

Old School Vision

Жижин Петр

Задача компьютерного зрения

- Понять, что представлено на изображении

Человек видит



Компьютер видит

54	58	255	8	0
45	0	78	51	100
85	47	34	185	207
22	20	148	52	24
52	36	250	74	214
158	0	78	51	247
	72	74	136	251
				74

Зачем это вообще нужно?

- Человек получает от 80% до 90% информации с помощью зрения
- Практическая значимость
 - 1.6 миллиарда картинок загружается в интернет каждый день
- Гипотеза об AI-полноте
- Сложная задача
 - Можно применить много методов

Что можно узнать по картинке?



Классификация

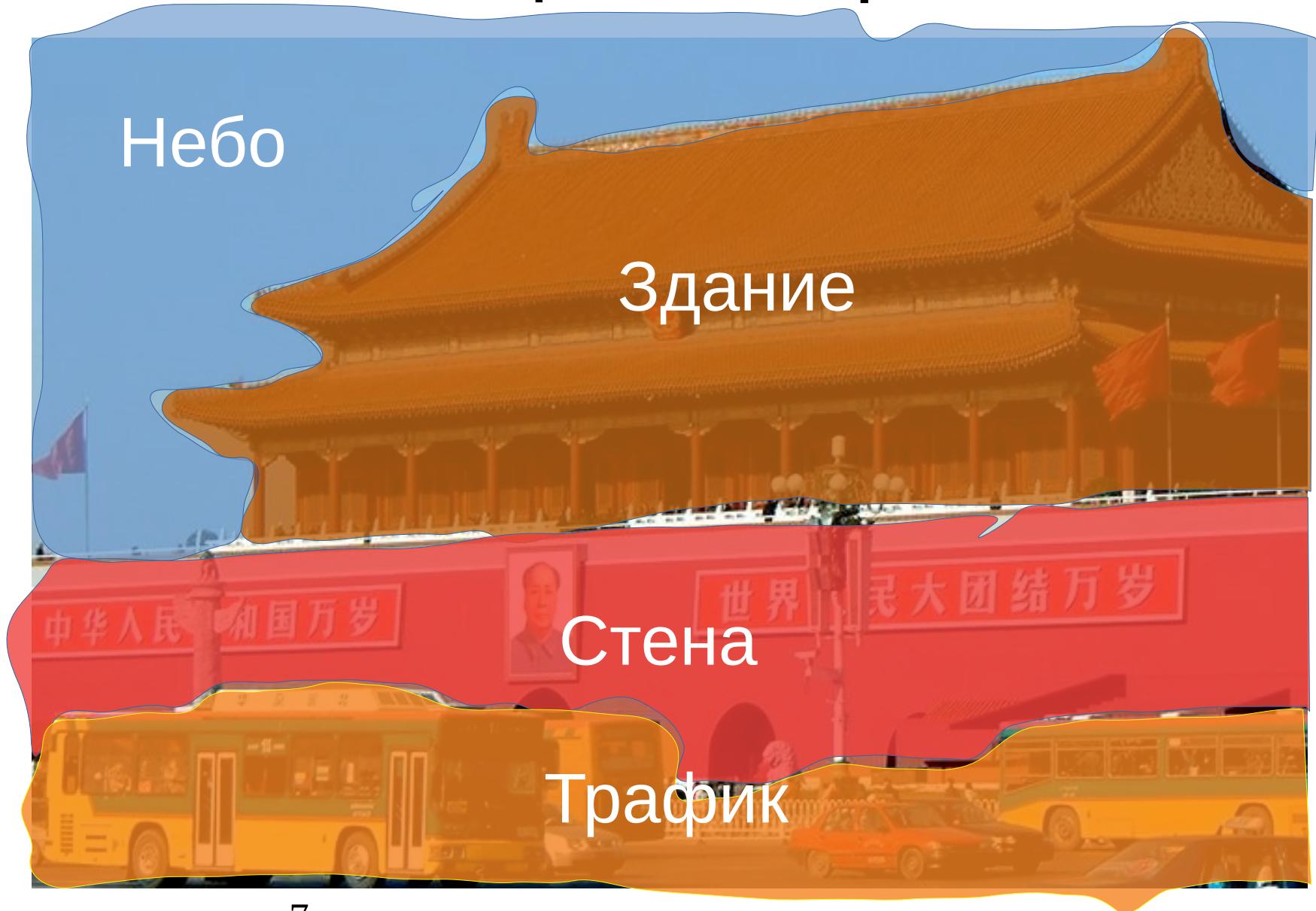
Пекин
Город
Улица
Пл. Танъаньмэнь



Обнаружение объектов



Сегментация изображения



Поиск похожих изображений



Контрольный вопрос:
Чем отличается задача поиска объекта от
задачи сегментации?

Почему же так сложно работать с изображениями?

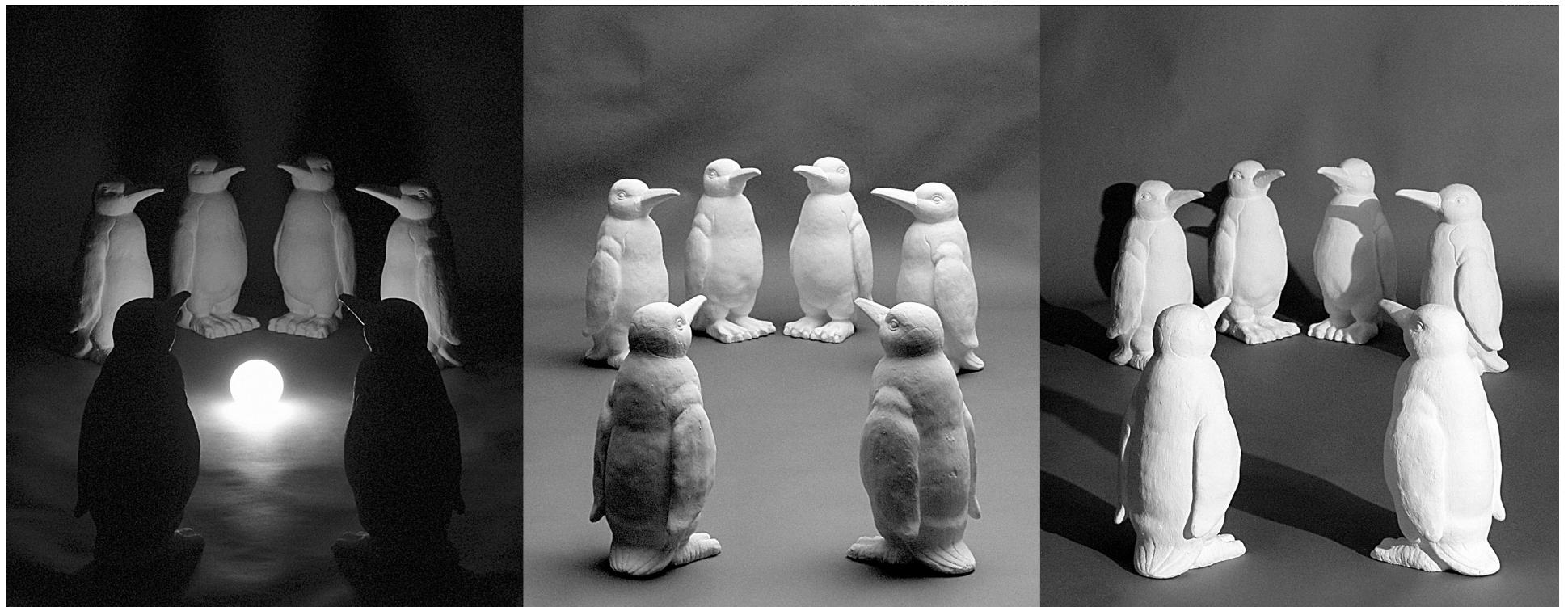
Сложности

- Различная точка съемки



Сложности

- Разница в освещении



Сложности

- Проблемы при съемке кадров



Сложности

- Разница в масштабе



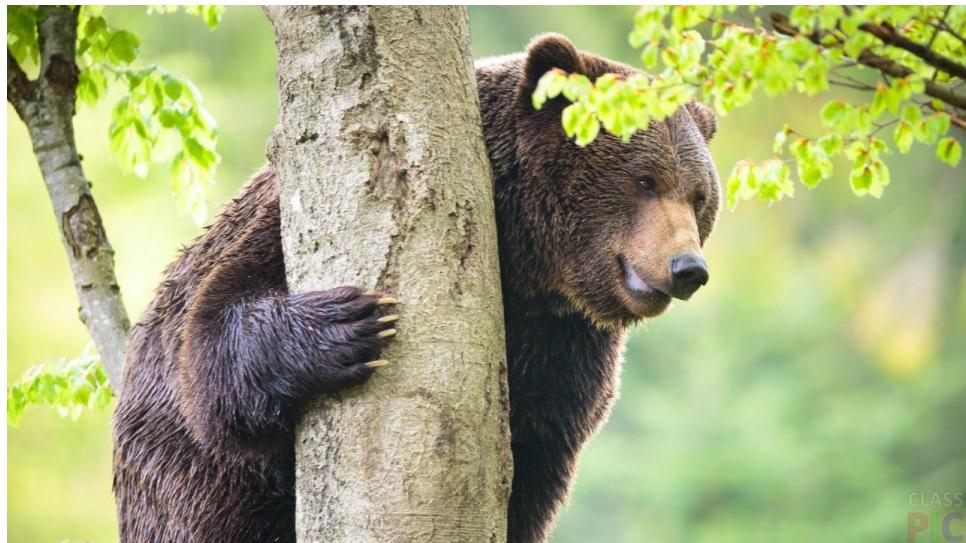
Сложности

- Деформации



Сложности

- Взаимные перекрытия



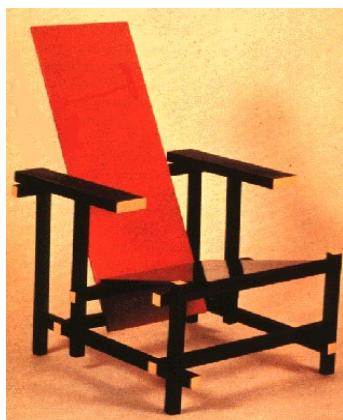
Сложности

- Маскировка

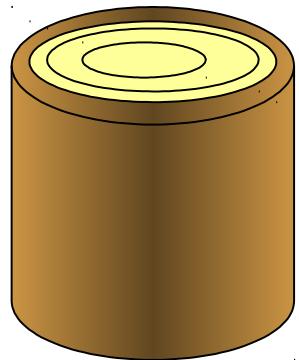


Сложности

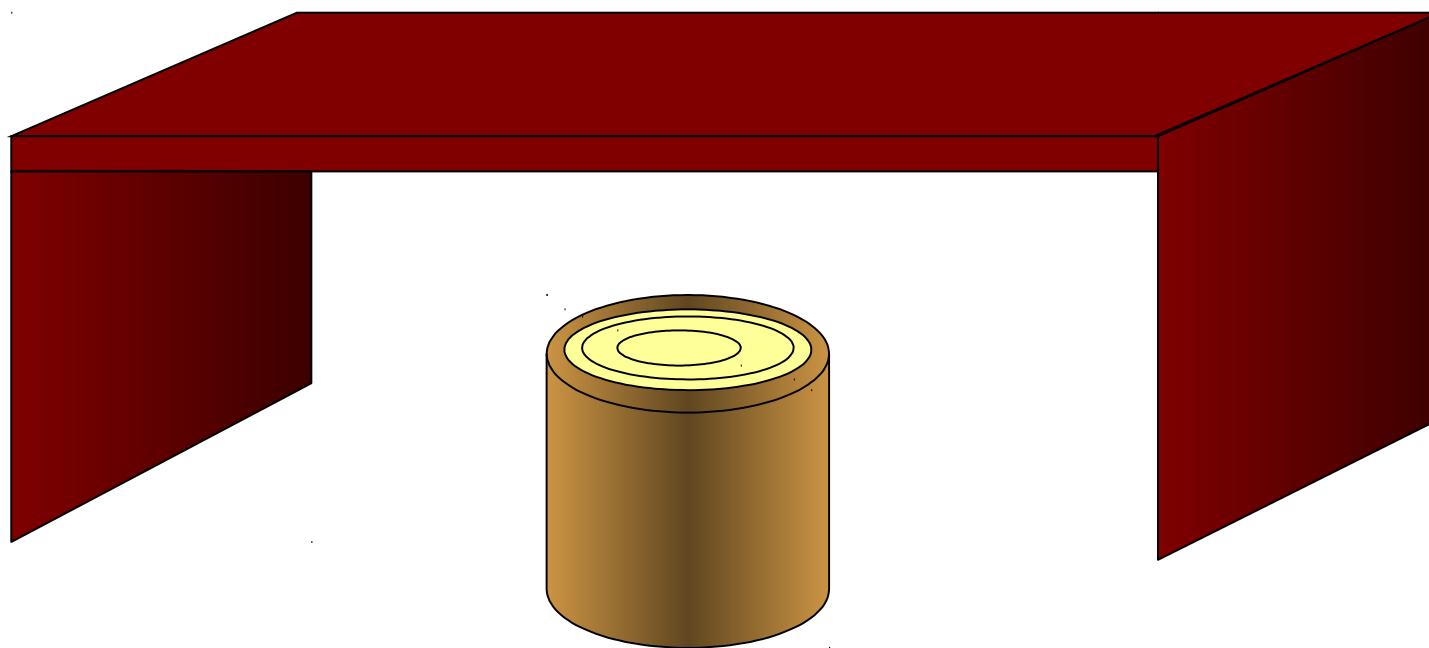
- Внутриклассовая изменчивость



Что изображено на картинке?



А так?



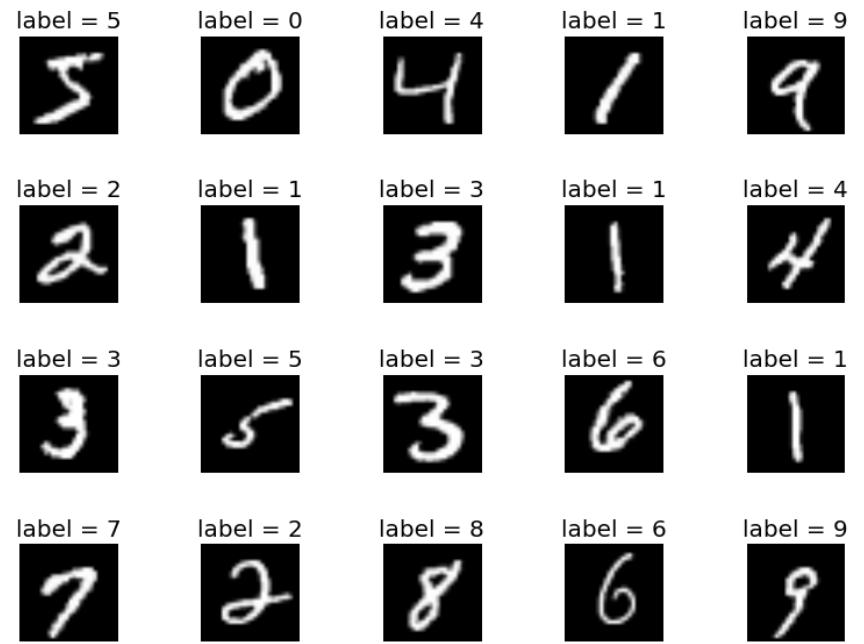
Датасеты для задач компьютерного зрения

- Классификация
 - Выход – один из классов картинки (какой класс изображен)
- Обнаружение объектов
 - Выход – набор прямоугольников с классами (что изображено в каждом прямоугольнике)
- Сегментация изображений
 - Выход – изображение, где каждый пиксель имеет класс
- Поиск похожих
 - Выход – похожие изображения
- Другие

Зачем вообще нужны датасеты?

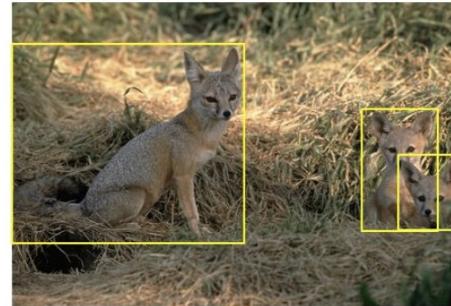
MNIST

- Ч/б изображения 28x28 с цифрами
- Предсказывается цифра на картинке
- 60 000 обучающая выборка
- 10 000 тестовая выборка

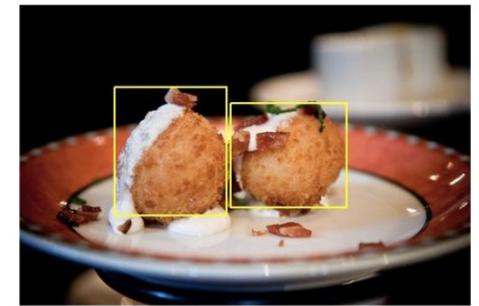


ImageNet

- 14 197 122 изображений
- 21 841 класс для классификации
- 1 034 908 изображений для обнаружения объектов
 - 3 000 классов



kit fox



croquette



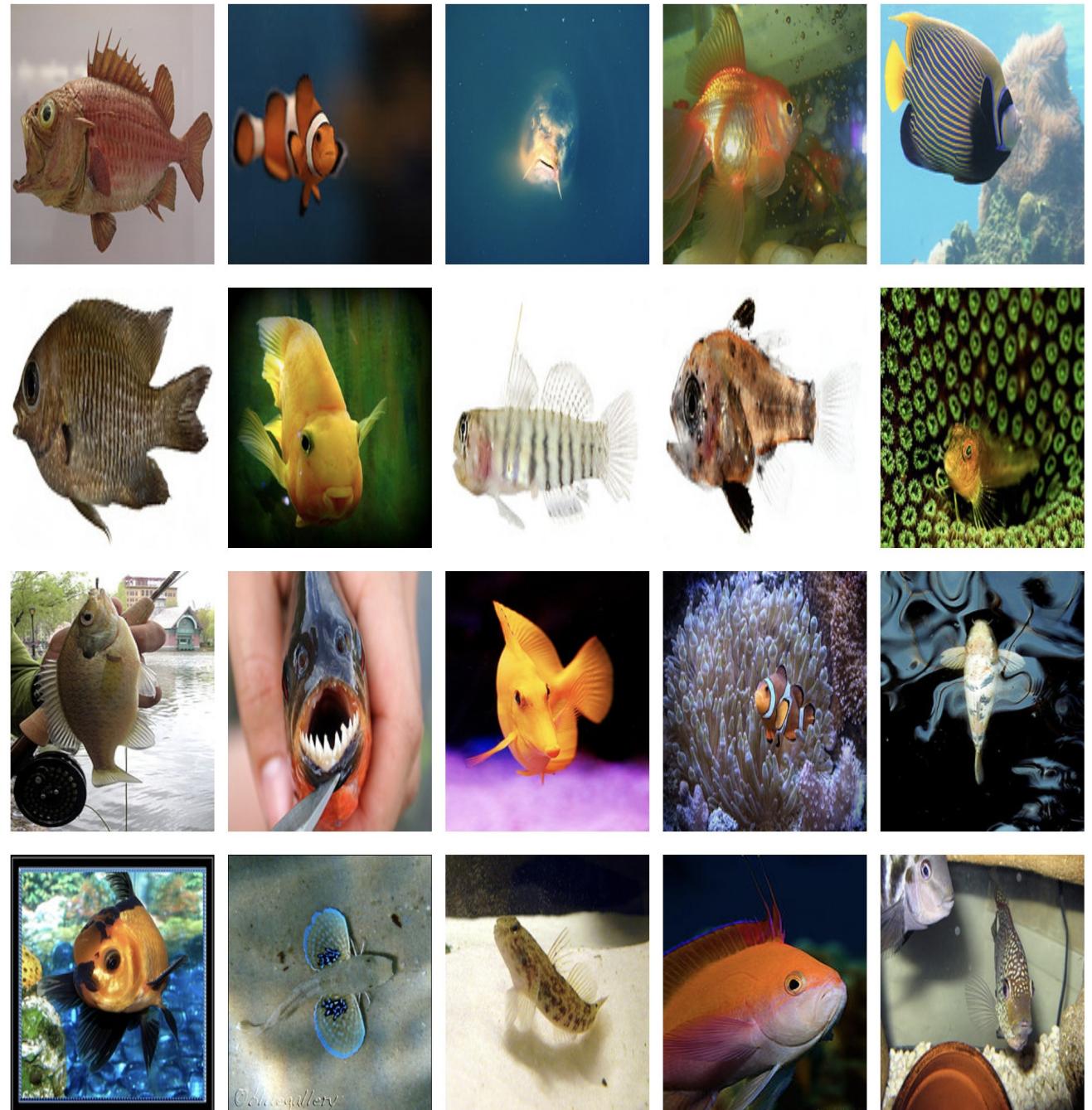
airplane



frog

ImageNet 2011 Fall Release (32326)

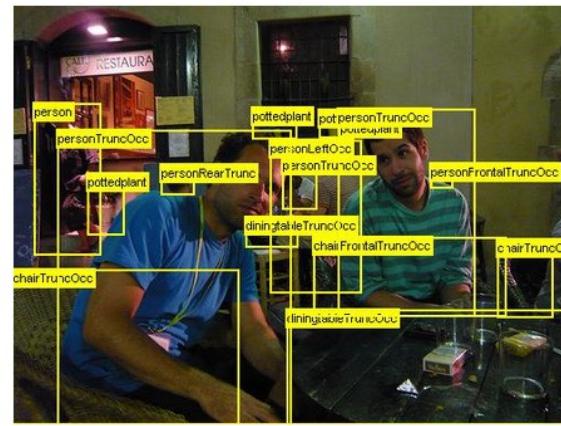
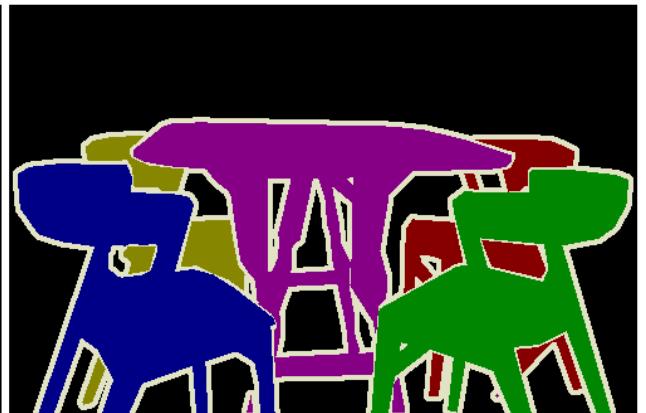
- ↳ plant, flora, plant life (4486)
- ↳ geological formation, formation (17)
- ↳ natural object (1112)
- ↳ sport, athletics (176)
- ↳ artifact, artefact (10504)
- ↳ fungus (308)
- ↳ person, individual, someone, someone (1)
- ↳ animal, animate being, beast, brute (1)
- ↳ invertebrate (766)
- ↳ homeotherm, homiotherm, homoiotherm (1)
- ↳ work animal (4)
- ↳ darter (0)
- ↳ survivor (0)
- ↳ range animal (0)
- ↳ creepy-crawly (0)
- ↳ domestic animal, domesticated (1)
- ↳ molter, moulter (0)
- ↳ varmint, varment (0)
- ↳ mutant (0)
- ↳ critter (0)



Pascal VOC

- 20 классов для классификации
- 11 530 изображений
- 27 540 прямоугольников
- 6 929 сегментаций

Pascal VOC



Microsoft Common Objects in Context (COCO)

- Сегментация
- 330 000 изображений
- 1 500 000 объектов
- 80 категорий
- 91 категорий объектов “stuff”



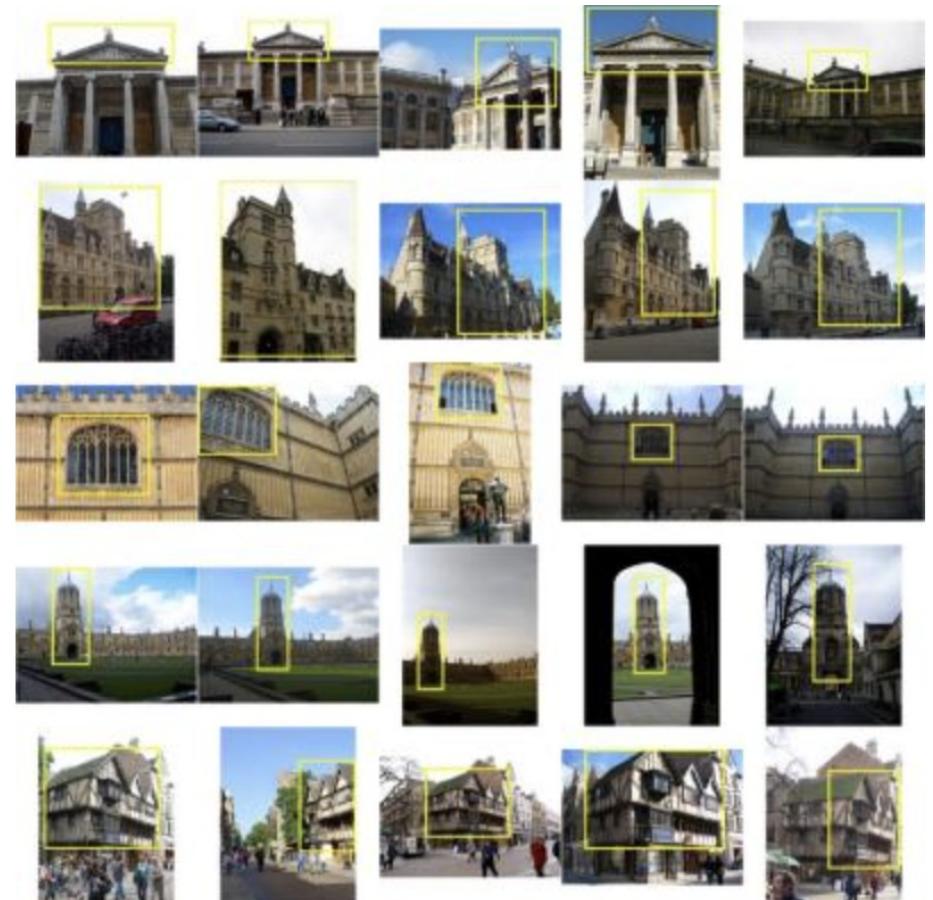
Cityscapes dataset

- Дорожные ситуации
 - 50 городов
 - Разное время дня и ночи
 - Разные месяца съемки
- Сегментация на 30 классов
 - Дороги, автомобили, пешеходы, знаки, дома и тд
- 25 000 изображений
 - 5 000 точно сегментированных
 - 20 000 ориентировочно



Oxford Buildings

- Здания Оксфорда с разных ракурсов
- 5062 изображений
 - Доступны 100 тысяч “отвлекающих” картинок
- 11 зданий
 - Разные ракурсы
- 55 поисковых запросов
 - Поиск похожих



INRIA Holidays dataset

- Изображения достопримечательно стей с отпусков разработчиков
- 1491 изображений
- 500 поисковых запросов

