# Object Recognition

in images

Image: two-dimensional matrix of pixels

Contains information, objects represented in it

Image: two-dimensional matrix of pixels

Contains information, objects represented in it

Extract this information?

Automate the process?

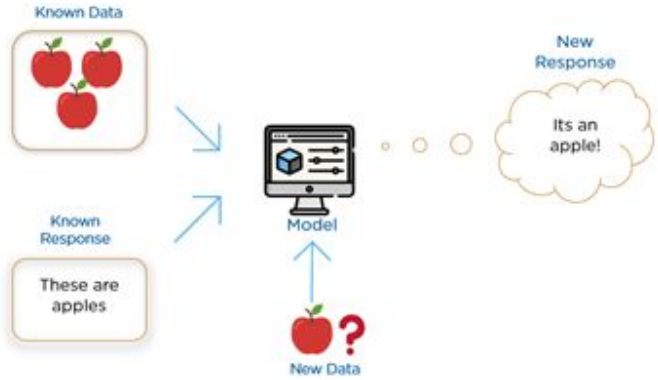
# Machine Learning

# Machine Learning

In particular, Neural Networks

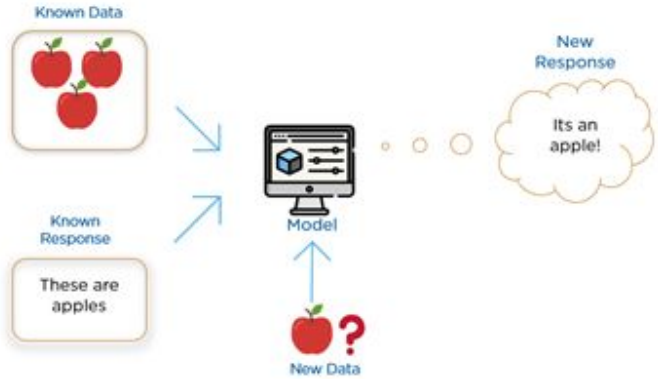
Neural network?

# Neural network



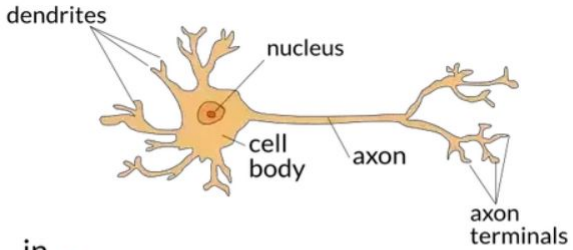

## Supervised learning

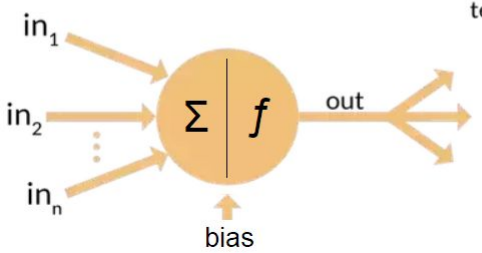
# Neural network

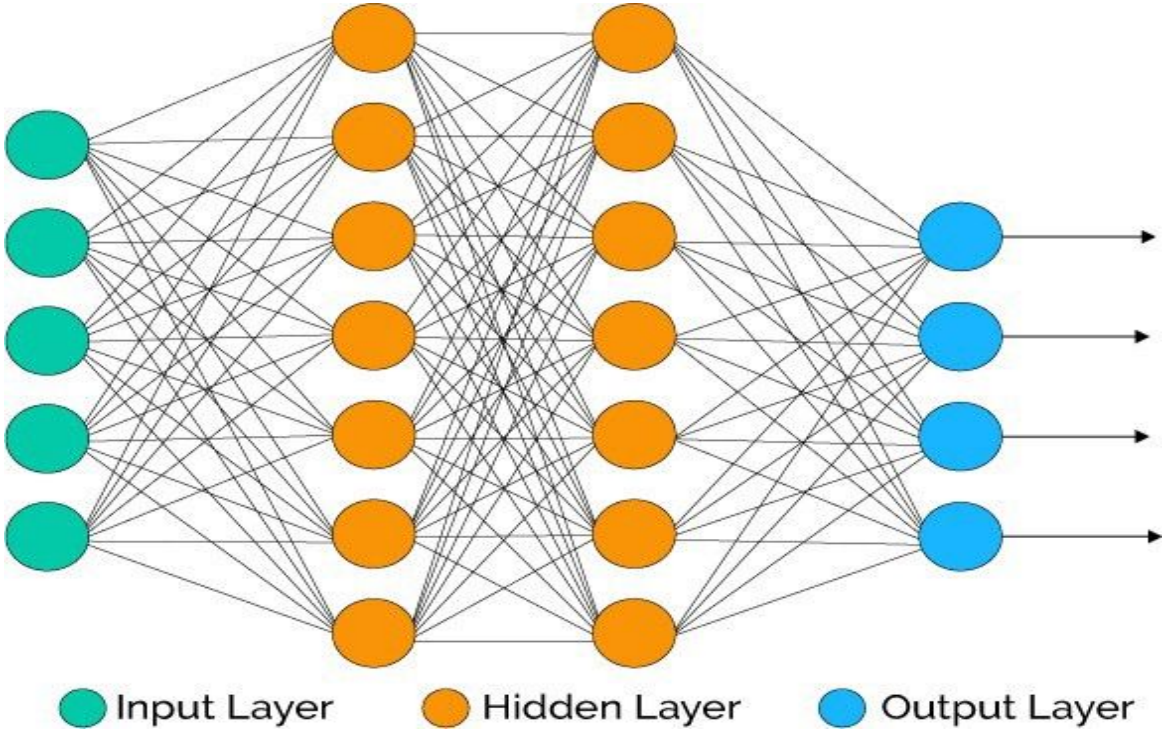



Mimics the brain

# Supervised learning






⇒ Interconnected layers of neurons

# To learn more:

- https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#3c290e7a1245
- https://www.techradar.com/news/what-is-a-neural-network
- https://www.datascience.com/blog/convolutional-neural-network

# Recipe for a good model

# Recipe for a good model

- A lot of training material (>1000 per class)

# Recipe for a good model

- A lot of training material (>1000 per class)

- A good neural network

# Recipe for a good model

- A lot of training material (>1000 per class)

- A good neural network
  - Many parameters

# Recipe for a good model

- A lot of training material (>1000 per class)

- A good neural network
  - Many parameters
  - lots of tweaking

# Recipe for a good model

- A lot of training material (>1000 per class)
- A good neural network
  - Many parameters
  - lots of tweaking
- Computational power

# Recipe for a good model

- A lot of training material (>1000 per class)
- A good neural network
  - Many parameters
  - lots of tweaking
- Computational power
  - GPUs

# Recipe for a good model

- A lot of training material (>1000 per class)
- A good neural network
  - Many parameters
  - lots of tweaking
- Computational power
  - GPUs
- A lot of time

# Recipe for a good model

- A lot of training material (>1000 per class)
- A good neural network
  - Many parameters
  - lots of tweaking
- Computational power
  - GPUs
- A lot of time
  - Retraining if the model isn't optimal

# How do we do that?

How do we do that?

# We don't

# Lots of training datasets online

Lots of training datasets online

Pretrained models available for them

# Image classification

# Image classification

persian_cat

bat

# Image classification

persian_cat

bat

# Image classification

persian_cat

toothbrush


IMAGENET


14m+ labeled images


14m+ labeled images

20k+ classes

# ILSVRC

# ILSVRC

- ImageNet Large Scale Visual Recognition Competition

# ILSVRC

- ImageNet Large Scale Visual Recognition Competition
- Yearly competition

# ILSVRC

- ImageNet Large Scale Visual Recognition Competition
- Yearly competition
- 2015 winner: ResNet50 -- available online

| Label: | Score: |
| --- | --- |
| platypus | 0.31578568 |
| puffer | 0.25567934 |
| scuba_diver | 0.1726397 |
| tree_frog | 0.049659904 |
| electric_ray | 0.036546305 |

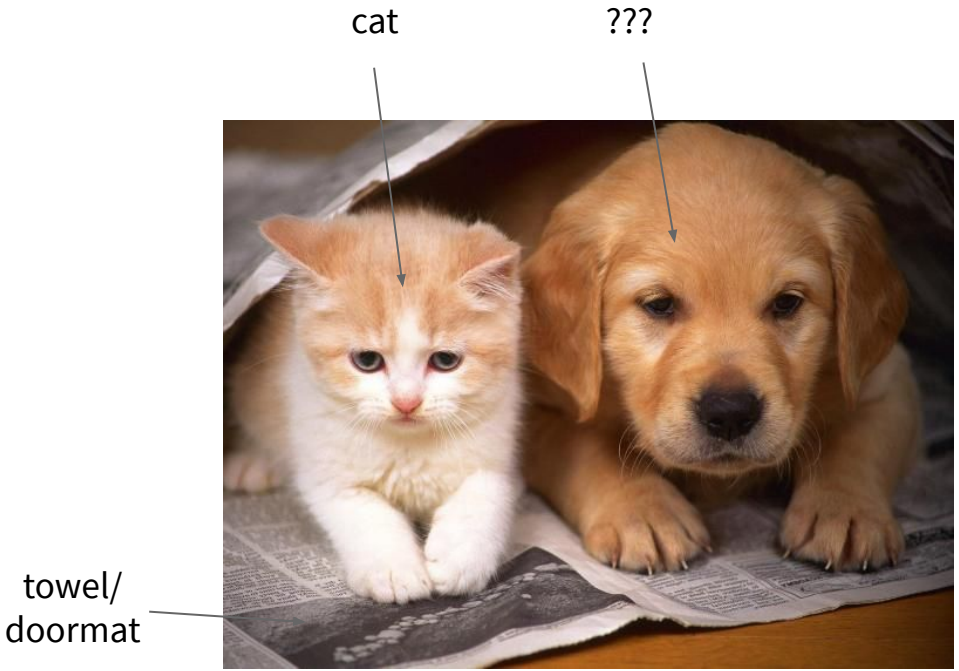
| Label: | Score: |
|---|---|
| platypus | 0.31578568 |
| puffer | 0.25567934 |
| scuba_diver | 0.1726397 |
| tree_frog | 0.049659904 |
| electric_ray | 0.036546305 |

| Label: | Score: |
|---|---|
| swimming_trunks | 0.855407178401947 |
| maillot | 0.03095098026096821 |
| miniskirt | 0.008445939049124718 |
| bathing_cap | 0.0081565510481596 |
| maillot (again) | 0.007052644621580839 |

# But...

# But...

# But...

# Huge dataset, many classes

# Huge dataset, many classes

BUT

Huge dataset, many classes

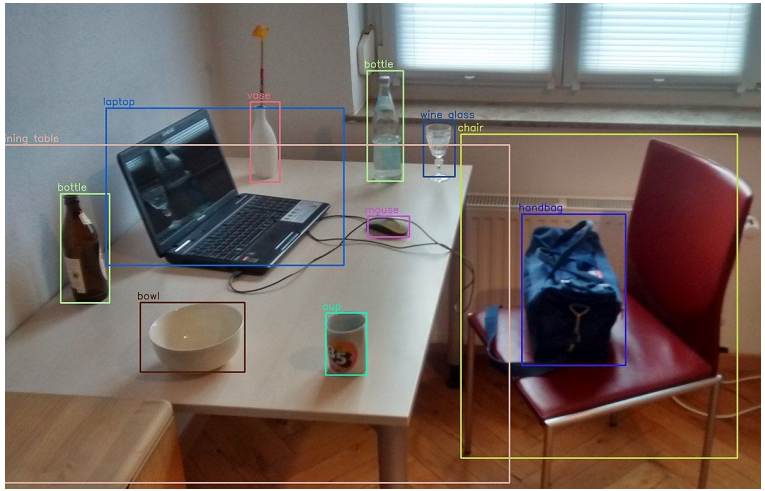
BUT

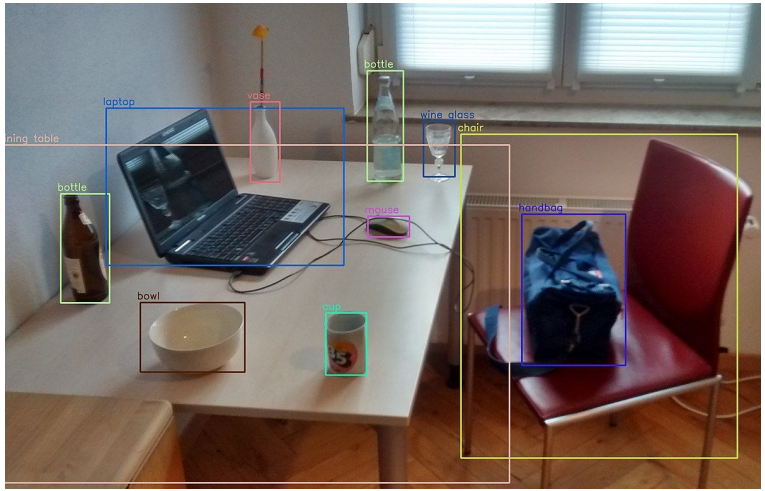
Made for single-label classification

# Object detection

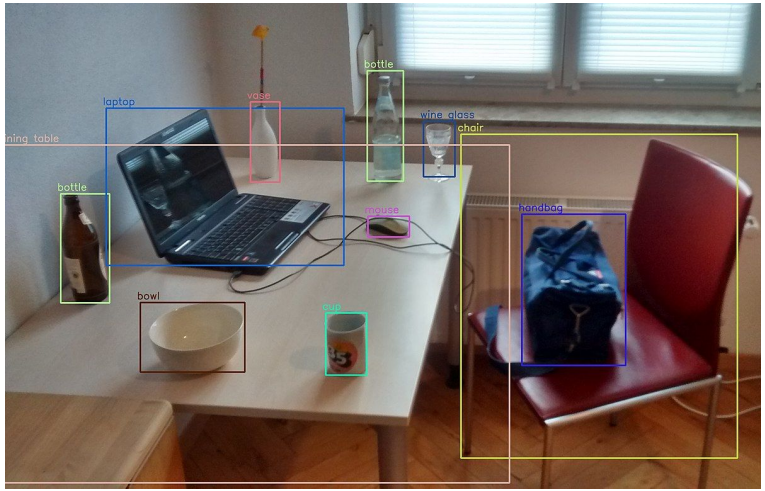
# Object detection

# Image segmentation


bottle
laptop
vase
wine glass
chair
ining table
bottle
mouse
handbag
bowl
cup

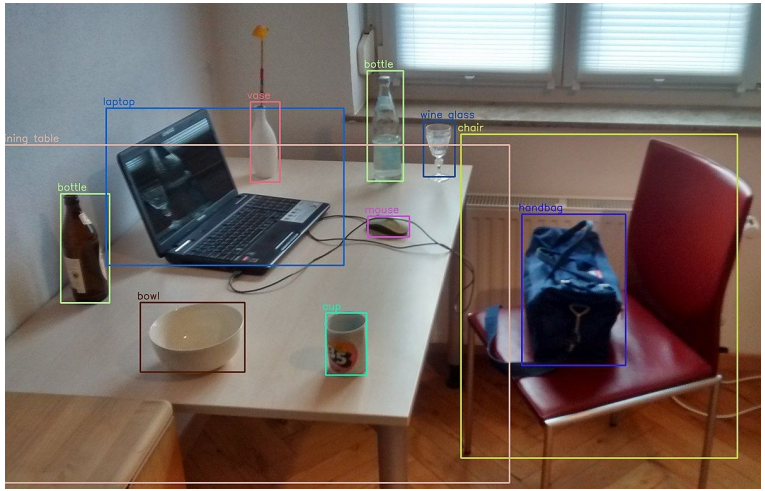

bottle
laptop
vase
wine glass
chair
bottle
mouse
handbag
bowl
cup

VS


bottle
laptop
vase
wine glass
chair
bottle
mouse
handbag
bowl
cup

VS


bus 0.853
truck 0.718
truck 0.726
person 0.967
car 0.777
person 0.969
person 0.857
person 0.940
cow 0.990
cow 0.788
cow 0.988
person 0.962
person 0.969
car 0.932
cow 0.742
car 0.941
person 0.918
car 0.756
cow 0.793




vs


# ⇒ multi-label classification


coco
Common Objects in Context


200k+ labeled images


200k+ labeled images

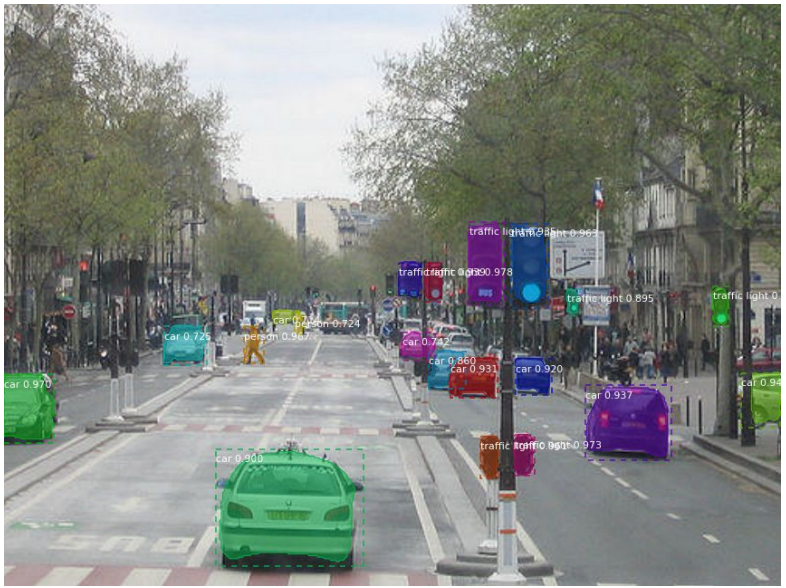
~80 classes

# Mask-RCNN pretrained model

https://github.com/matterport/Mask_RCNN

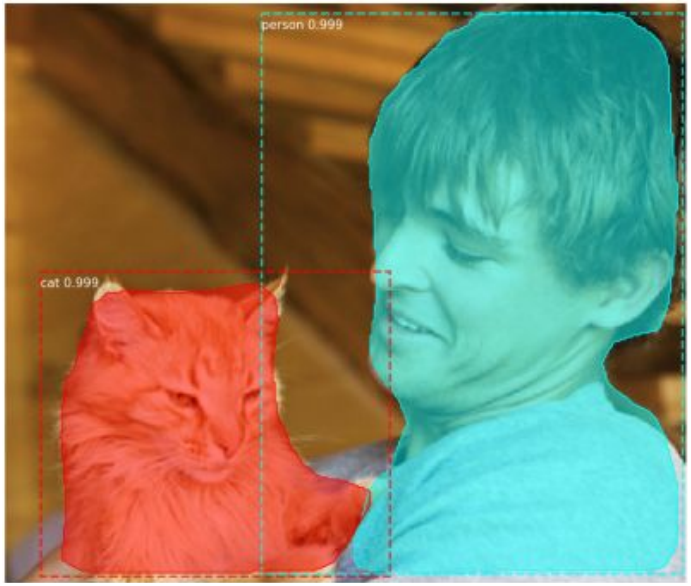
# Mask-RCNN pretrained model

https://github.com/matterport/Mask_RCNN

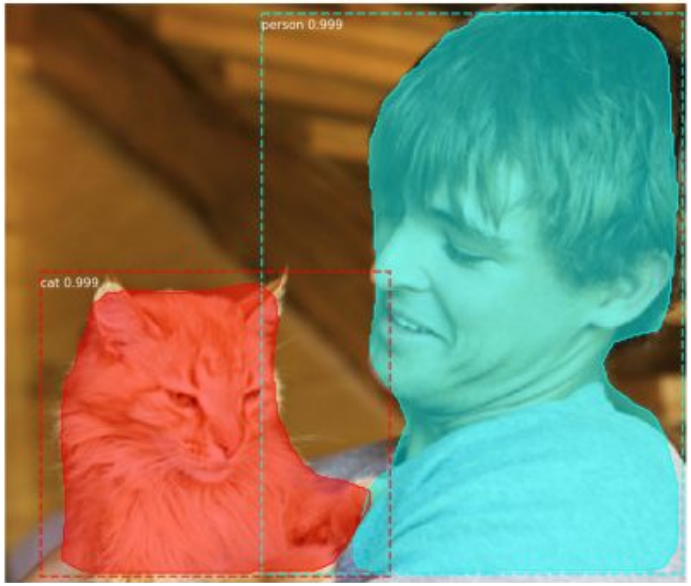


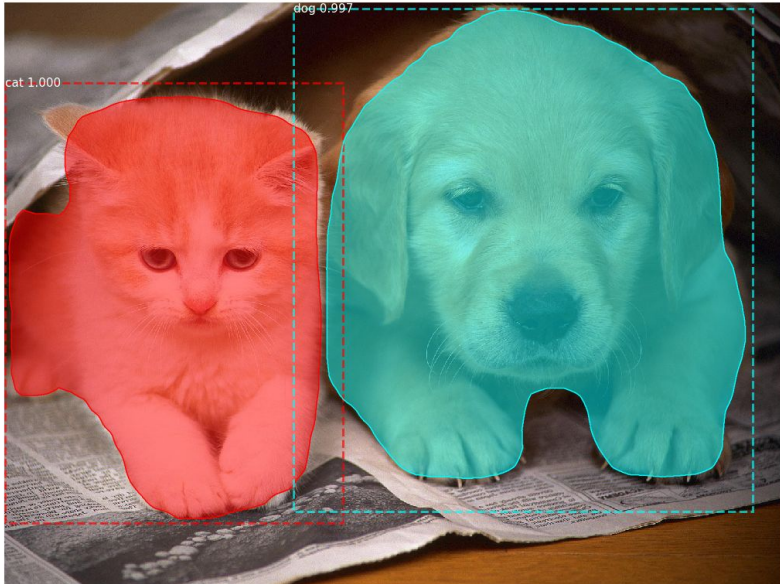
# Good results…

# Good results...



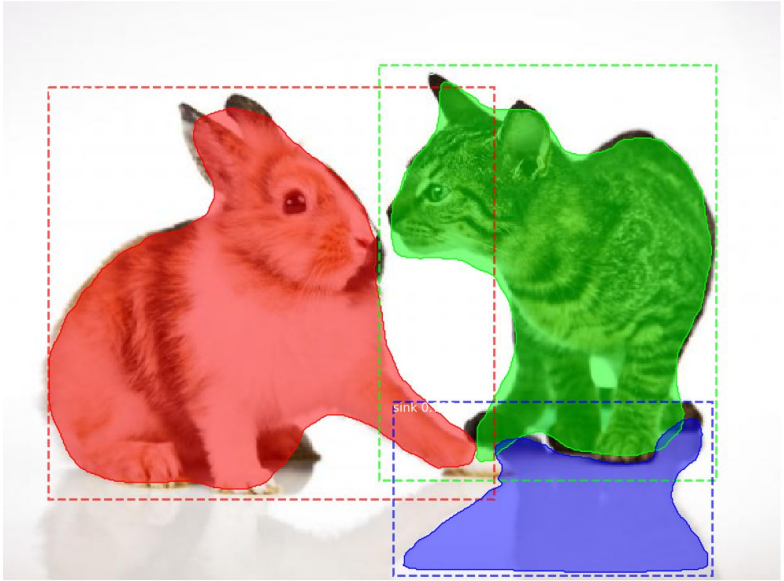
# Good results...



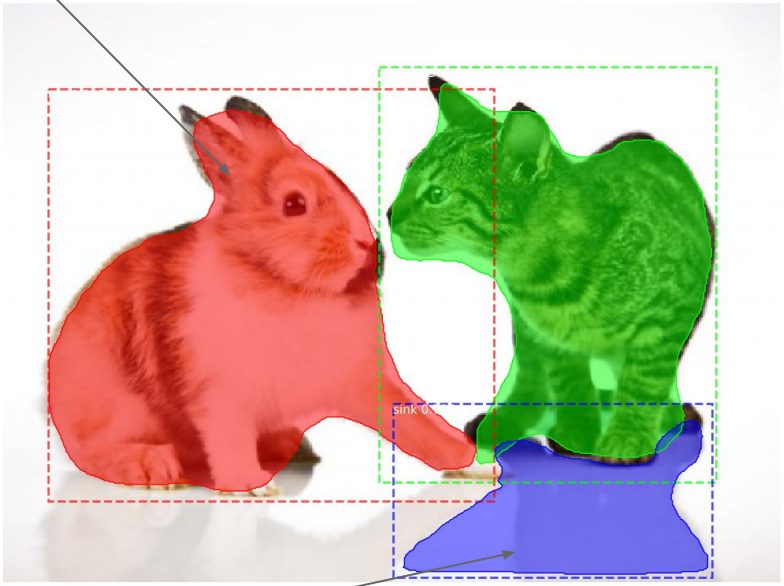



# … mostly

# … mostly



# … mostly



# Multi-label classification

# Multi-label classification

BUT

# Multi-label classification

BUT

# Harder to create a good dataset

# What's left to do

# What's left to do

- Look into other datasets

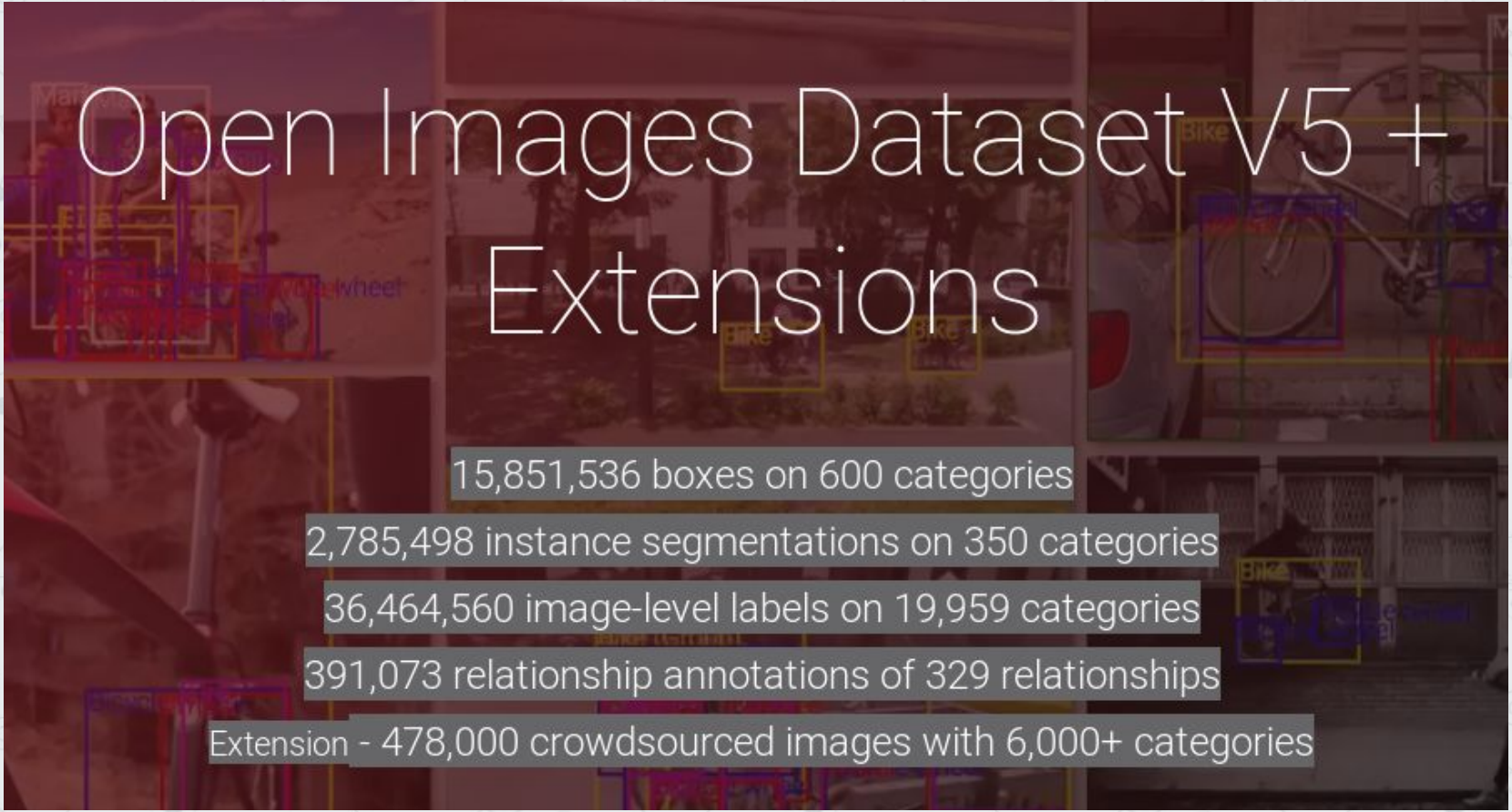
- Look into other datasets
- Find additional models

- Look into other datasets
- Find additional models
- Investigate model retraining

# A promising axis of research

A promising axis of research

# Google's Open Images


Open Images Dataset V5 +
Extensions
15,851,536 boxes on 600 categories
2,785,498 instance segmentations on 350 categories
36,464,560 image-level labels on 19,959 categories
391,073 relationship annotations of 329 relationships
Extension - 478,000 crowdsourced images with 6,000+ categories

# Open Images

# Open Images

- A recent dataset (V1 in 2016!)

# Open Images

- A recent dataset (V1 in 2016!)
- Models harder to find

# Open Images

- A recent dataset (V1 in 2016!)
- Models harder to find
- Looks like our best option so far

*Thank you for your attention.*

*Tho tion.*