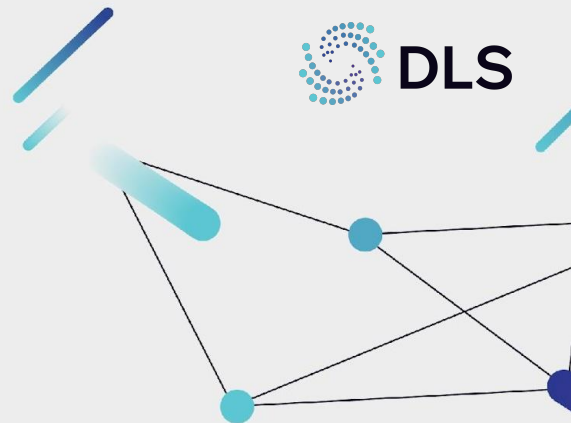
Получим:  
**dog : 0.8**  
**frisbee : 0.7**



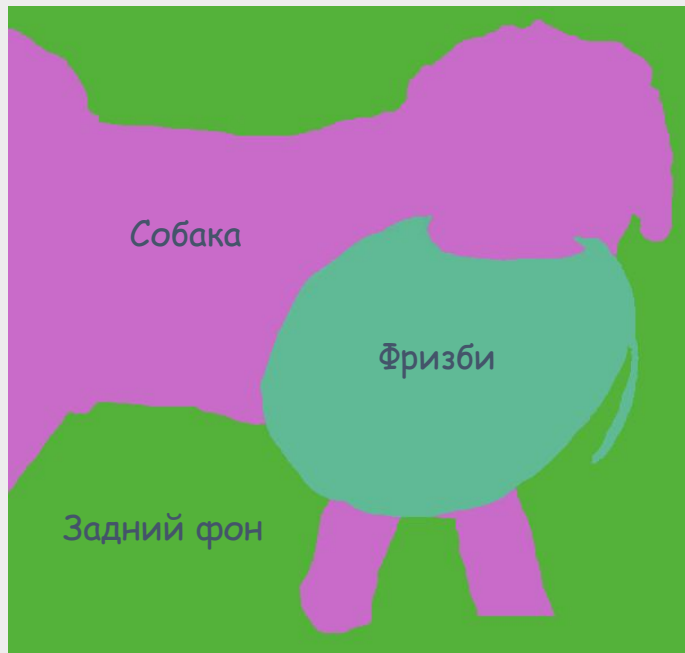
# Ресар. Семантическая сегментация.



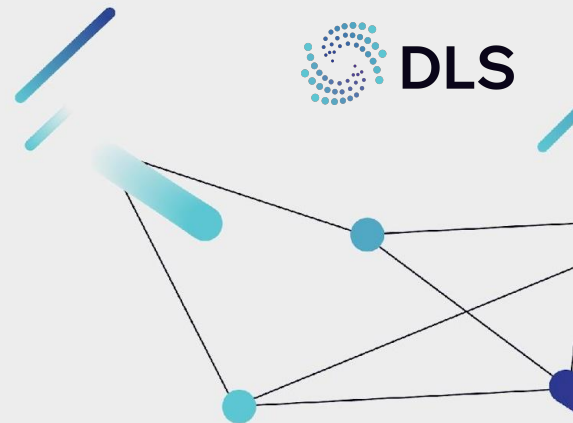
Семантическая сегментация позволяет определить все объекты на картинке.



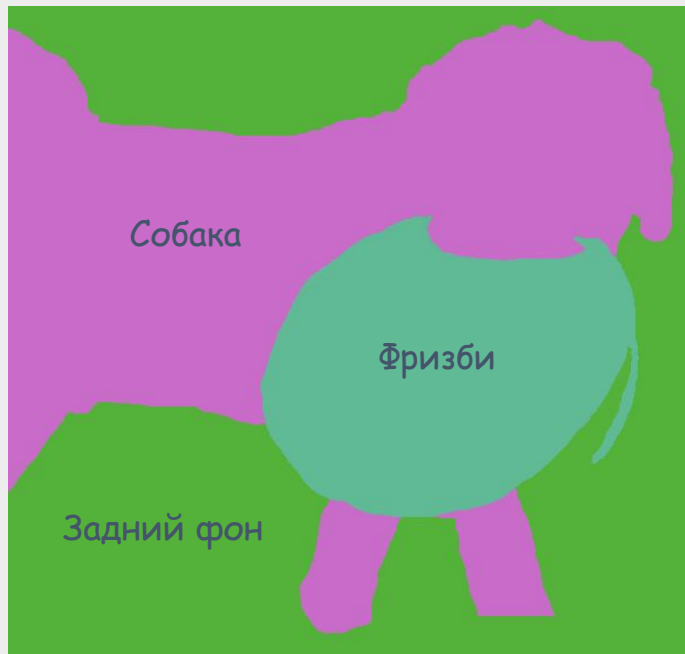
# Ресар. Семантическая сегментация.



Семантическая сегментация позволяет определить все объекты на картинке.

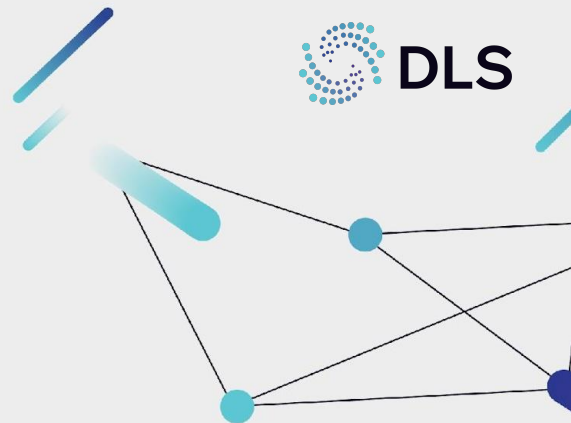


# Ресар. Семантическая сегментация.



Семантическая сегментация позволяет определить все объекты на картинке.

Но у неё есть один недостаток.



# Ресар. Семантическая сегментация.



Возьмем картинку посложнее.

Как будет выглядеть маска семантической сегментации?

# Ресар. Семантическая сегментация.



А что если нам нужно узнать сколько собак на картинке?

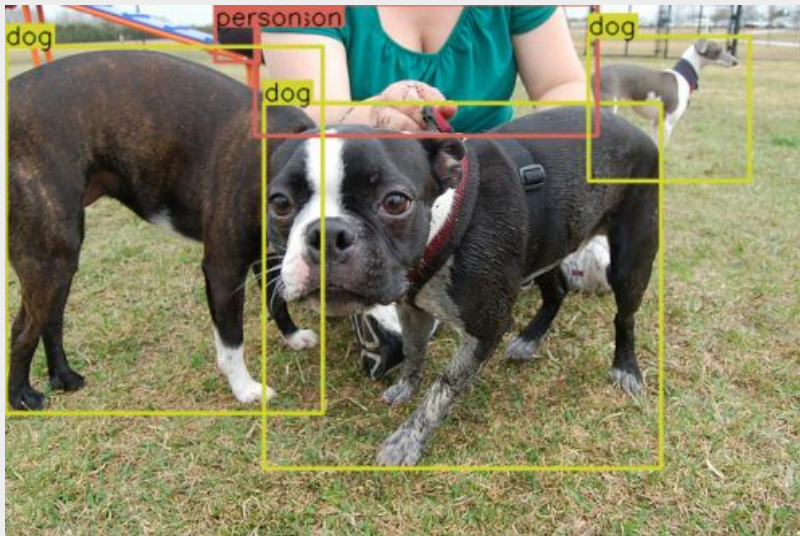


# Задача детекции



Детекция позволяет определить не только класс объекта, но и узнать где он находится на изображении.

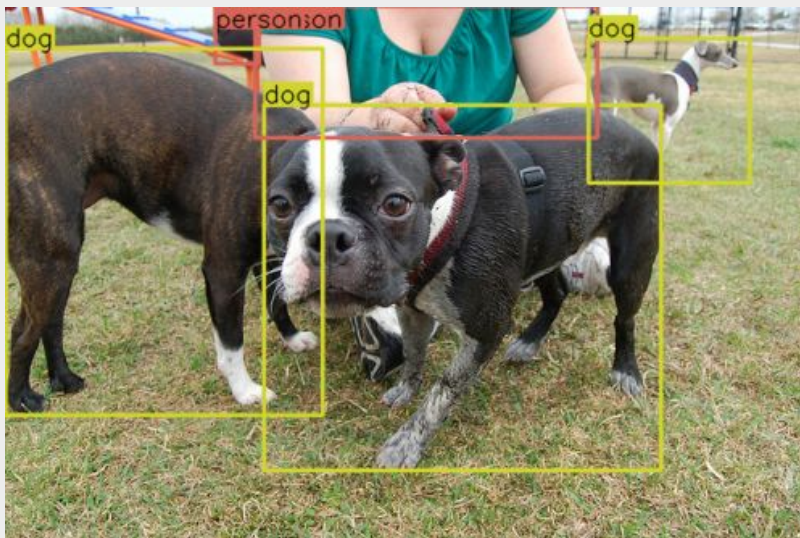
# Задача детекции



Детекция позволяет определить не только класс объекта, но и узнать где он находится на изображении.

Рамка, вокруг объекта называется bounding box (bbox).

# Задача детекции



В отличие от других задач, в детекции выход всегда разный:

На картинке может быть сразу несколько объектов, а может не быть ни одного.

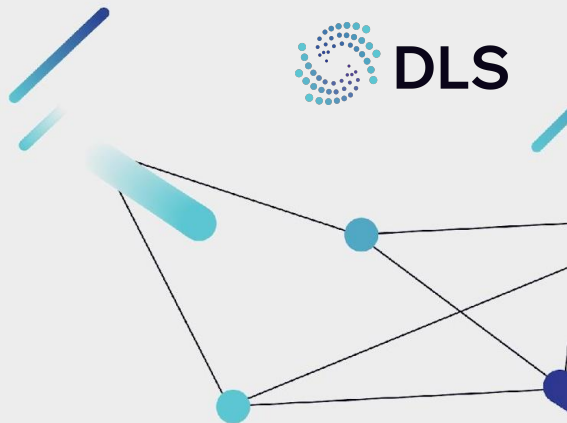
Один из главных вопросов в задаче детекции: Как учесть разное количество ббоксов при построении модели?

# Bounding Boxes. Описание.



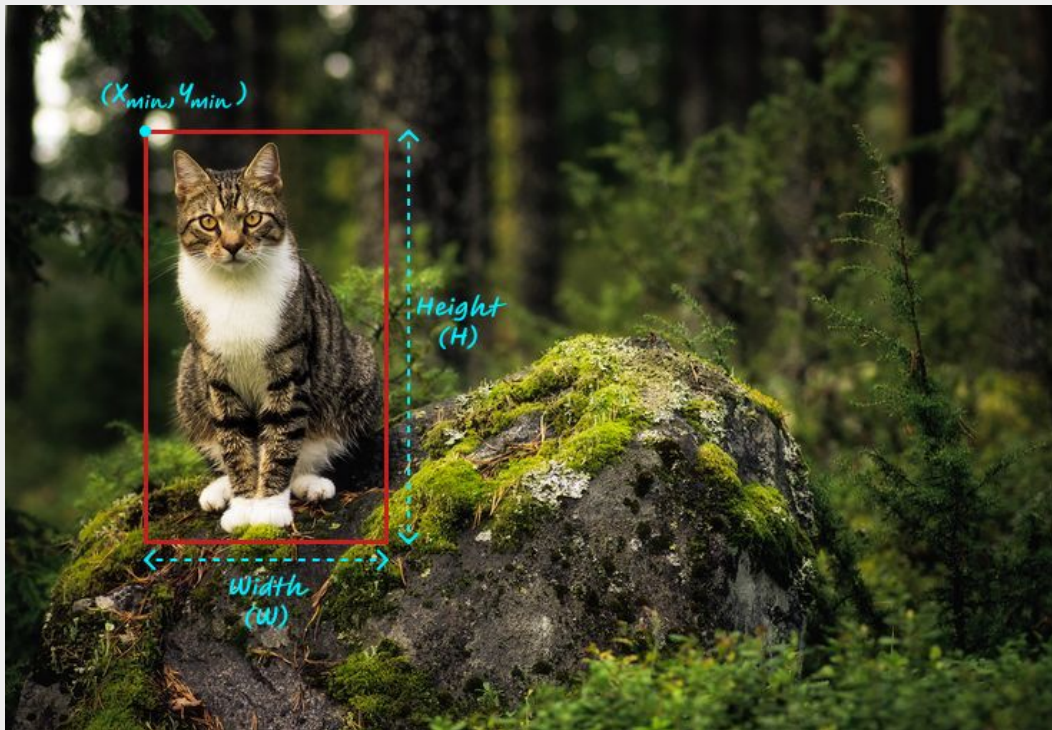
За начало координат принято выбирать верхний левый угол.

Исключение - `matplotlib`, там используется нижний левый угол.





# Bboxes. COCO format.



В COCO формате, bounding box  
описывается как:  
 $\text{Bbox} = [\text{Xmin}, \text{Ymin}, \text{W}, \text{H}]$

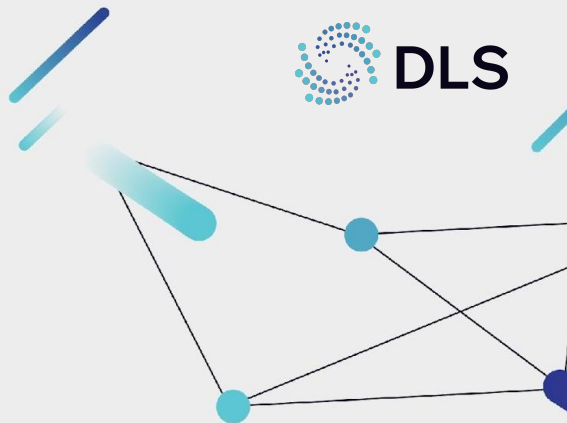
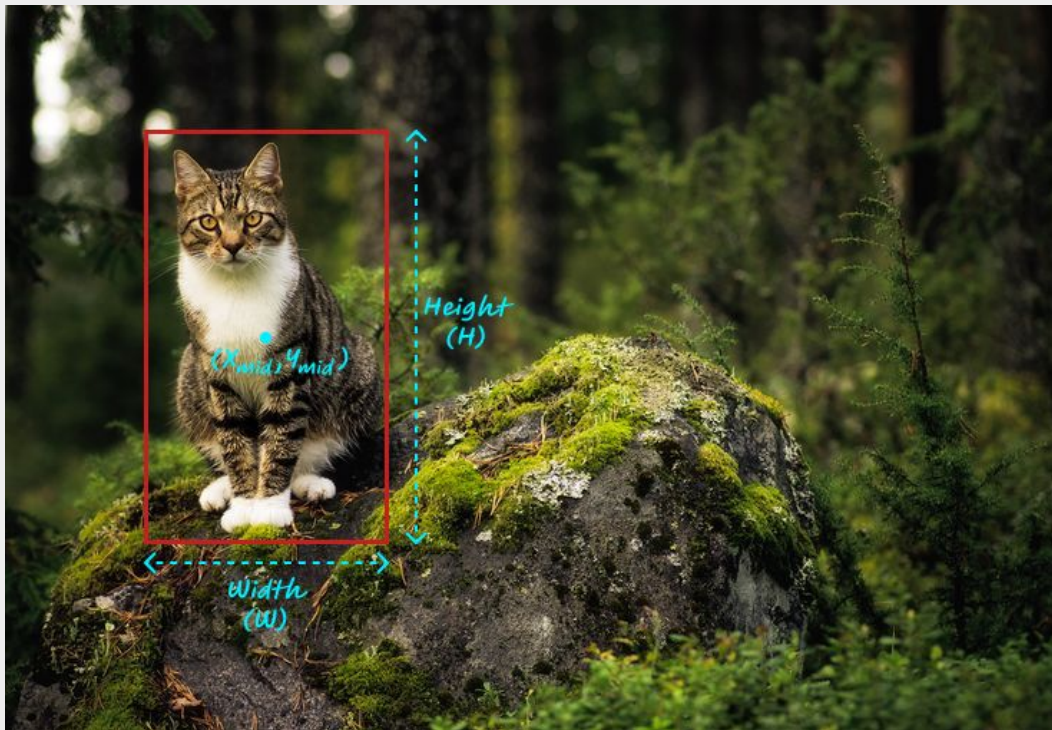
# Bboxes. Pascal VOC format.



В VOC формате, bounding box  
описывается как:  
 $Bbox = [Xmin, Ymin, Xmax, Ymax]$



# Bboxes. YOLO format.



В YOLO формате, bounding box описывается как:

Bbox = [Xmid, Ymid, W, H]

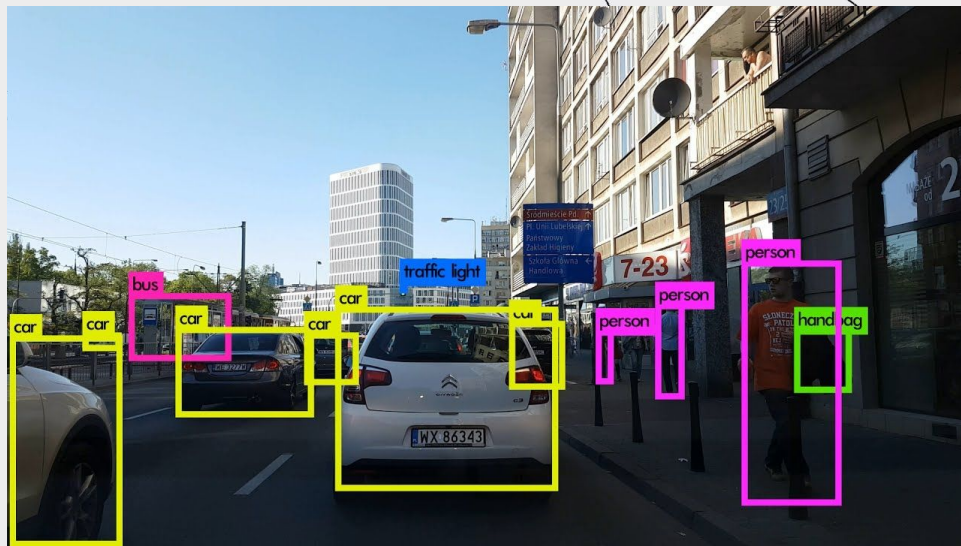
# Применение детекторов в ЖИЗНИ



# Применение. Self-driving cars.

Одна из самых популярных областей для применения детекторов: автономный транспорт.

Тут необходимо распознавать все вокруг: машины, знаки, людей и тд.



Source: [https://www.researchgate.net/publication/358106838\\_Perception\\_and\\_Planning\\_in\\_Autonomous\\_Car](https://www.researchgate.net/publication/358106838_Perception_and_Planning_in_Autonomous_Car)

# Применение. Self-driving not only cars.

При этом автономный транспорт не ограничивается автомобилями.

- Driverless tractors
- Driverless trains
- Driverless drones
- И тд



Source: CNH Industrial/Case IH



# Применение. Контроль безопасности.

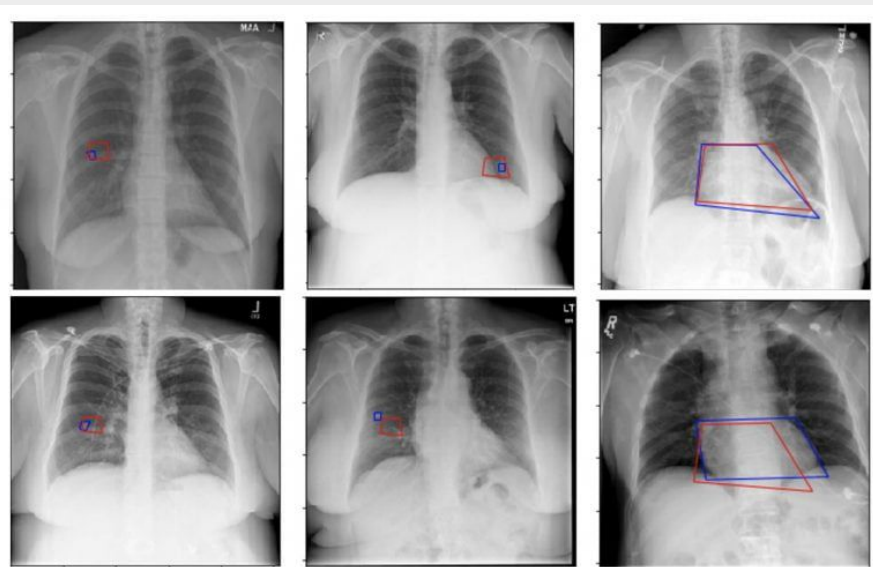


Можно автоматически  
проверять технику  
безопасности на производстве:

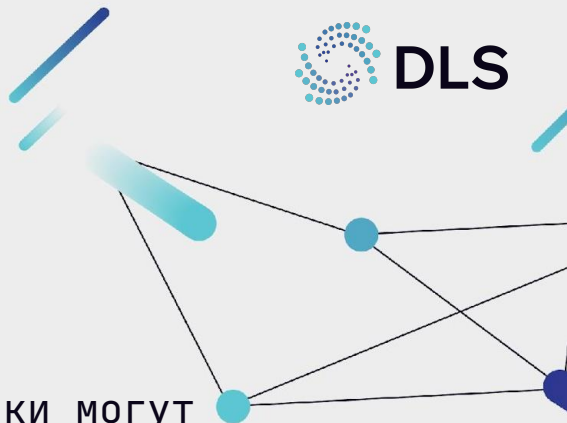
- Внешний вид рабочего
- Контроль безопасных зон
- И тд.

Source: [https://www.researchgate.net/figure/Examples-of-identified-danger-zones-and-different-safety-statuses-of-workers-The-accuracy\\_fig1\\_334128626](https://www.researchgate.net/figure/Examples-of-identified-danger-zones-and-different-safety-statuses-of-workers-The-accuracy_fig1_334128626)

# Применение. Медицина.



Системы-помощники могут анализировать снимки и отправлять врачу на анализ все подозрительные участки.





# Применение. Ретейл.



Source: <https://www.labelvisor.com/enhancing-retail-analytics-with-yolov8-object-detection/>

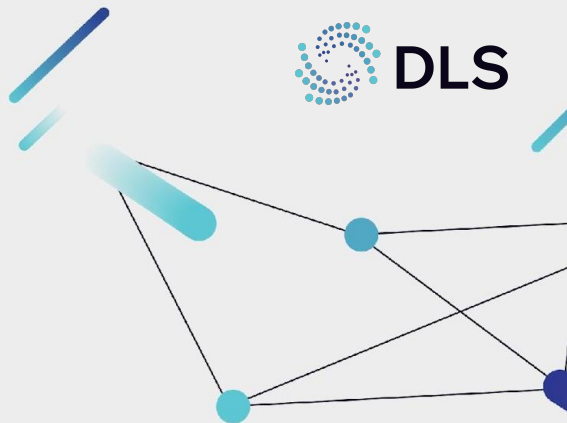
Области применения:

- Автоматический поиск вещей по фото
- Контроль качества товара на полках
- Контроль количества товара на полках

# Традиционные детекторы

# Как раньше решали задачу детекции?

1. HOG (Histogram of oriented gradients)
2. SIFT (Scale-Invariant Feature Transform)
3. Haar-каскады



# Как раньше решали задачу детекции?

1. HOG (Histogram of oriented gradients)  
Работает на основе анализа градиентов изображения и гистограмм направлений.
2. SIFT (Scale-Invariant Feature Transform)
3. Haar-каскады

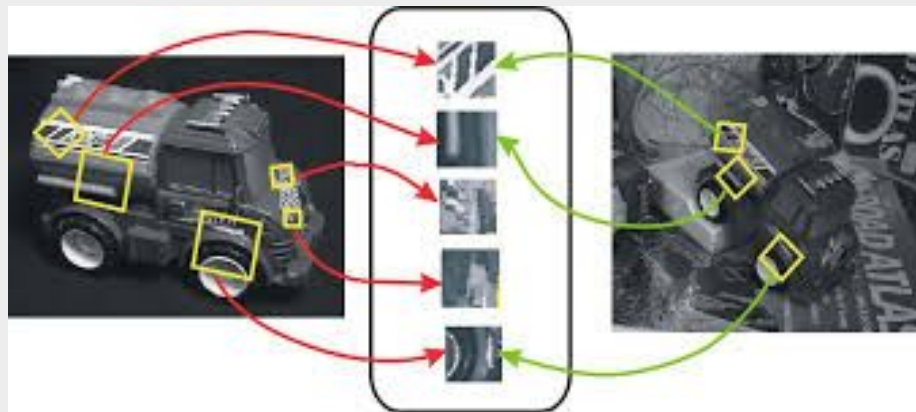


# Как раньше решали задачу детекции?

1. HOG (Histogram of oriented gradients)
2. SIFT (Scale-Invariant Feature Transform)

Алгоритм для поиска ключевых точек и составления их признакового описания.

3. Haar-каскады

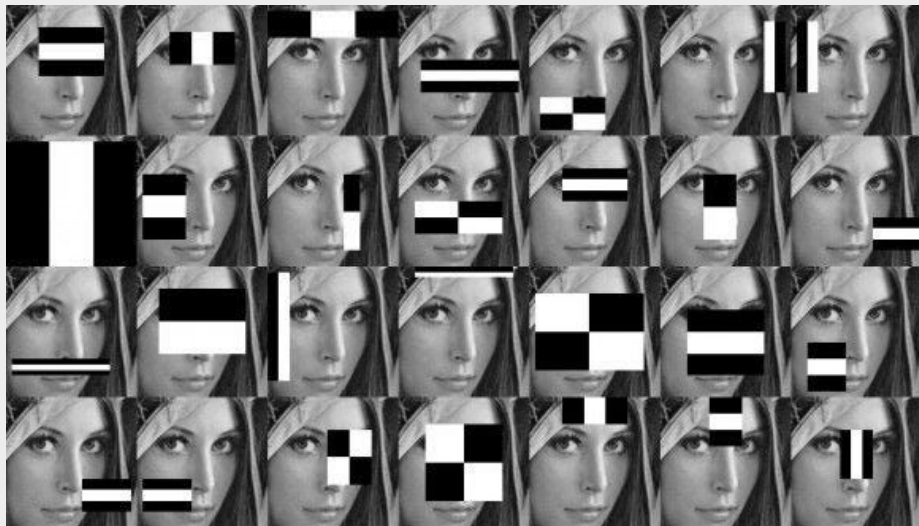


# Как раньше решали задачу детекции?



1. HOG (Histogram of oriented gradients)
2. SIFT (Scale-Invariant Feature Transform)
3. Haar-каскады

Метод, основанный на основе анализа контрастов изображения с помощью простых признаков (Хаара).



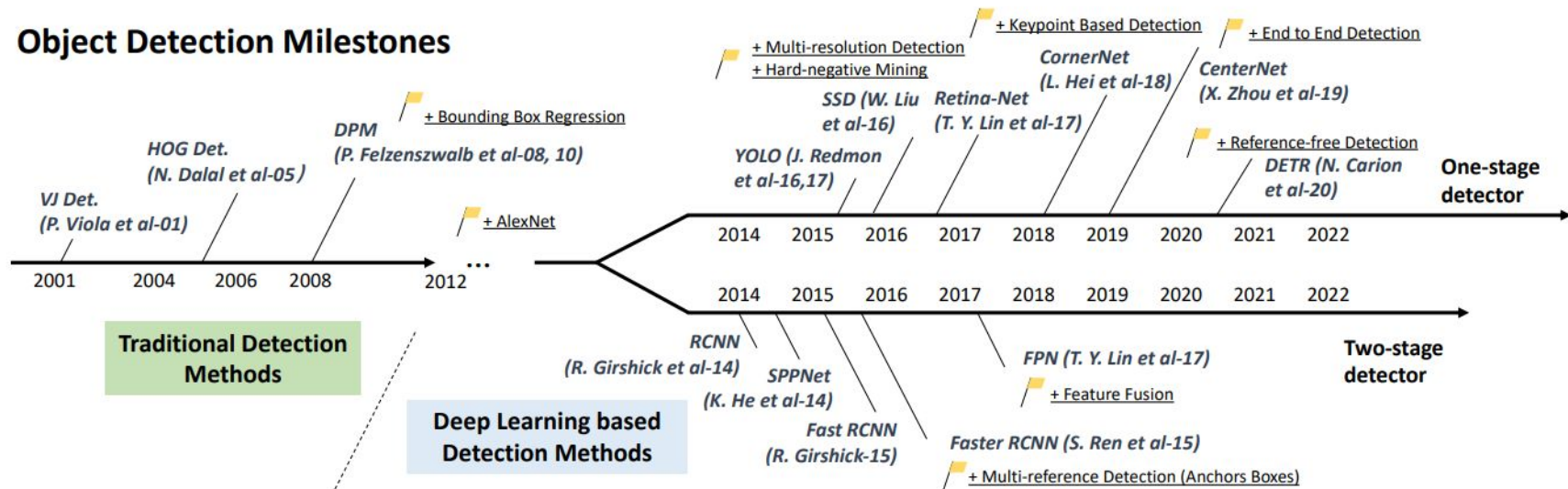


# Modern Detectors Overview



# Нейросетевые подходы

## Object Detection Milestones



# Datasets

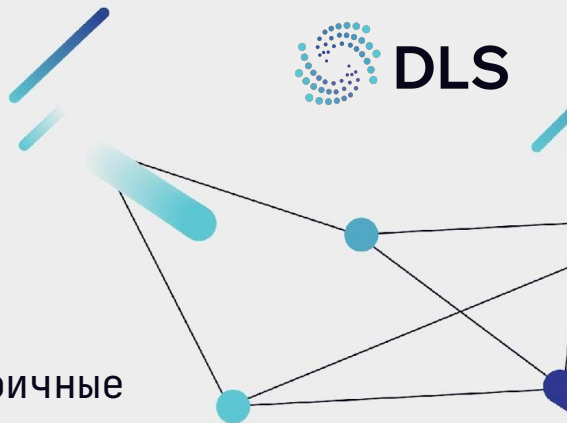
# Датасеты для детекции

Датасеты-бенчмарки:

- Pascal VOC 2012
- COCO 2017
- Open Images

Датасеты, специфичные  
для задачи:

- KITTI, BDD100K
- WiderFace
- DOTA
- CrowdHuman



# Датасеты. Pascal VOC 2012.

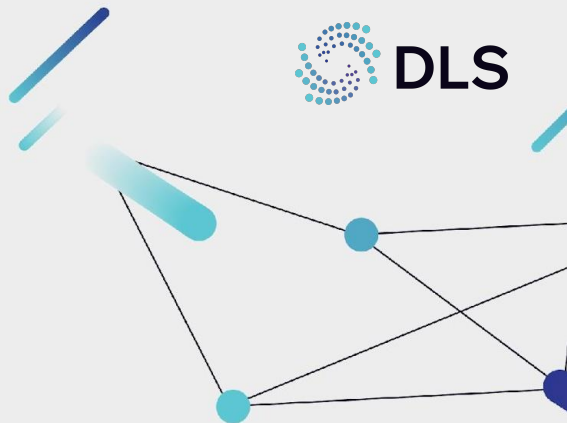


Датасет для задач: detection, instance and semantic segmentation.

Всего 21 класс.

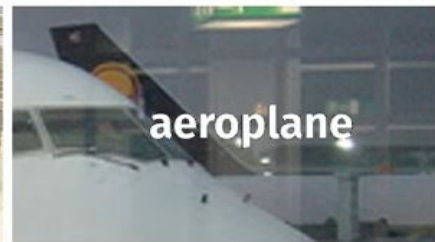
Train part: 2,913 , test: 1,456 , val: 1,449 images

Размеченных объектов: 19,694





PASCAL VOC 2012 · 21 classes



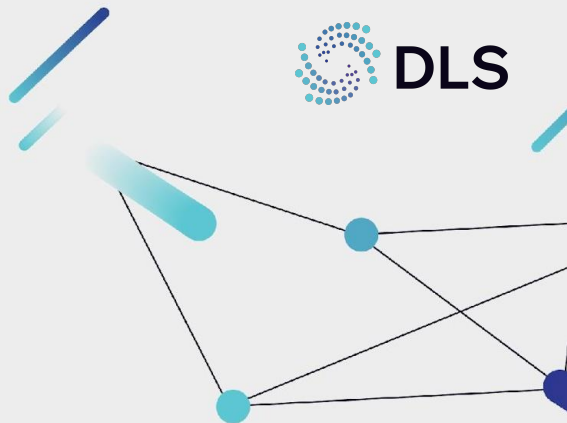
# Датасеты. COCO 2017.

Датасет для задач: detection, instance and semantic segmentation.

Всего 80 классов.

Train part: 118,287, test: 40,670, val: 5,000 images

Размеченных объектов: 2,099,063





# Датa

COCO 2017 · 80 classes

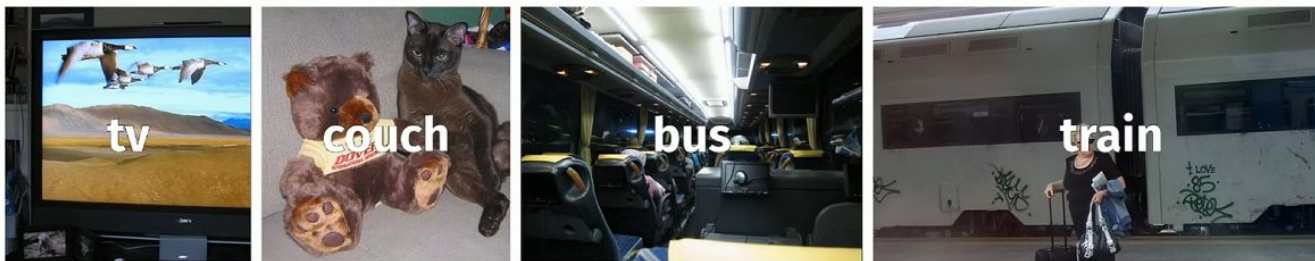


Датасет д  
segmentat

Всего 80

Train par

Размеченн



# Датасеты. Open Images.



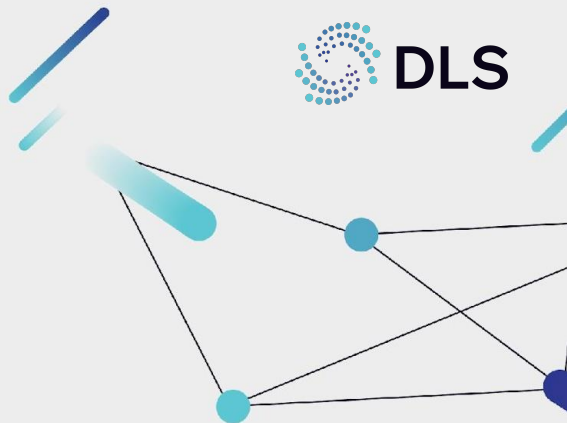
Всего существует 7 версий датасета.  
Последняя вышла в 22 году.

Датасет для задач: detection, instance and semantic segmentation, visual relationships and localized narratives.

Всего 600 классов.

Train part: 1,743,042 , test: 125,436 , val: 41,620 images

Размеченных объектов: 16M





# Дата

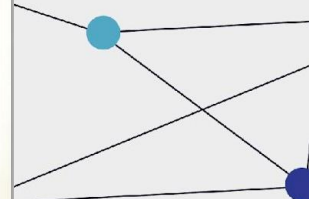
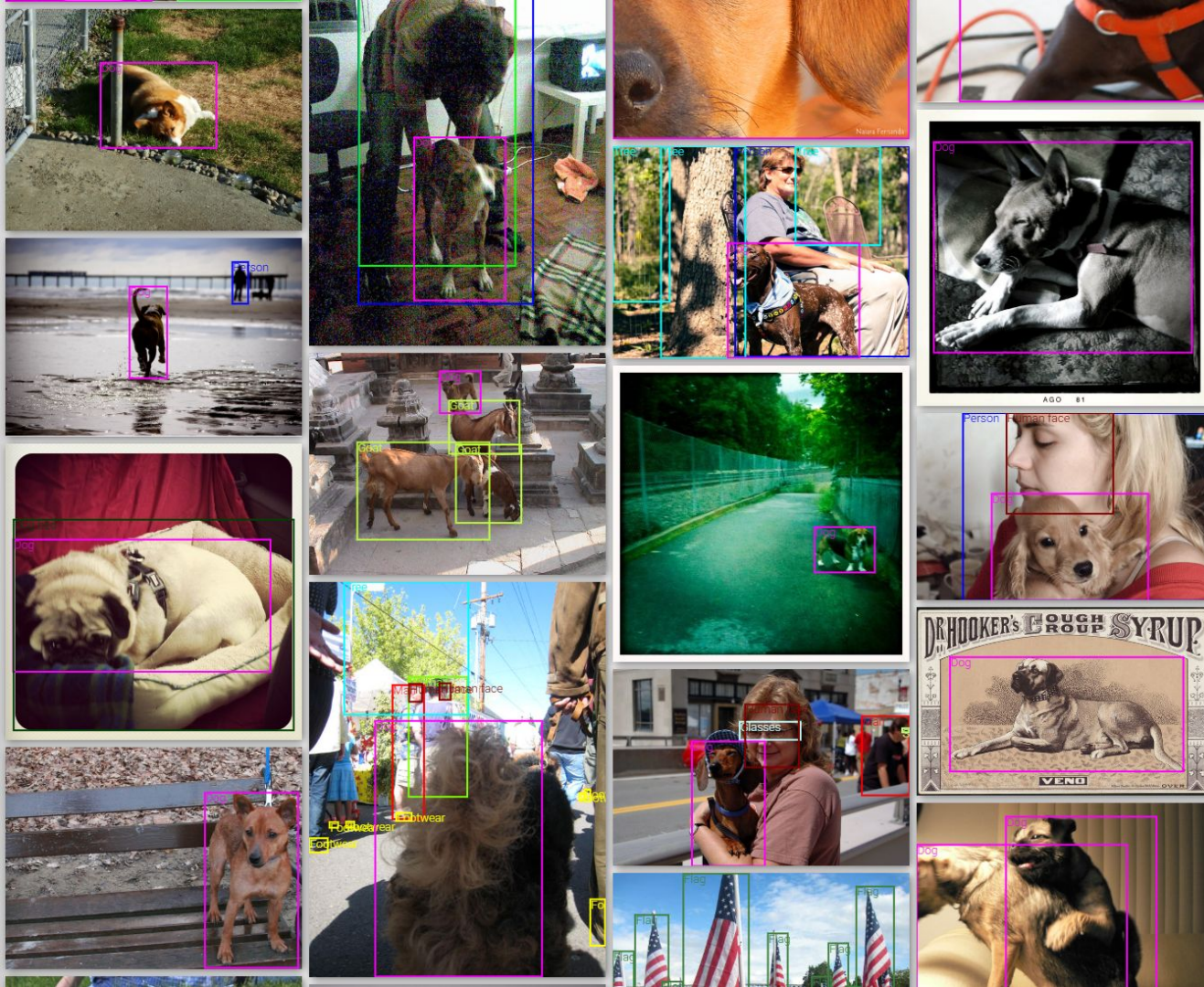
Всего су  
Последня

Датасет д  
segmentat

Всего 600

Train par

Размеченн





# Self driving. BDD100K.



Всего 12 классов: car, drivable area, lane, traffic sign, traffic light, person, truck, bus, bike, rider, motor, and train.

Train part: 70,000 , test: 20,000 , val: 10,000 images.

Размеченных объектов: 2,221,128

Для все картинок есть метаданные, такие как "погода", "время суток", и тд.



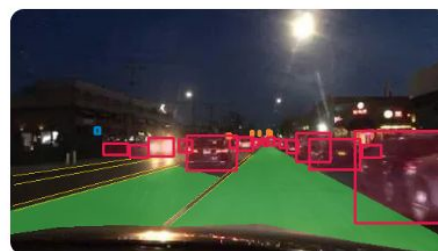
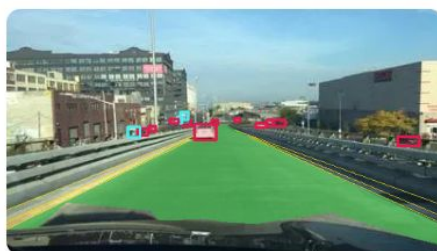
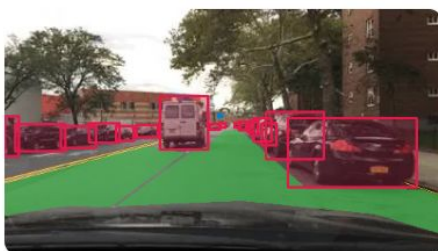
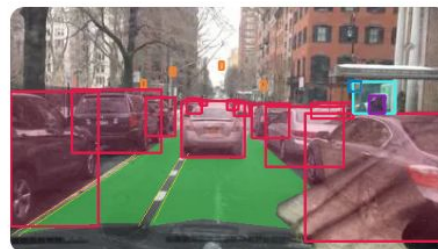
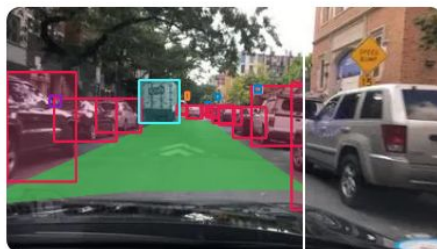
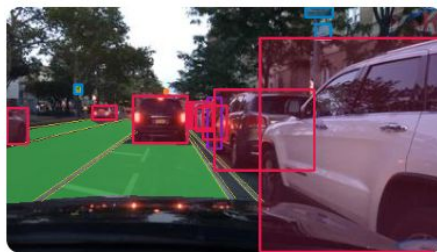
# Self

Всего 12  
traffic  
train.

Train ра

Размечен

Для все  
суток",



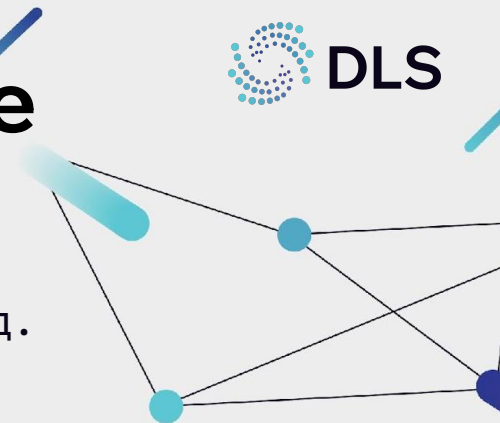
# Распознавание лиц. WiderFace



Датасет содержит фотографии с лицами в 12ти различных ситуациях: в толпе, в гриме, с выражением, в маске и тд.

Train part: 12,881 , test: 16,102 , val: 3,220 images.

Размеченных объектов: 393,703



# Распознавание лиц. WiderFace

Scale



Pose



Occlusion



Expression



Makeup



Illumination



# Снимки со спутников. DOTA v2.0



Всего 18 классов, описывающих машины, здания, дороги, мосты и тд.

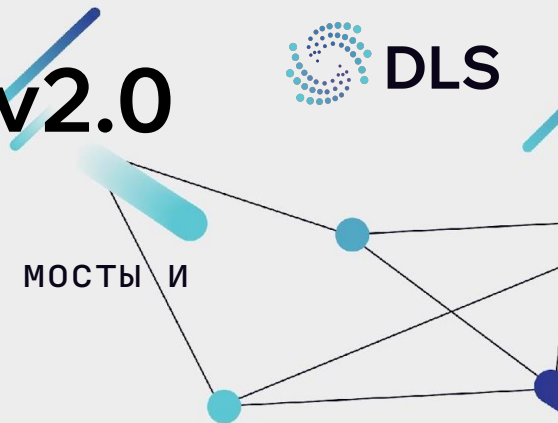
Train part: 1,830 , test: 2792 , val: 593 images.

Также есть часть test-challenge, содержащая 6053 изображений.

Картинки в датасете очень разного размера (от 800 до 20,000 пикселей)

Размеченных объектов: 1,793,658

Присутствует мета информация о времени и дате съемки, расстоянии до земли и тд.





# СНИМ

DOTA · 18 classes

airport



Всего 18  
тд.

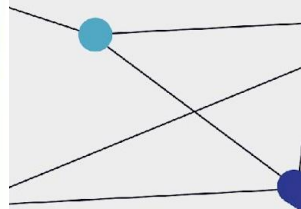
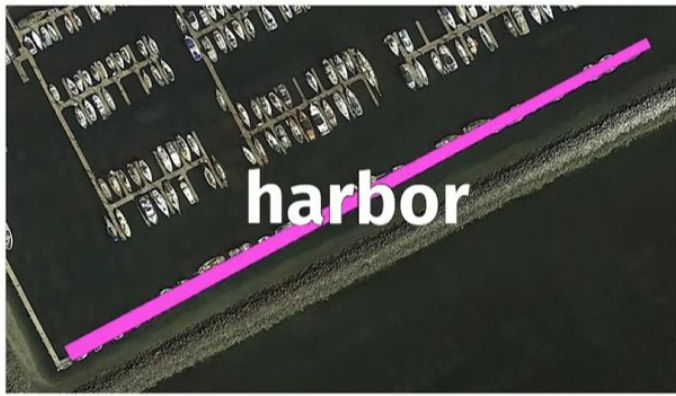
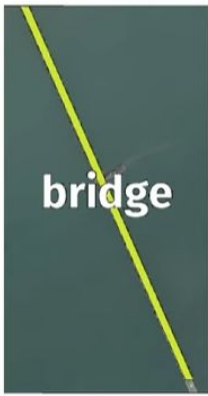
Train par

Также ест

Картинки  
пикселей)

Размеченн

Присутств  
до земли

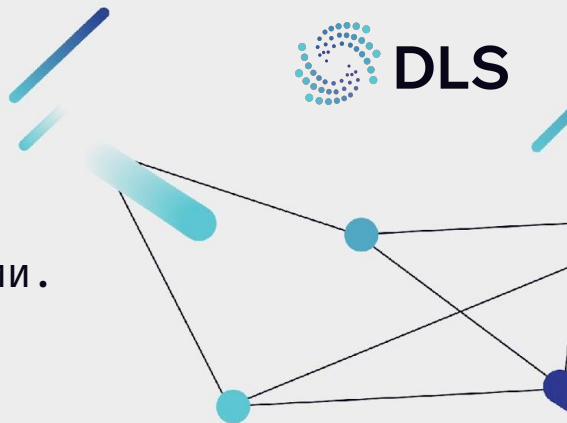


# CrowdHuman.

Большие группы людей в различных ситуациях и освещении.  
Содержит отдельные ббоксы для тела и лица.

Train part: 15,000 , test: 5000 , val: 4,370 images.

Размеченных объектов: 470,000

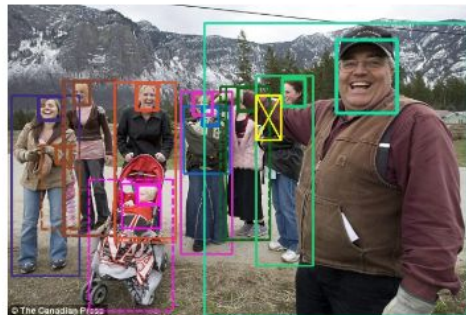


# Cro

Большой  
Содержание

Training

Размер

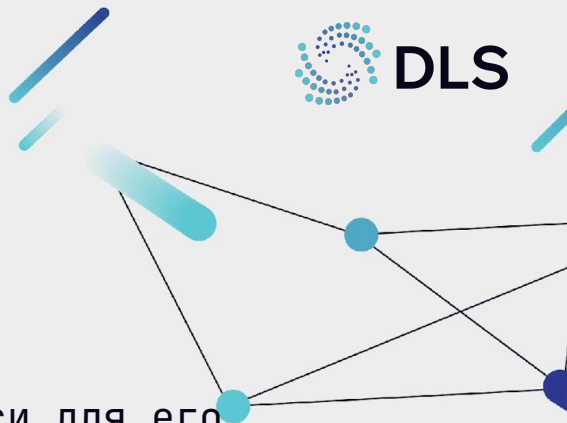


## DLS

# Заключение

Сегодня мы узнали:

1. В чем заключается задача детекции,
2. Где могут быть полезны детекторы,
3. Что такое Bounding box и какие есть форматы записи для его описания,
4. Какие методы применялись до нейросетей,
5. Какие есть детекционные датасеты.



# В следующем видео

1. Познакомимся с общим видом практически всех архитектур детекции,
2. Разберем как детектор обучается и делает предсказания,
3. Посмотрим на метрики и аугментации присущие для задачи.

