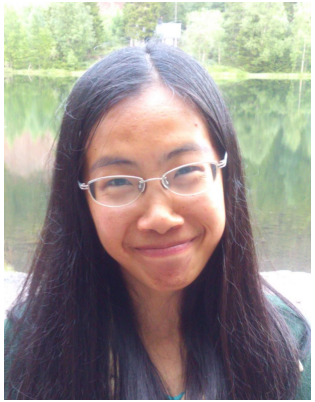
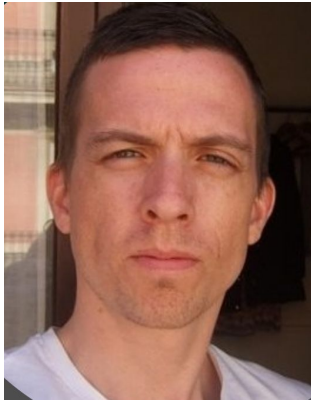
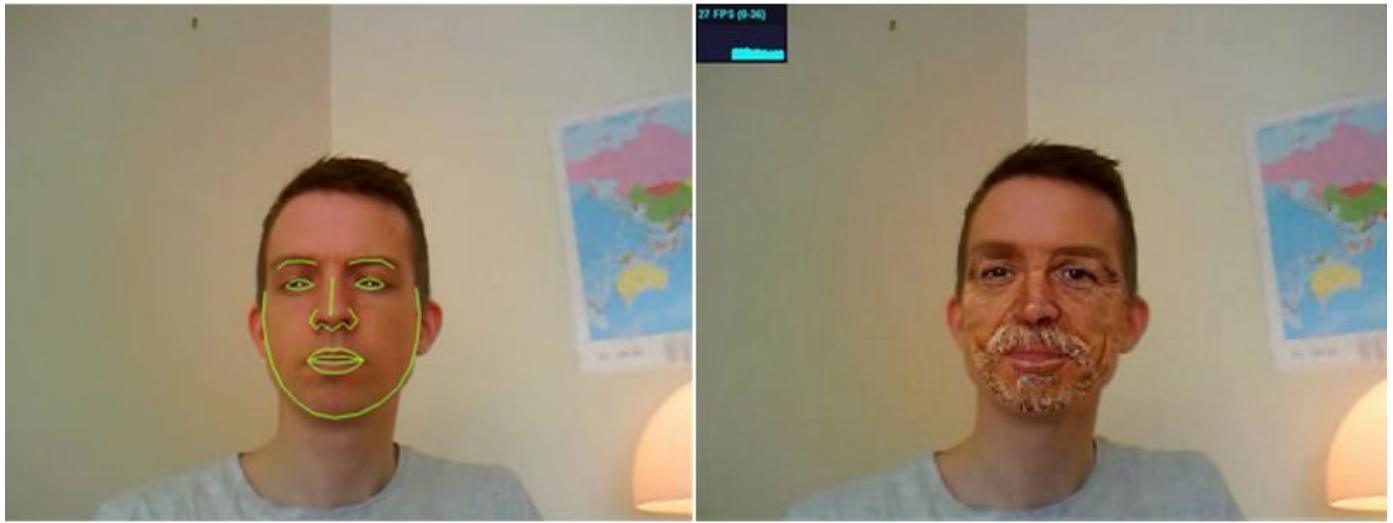
ML Workshop Day 1



Ning Zhou



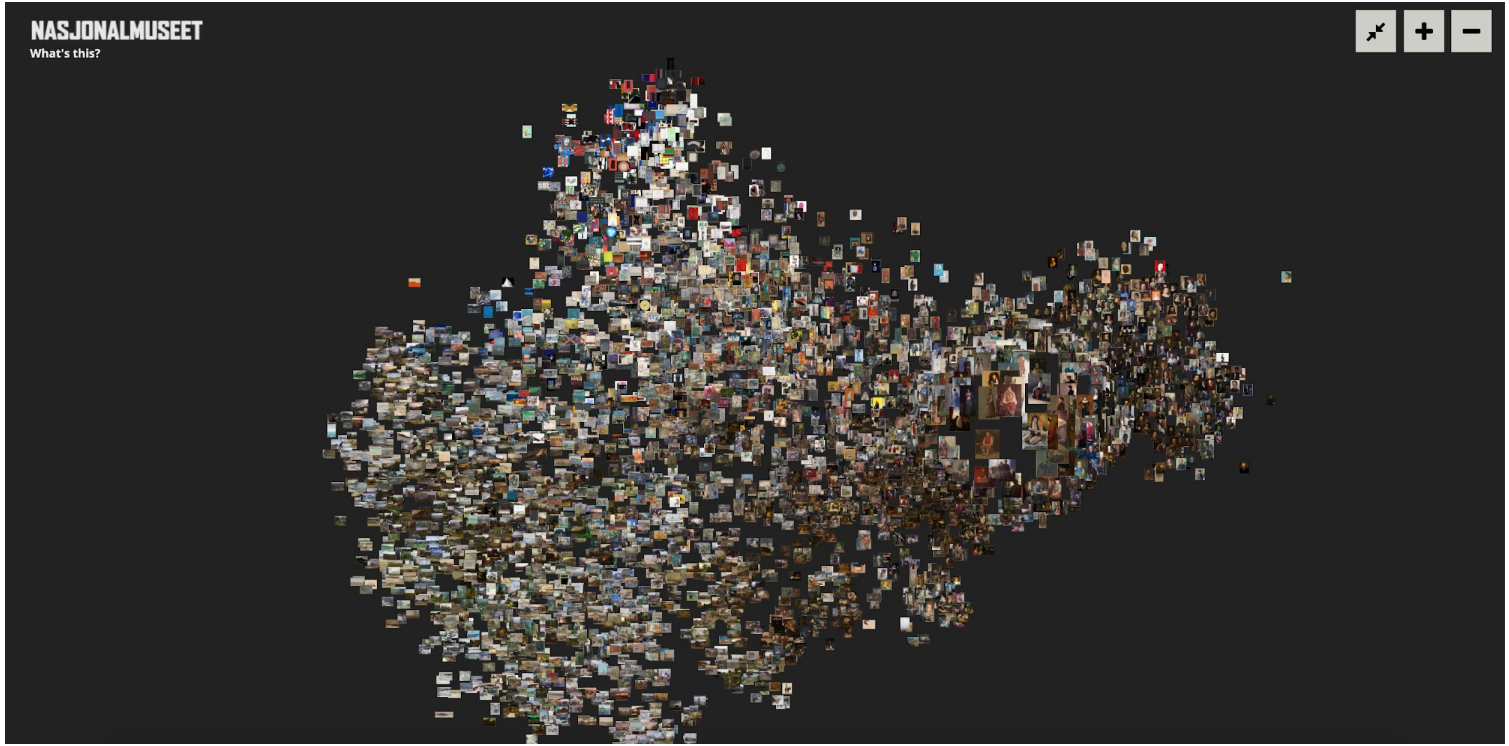
Audun Mathias Øygard



“clmtrackr”

Realtime face tracking library in javascript

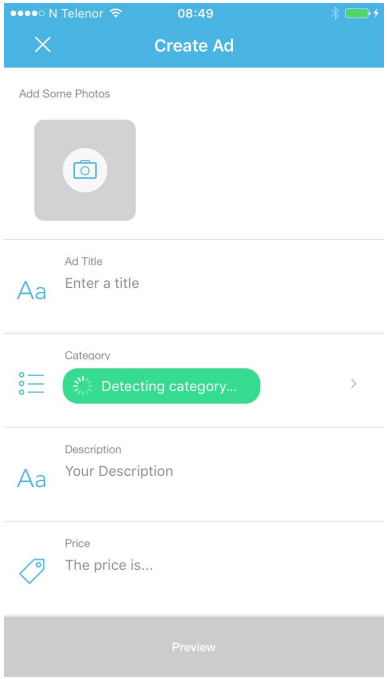
https://github.com/auduno/clmtrackr

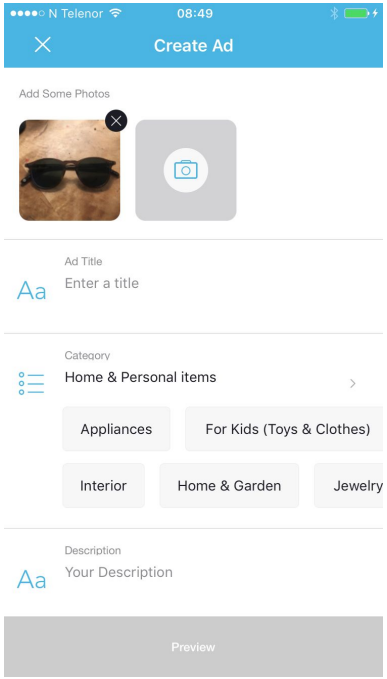


Visualizing the collection of Nasjonalmuseet

(Collaboration with Bengler)

http://vy.nasjonalmuseet.no





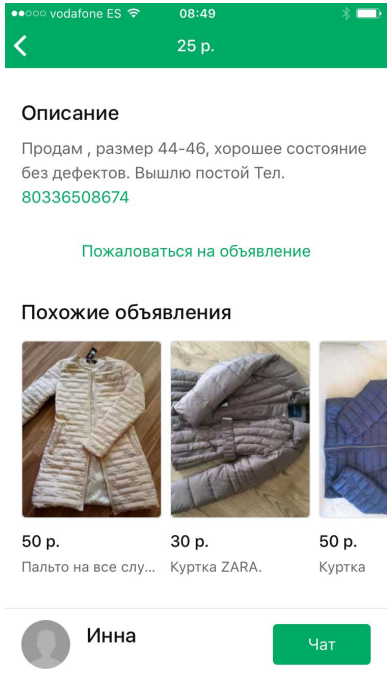


Image recognition systems
Schibsted marketplaces

Workshop outline

Day 1:

- Morning
  - Intro to ML/AI
  - Intro to p5.js
  - ML Basics
- Lunch
- Afternoon
  - Intro to ML5.js
  - Hands-on with pretrained models
  - Training models with Teachable Machine
  - Hands-on with trained models

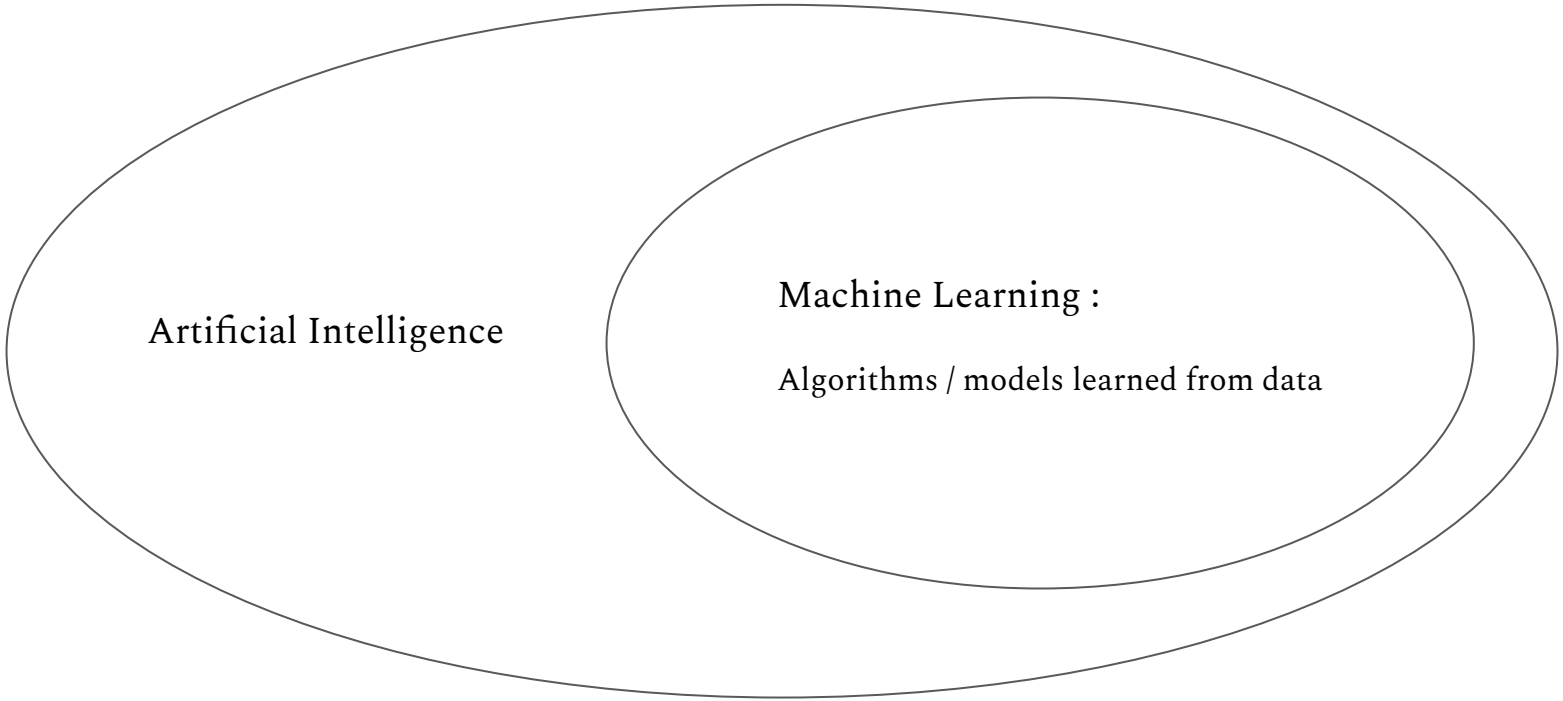
Day 2:

- Morning
  - Recap of Day 1
  - Hands-on continued?
  - Search and recommenders
- Lunch
- Afternoon
  - Generative ML: basics
  - Hands-on with generative models

Day 3:

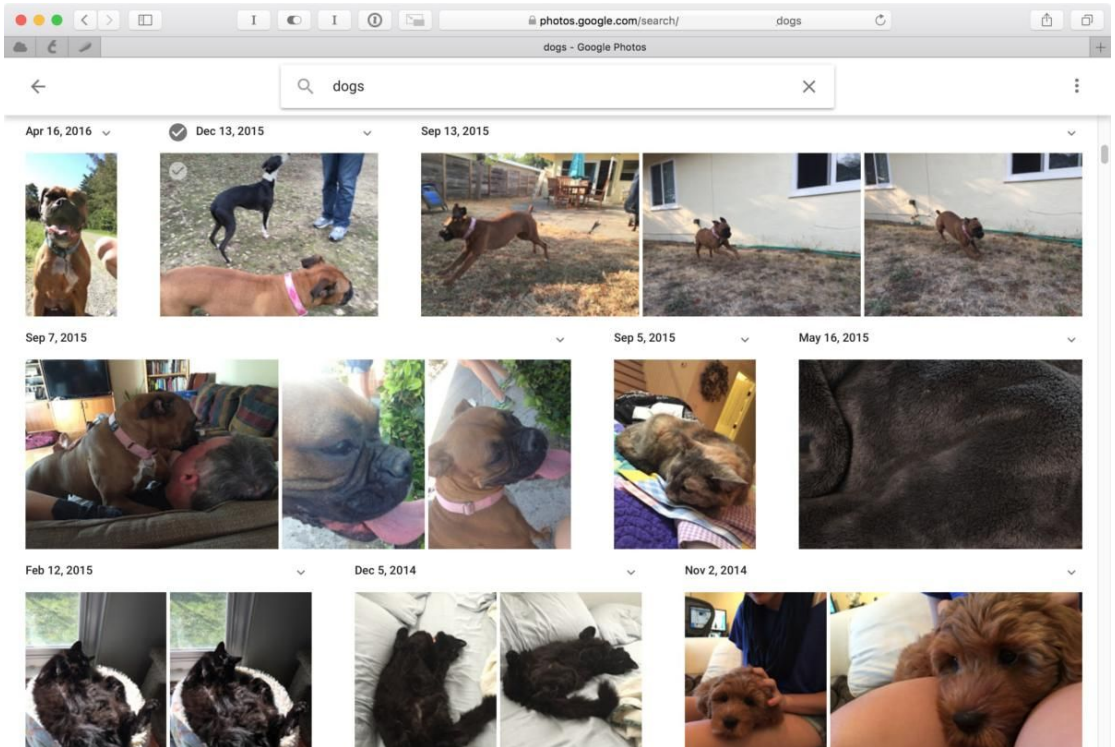
- Morning
  - Recap of day 2
  - Generative ML : GANs
  - Hands-on session
  - Generative ML for design : issues
- Lunch
- Afternoon
  - Things to consider when using ML in reality
  - Discussion and further work
  - Recap of the workshop

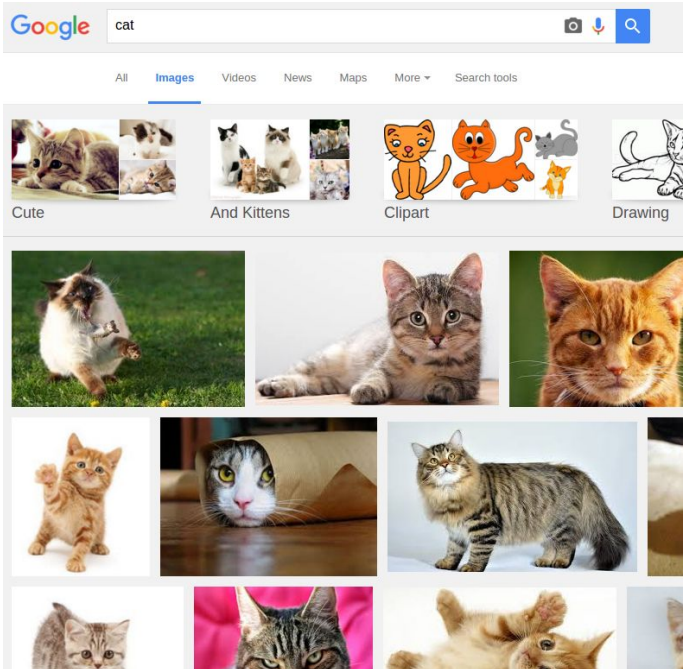
What is Machine Learning?


Artificial Intelligence
Machine Learning :
Algorithms / models learned from data

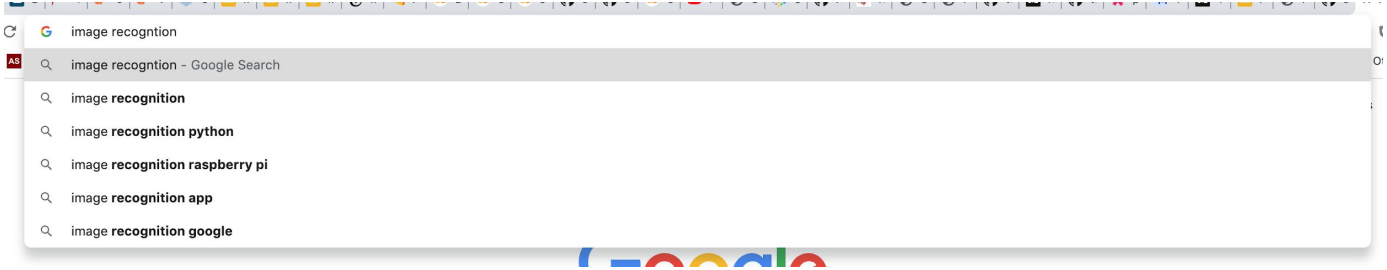
Voice recognition

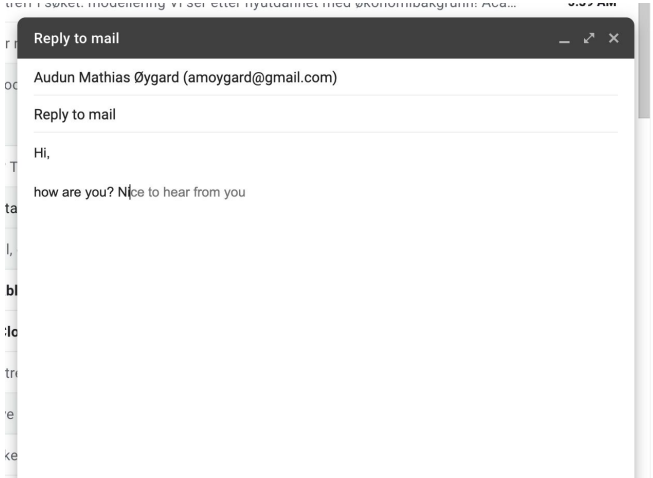
Image search

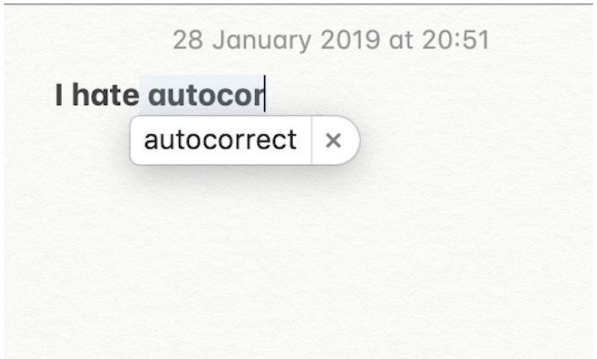




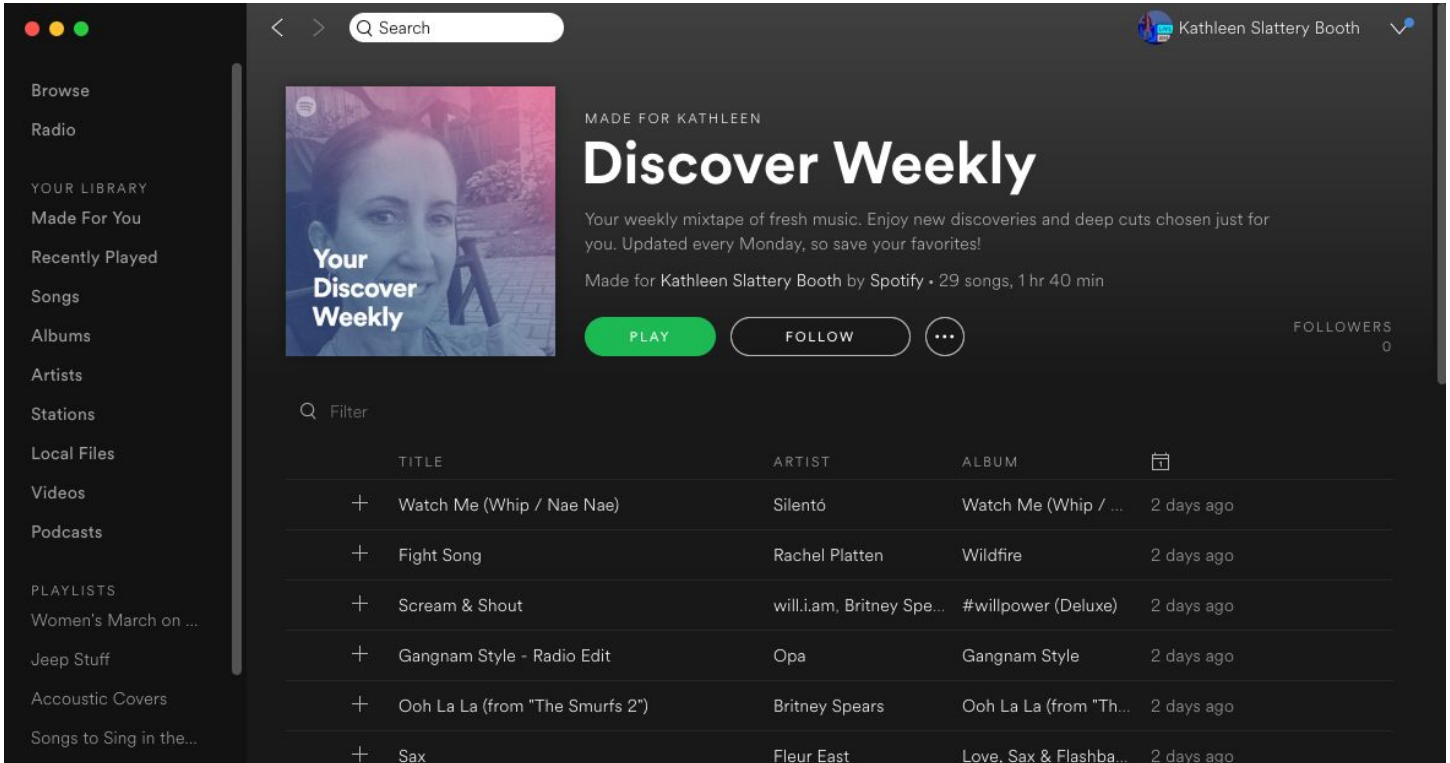
Text suggestion



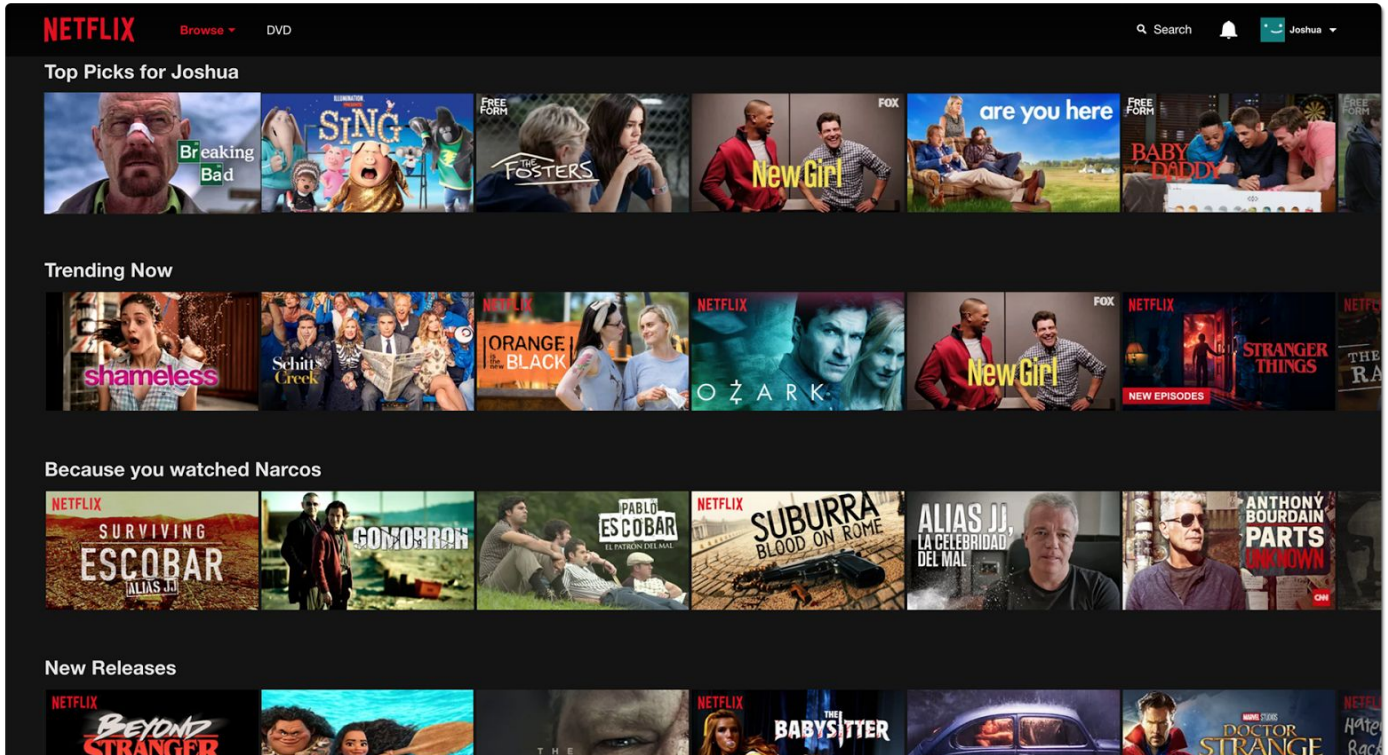




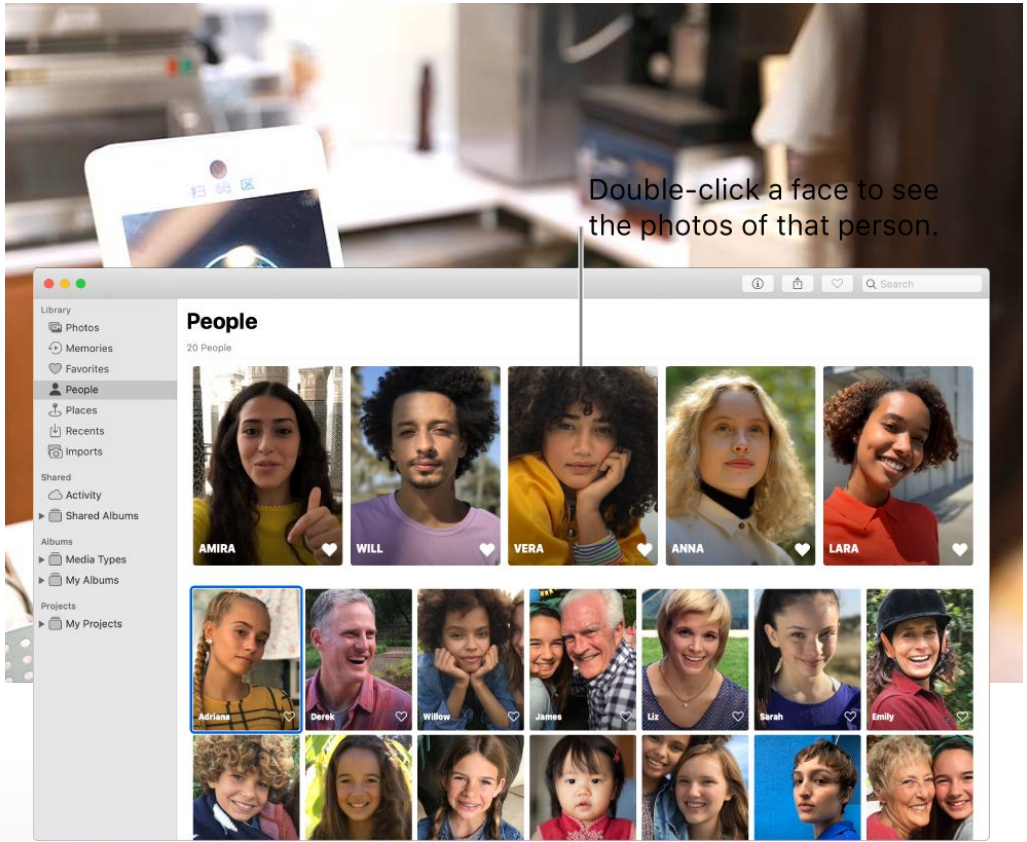
Recommendations



Recommendations

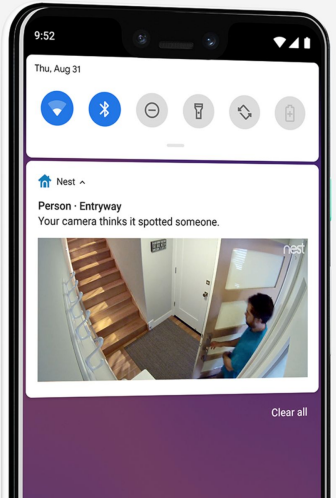


Face recognition



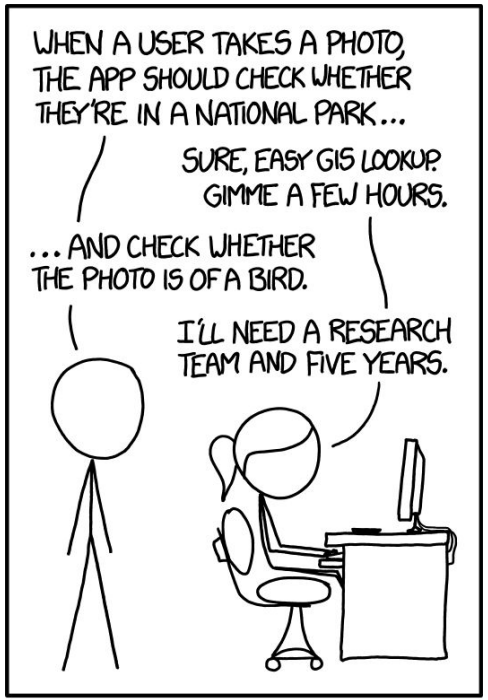
“Self-driving” cars

Smart home devices

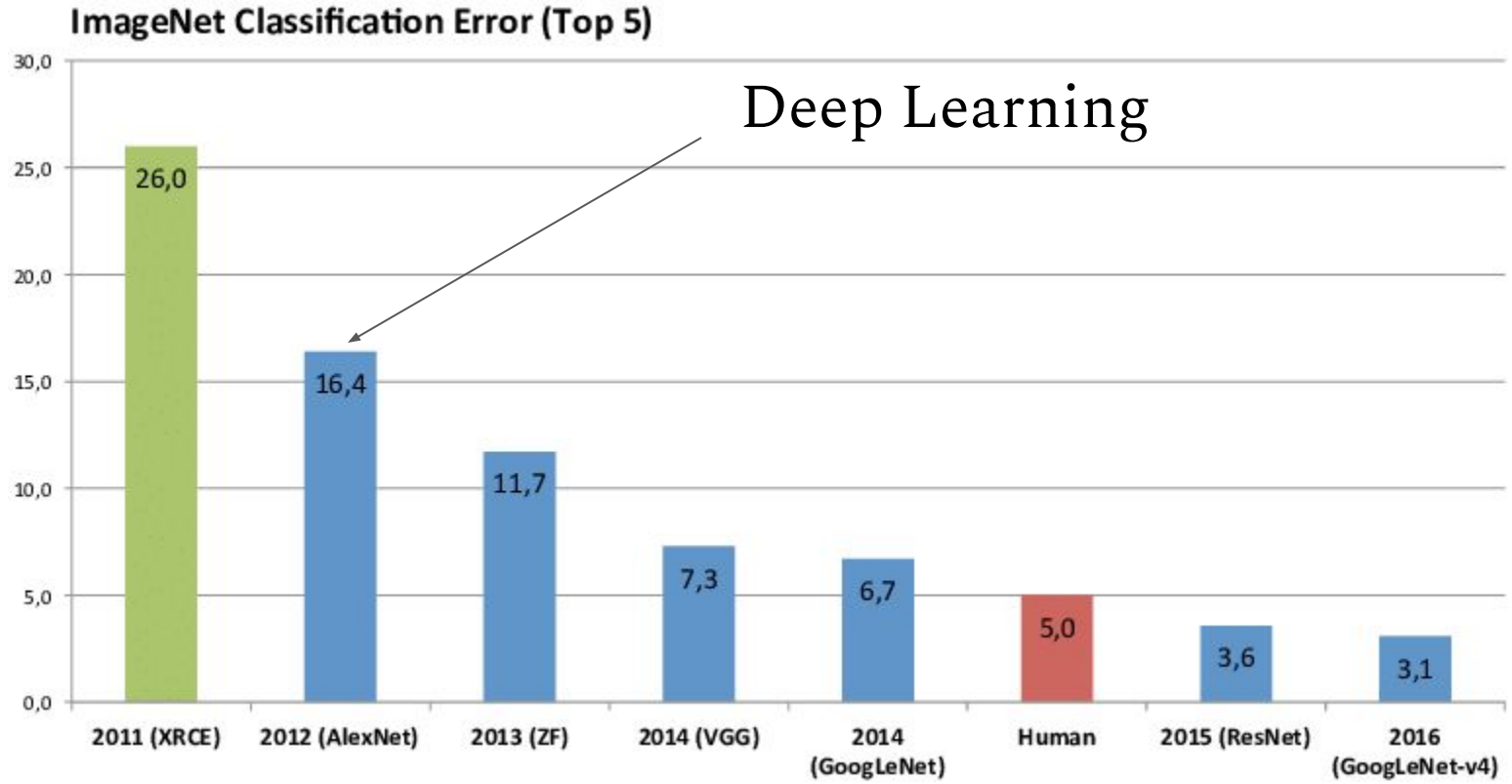


Automated photo-editing

State of image recognition (2014)



Improvements in Image Recognition



AlphaGo (2016)

Recent improvements in Machine Learning

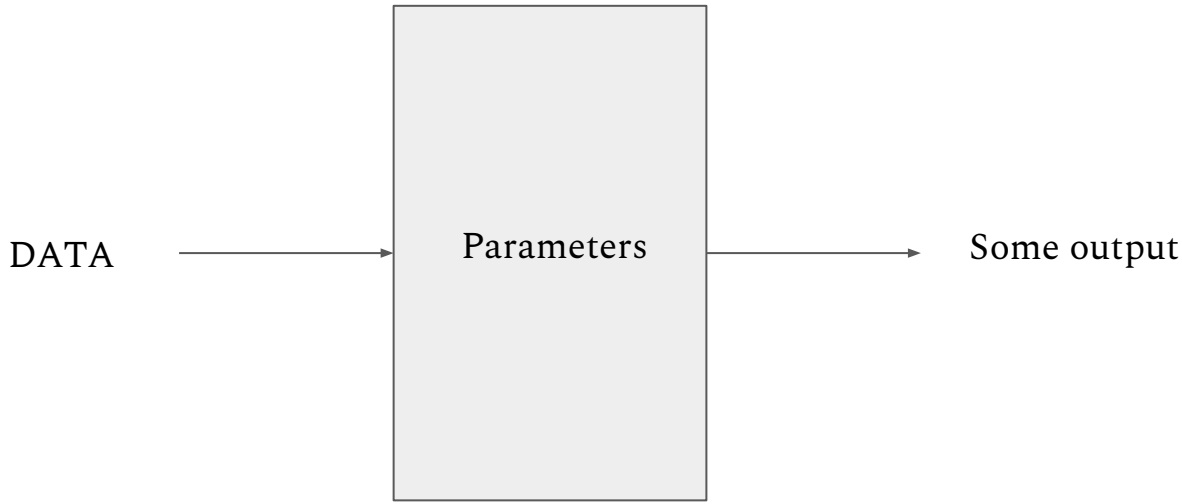
- New and improved methods
- More processing power (GPUs)
- Lots of more labelled data available and standard datasets
- More mature tooling

Intro to p5.js

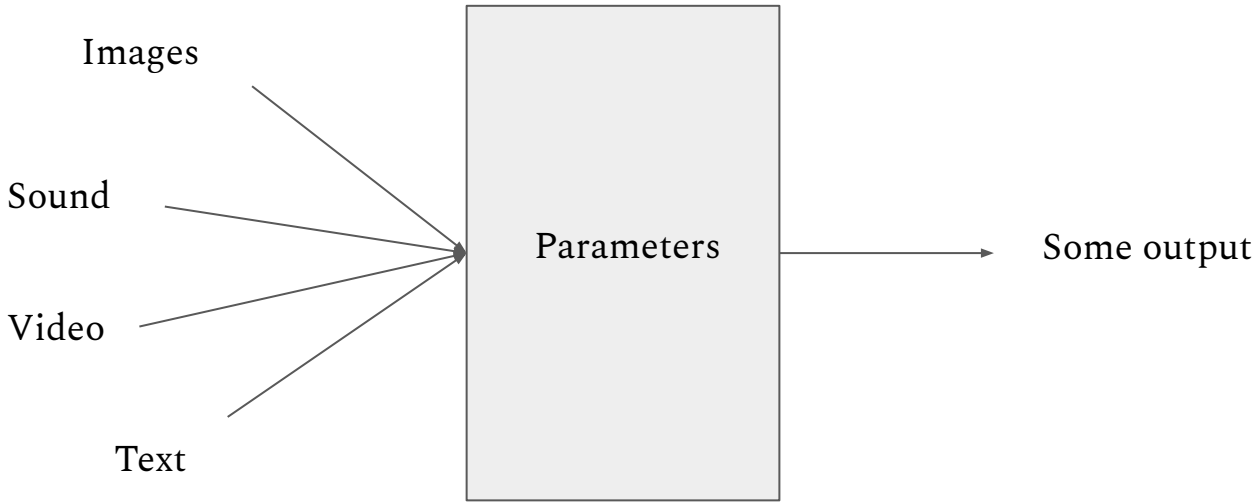
https://p5js.org/

Intro to Machine Learning

What is a model?



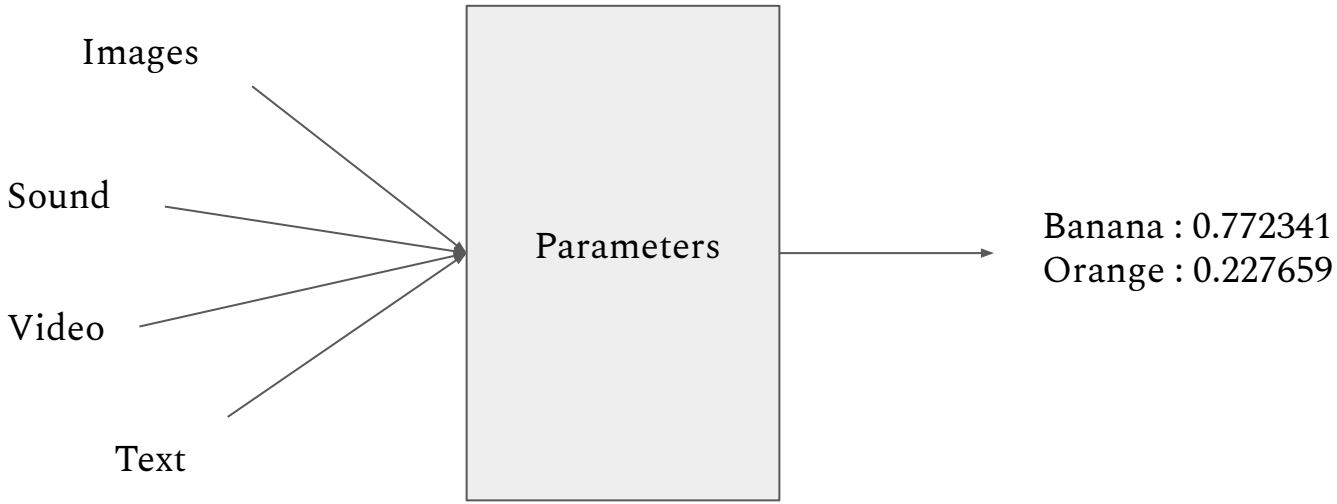
What is a model?



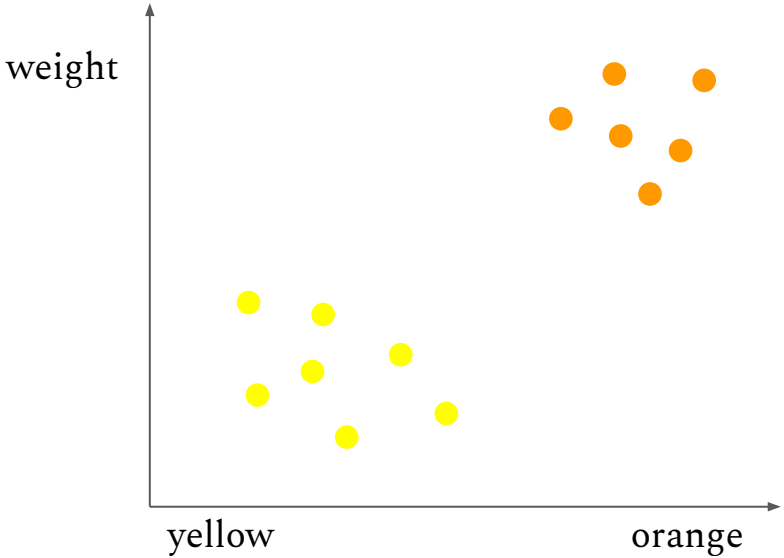
What is classification?

- Predicting what *class* an object belongs to based on given input

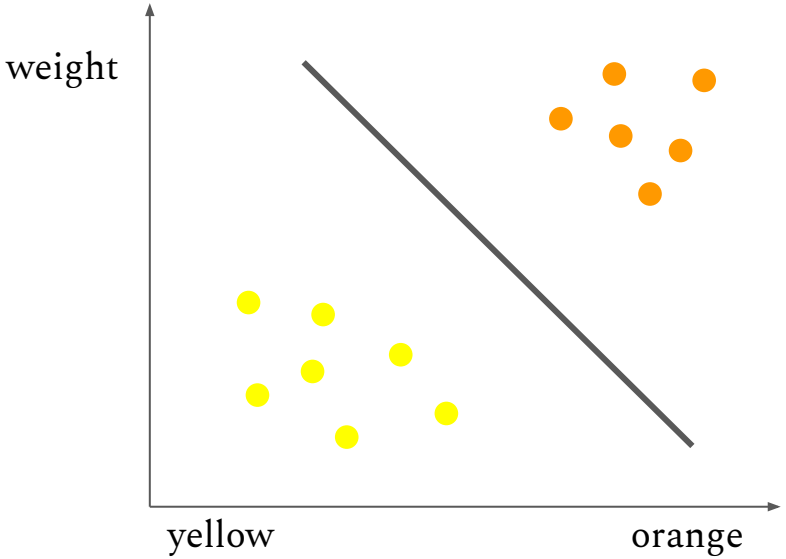
Classification model



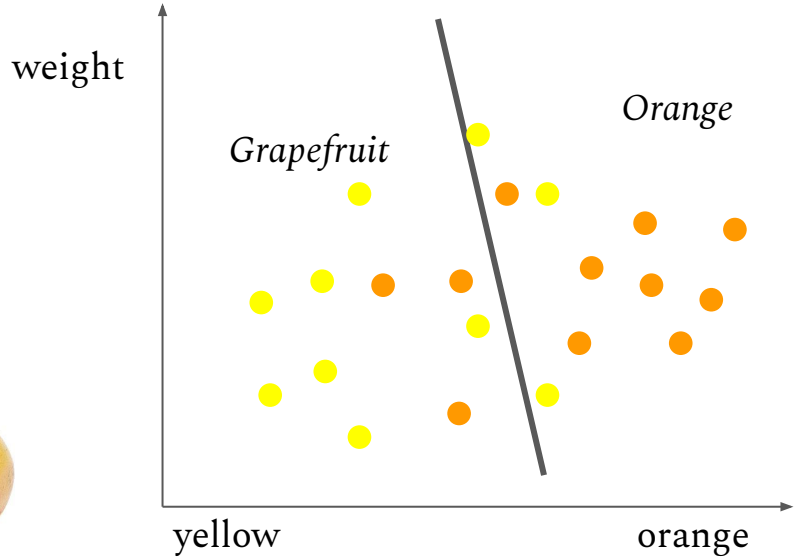
Example of simple classification



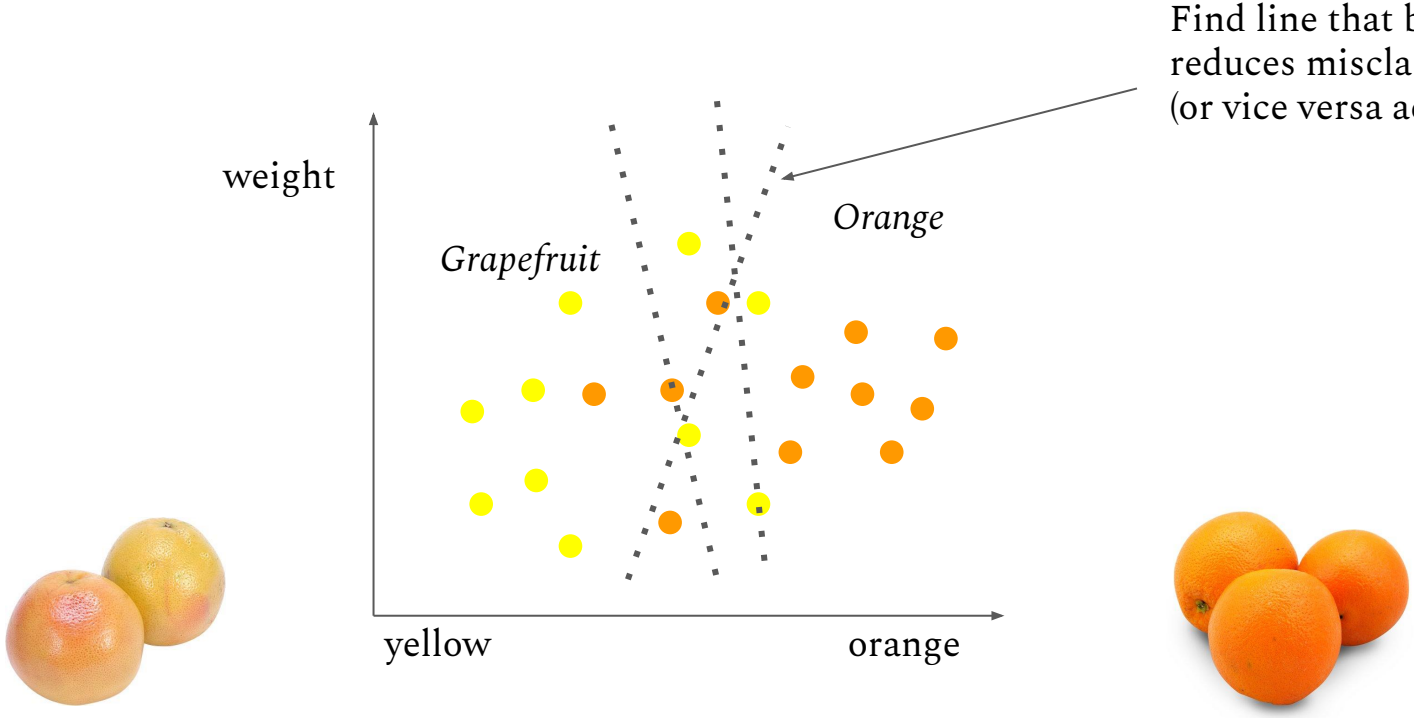
Linear separation



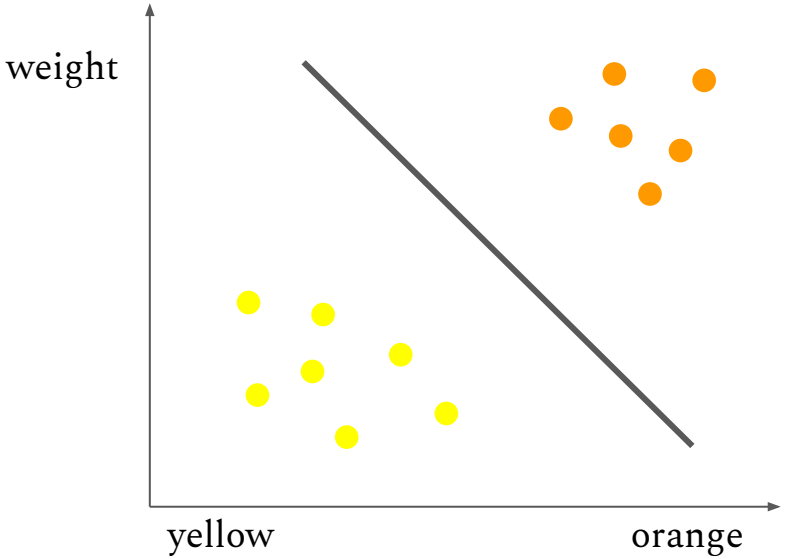
Misclassifications



“Training” the model



100% accuracy



70% accuracy

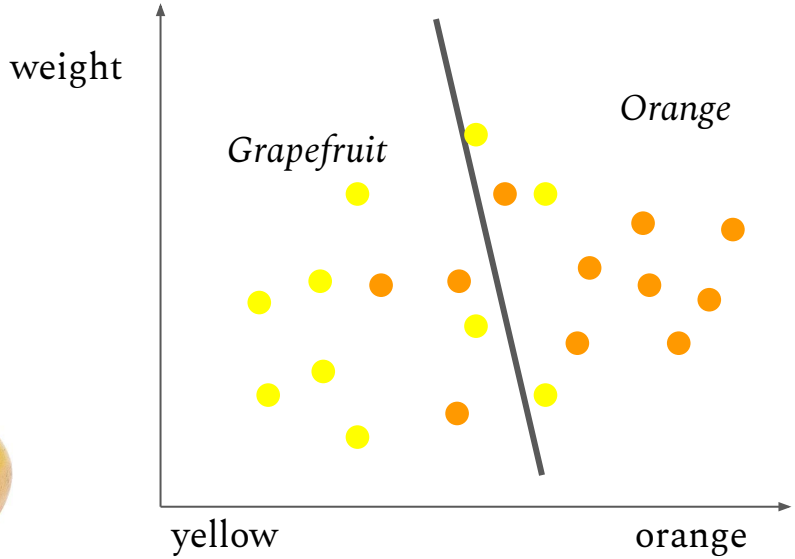
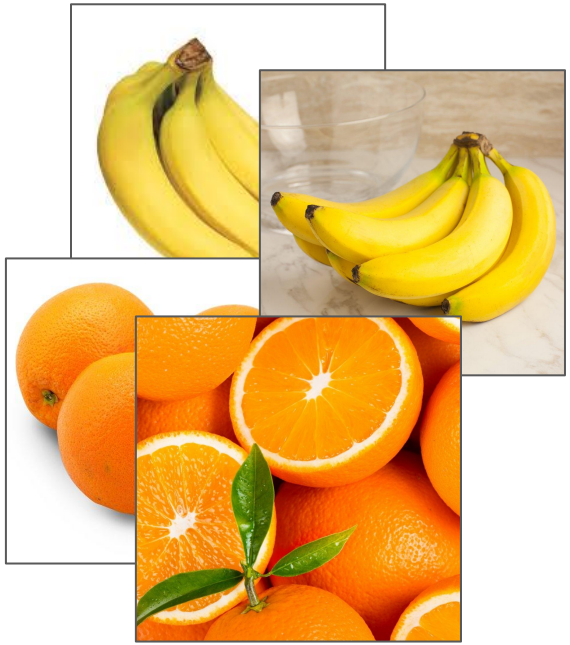
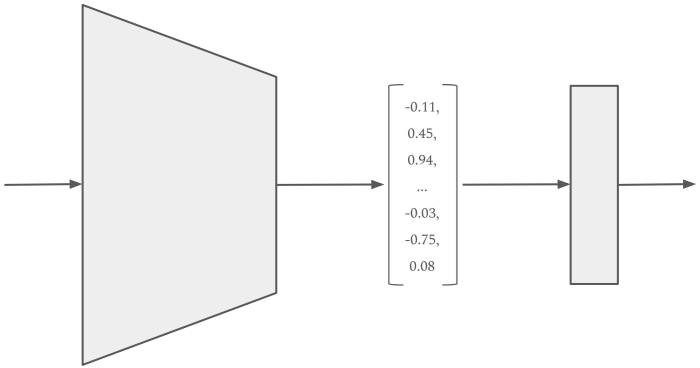


Image classification

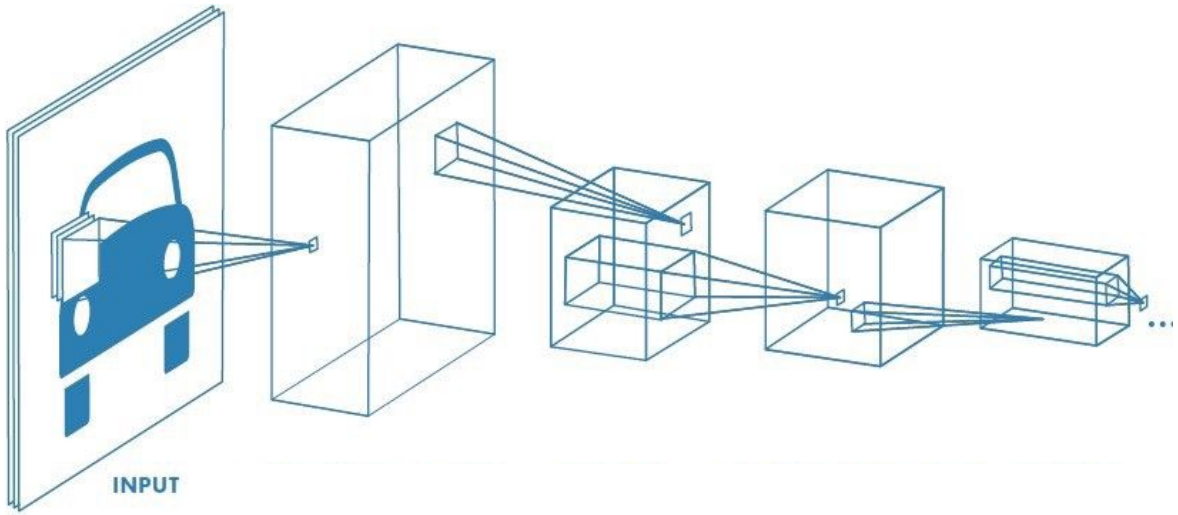




Convolutional
Neural
Network

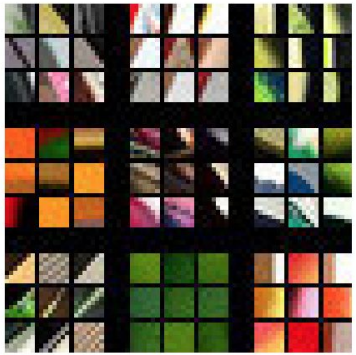
Banana or Orange?

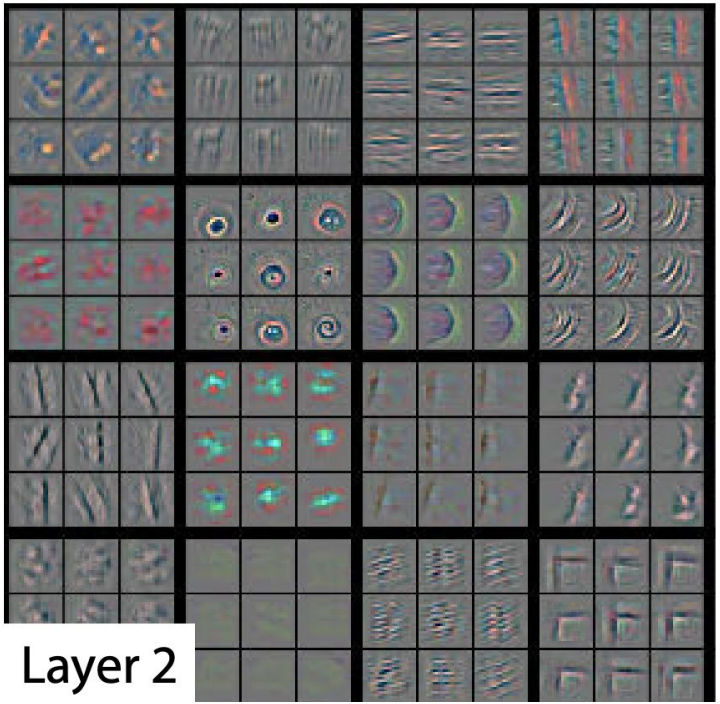
Convolutional neural networks (images)



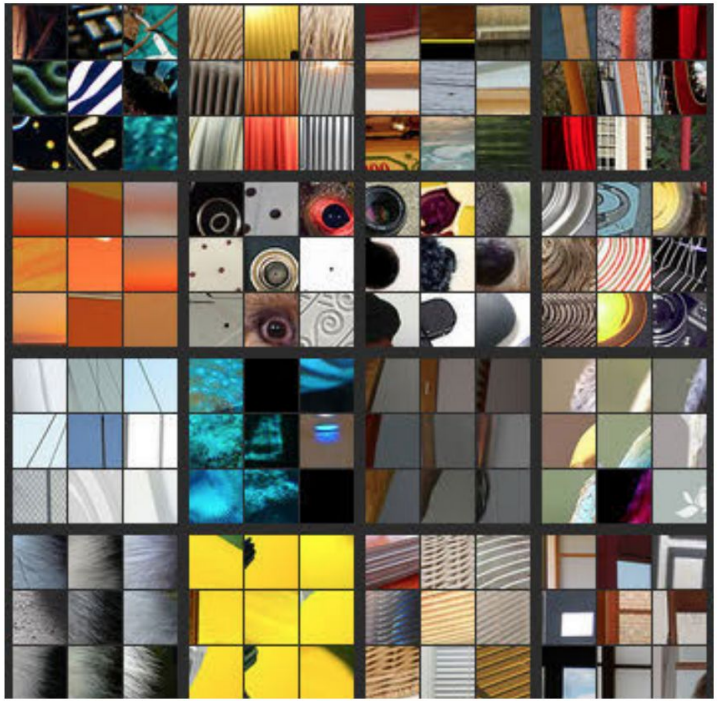
Convolutional neural networks

Layer 1

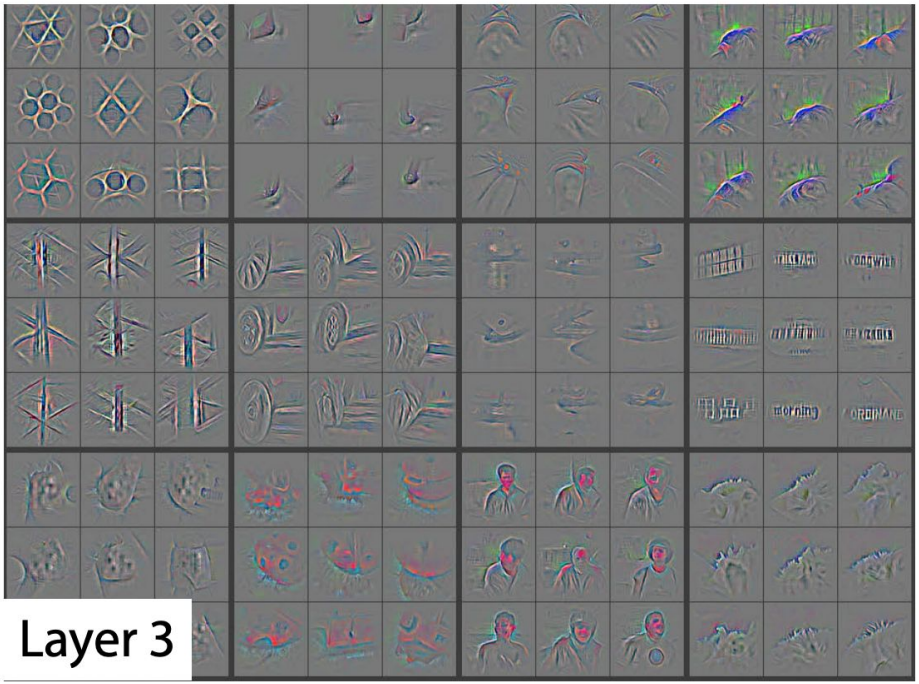


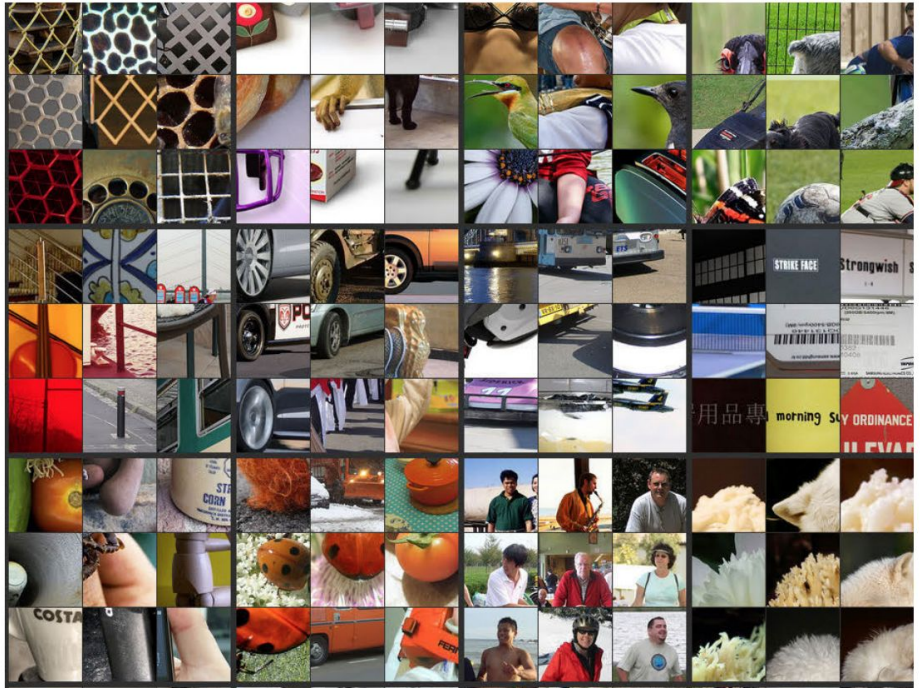


Layer 2

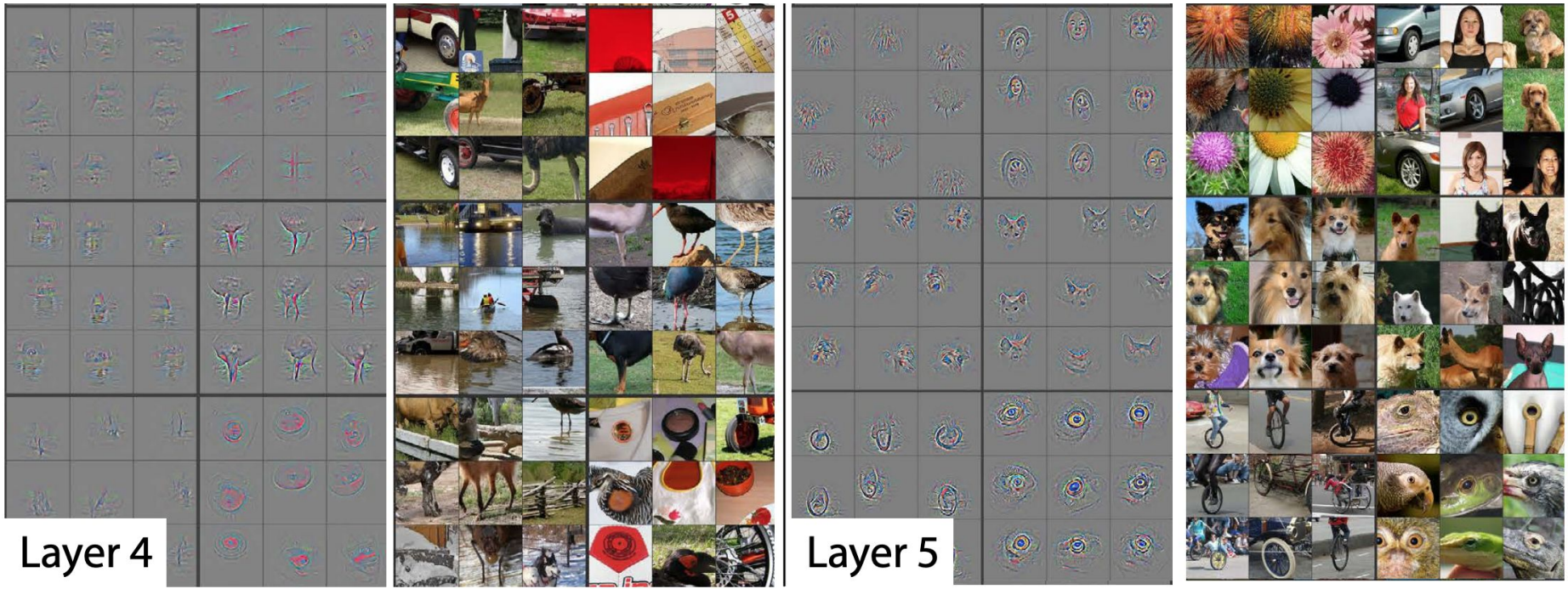


Convolutional neural networks



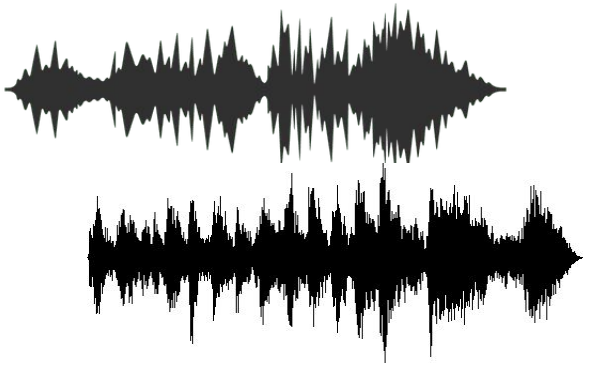


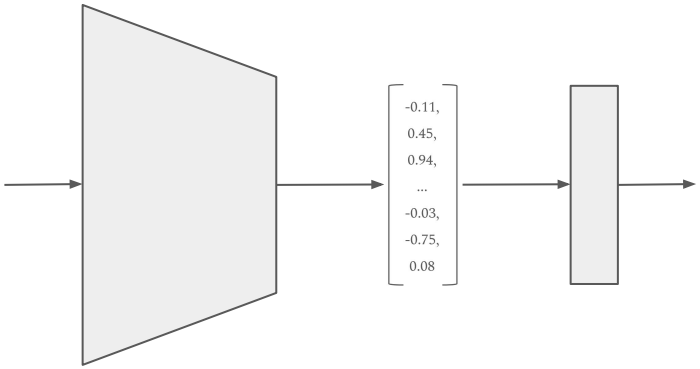
Convolutional neural networks



Source : “Visualizing and understanding Convolutional Networks”, Zeiler et al (2013)

Sound classification



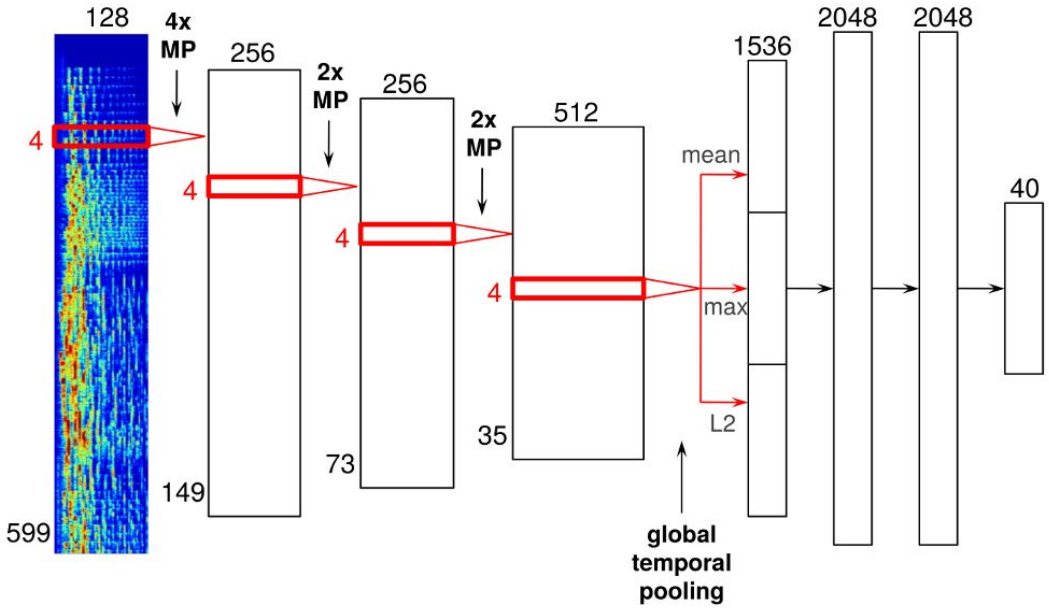


Convolutional Neural Network

Which voice command?

And also video etc.

Convolutional neural networks (sound)

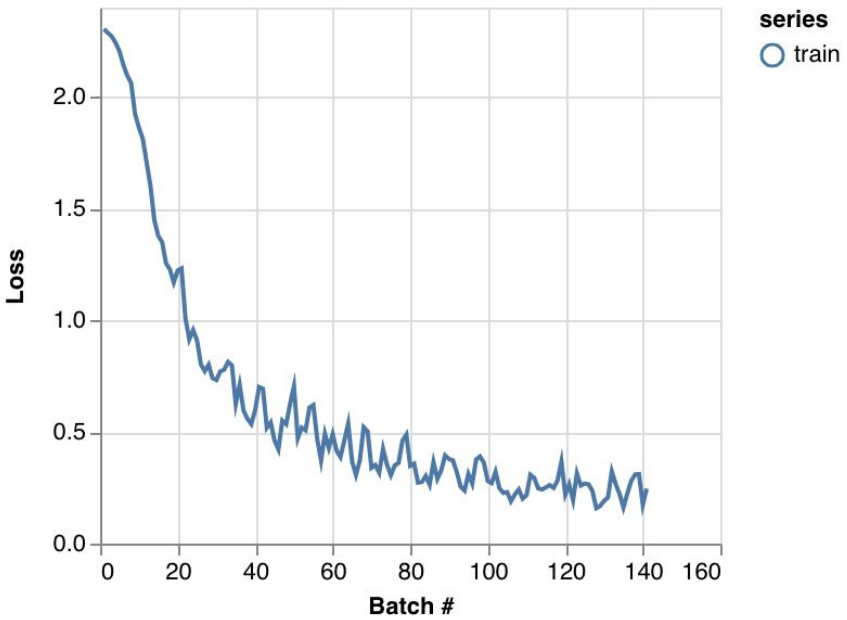


IMAGENET Dataset

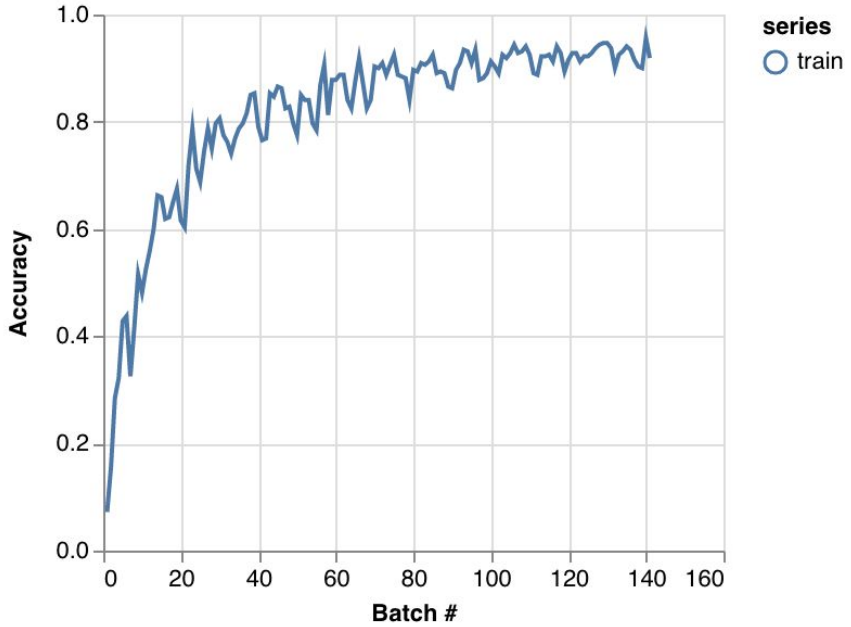
- Over 24 million labelled images
- Collected by Fei-Fei Li ++ at Stanford (2009)
- Many different classes
- 1000 classes are often used as reference

Training on large datasets

last loss: 0.246



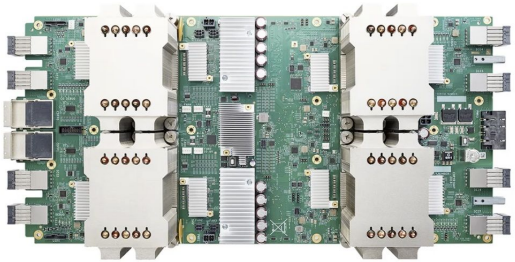
last accuracy: 91.9%



Training hardware

CPU

GPU

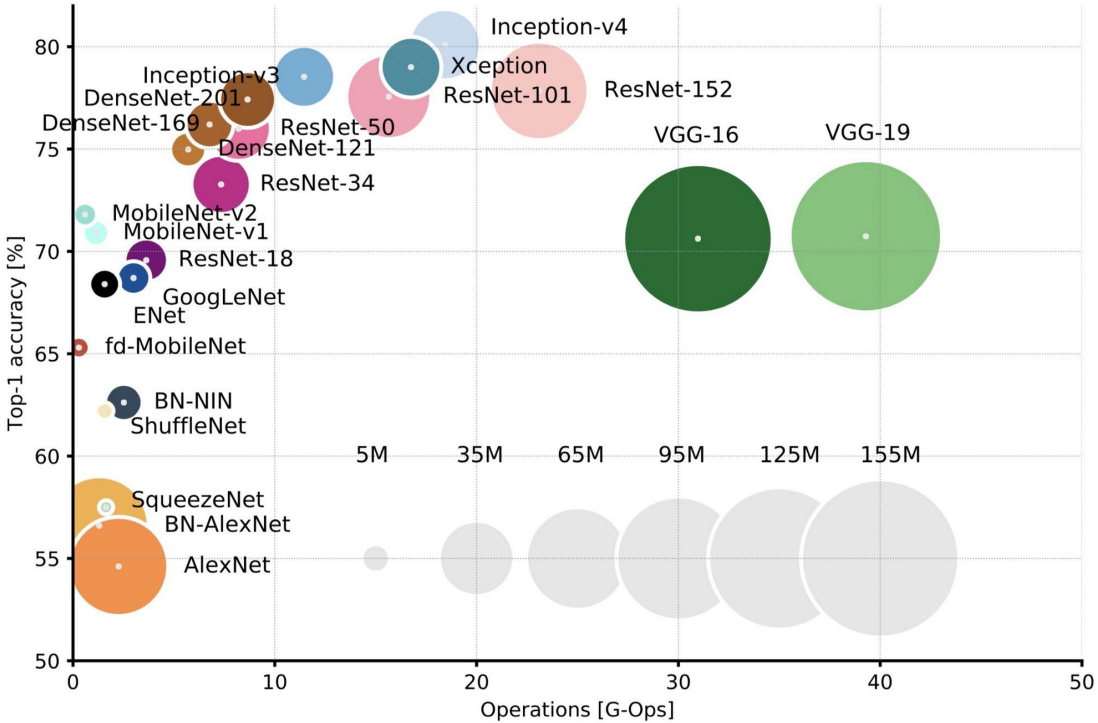


TPU

(2016)

Some common models

- Inception (2014)
- ResNet (2015)
- WaveNet (2016)
- MobileNet (2017)
- Transformer



Source : https://towardsdatascience.com/neural-network-architectures-156e5bad51ba

Common frameworks for ML

(Facebook)
Python, C++

(Google)
Python, C++, Java, Javascript

+ many, many more

Intro to ml5js

https://ml5js.org/

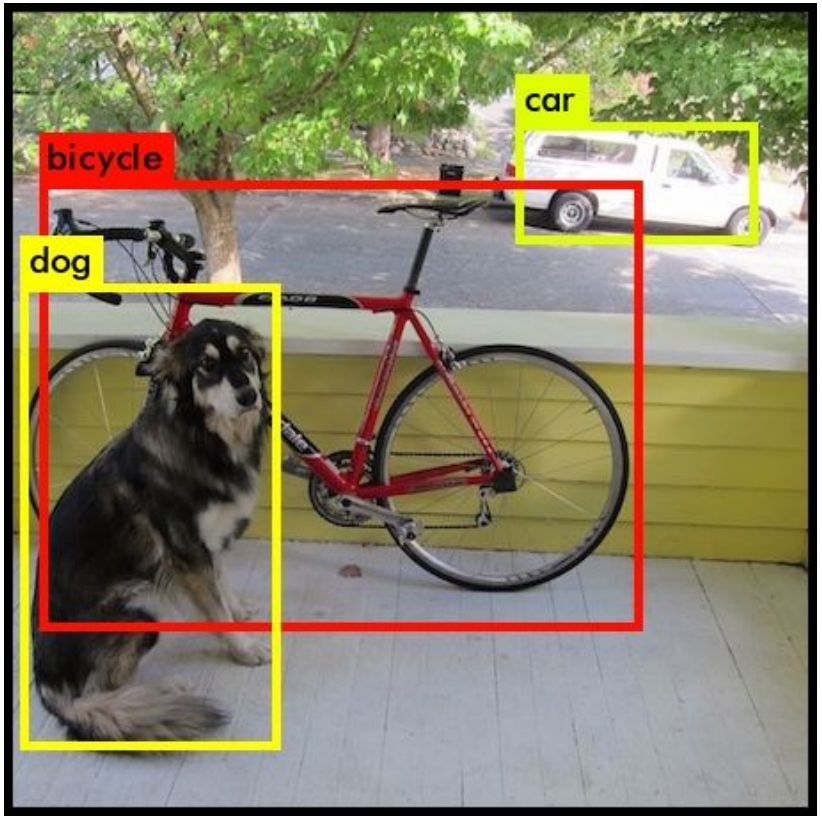
Lunch

Image classifier

- “MobileNet”
- Trained on ImageNet dataset
- Classifies image into 1000 different classes

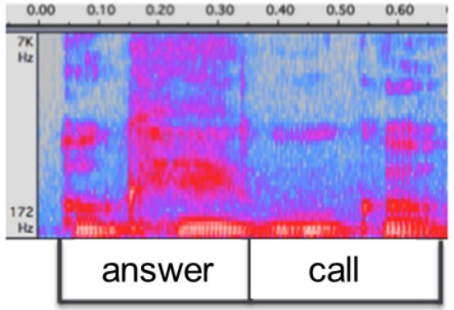
Object detection

- “YOLO” (You only look once)
- “Coco-SSD” (Coco Single-shot multibox detector)
- Localizes and classifies objects
- Detects 91 different classes from the COCO dataset (“Common Objects in Context”)



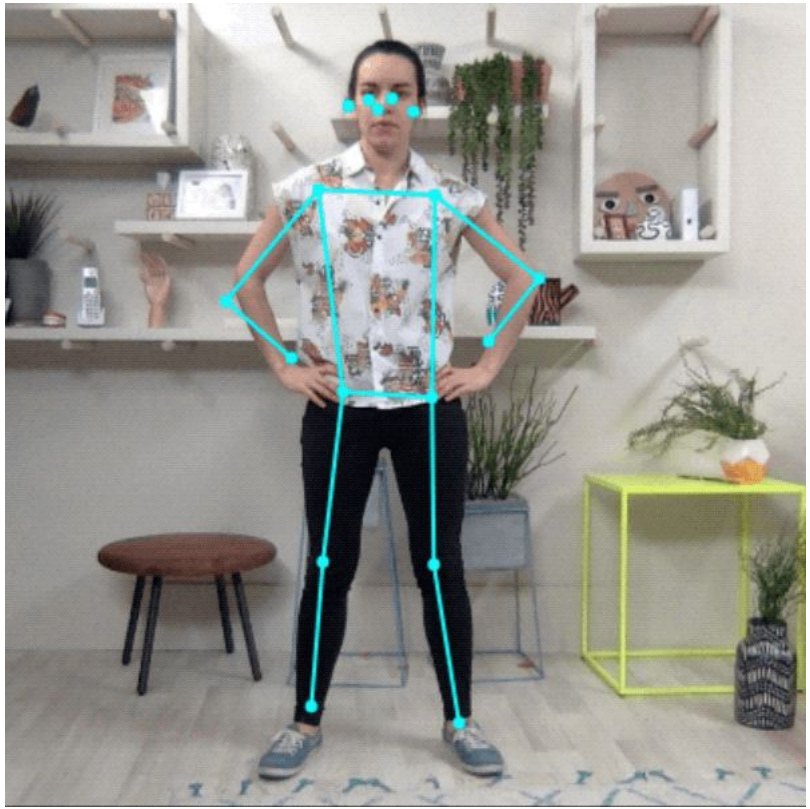
Sound classifier

- Detects 1-second keyphrases
- 18 words: “zero”, “one”, “two”, “three”, “four”, “five”, “six”, “seven”, “eight”, “nine”, “up”, “down”, “left”, “right”, “go”, “stop”, “yes” and “no”
- Converts sound to spectrogram image and uses small convolutional network

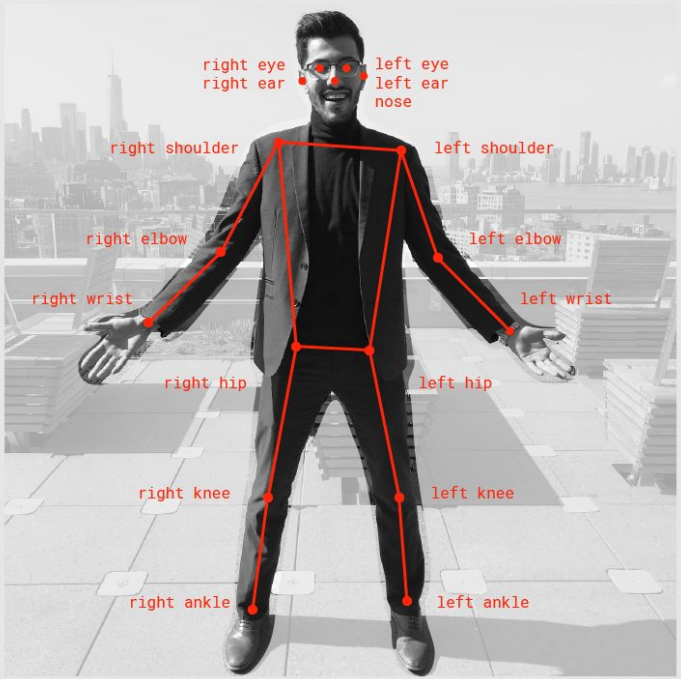


PoseNet

- Detects 17 body keypoints
- Is able to track both one person and multiple people

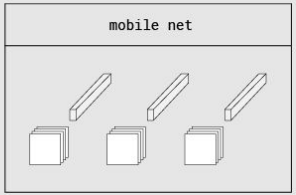


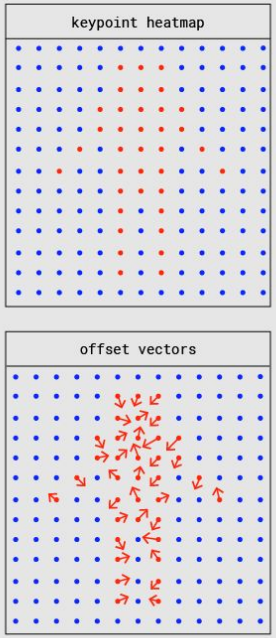
17 Pose Keypoints
Returned by PoseNet

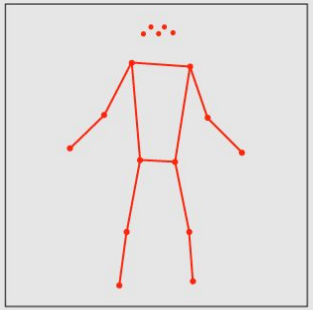


Single-Pose
Detection Algorithm

input image







pose estimation

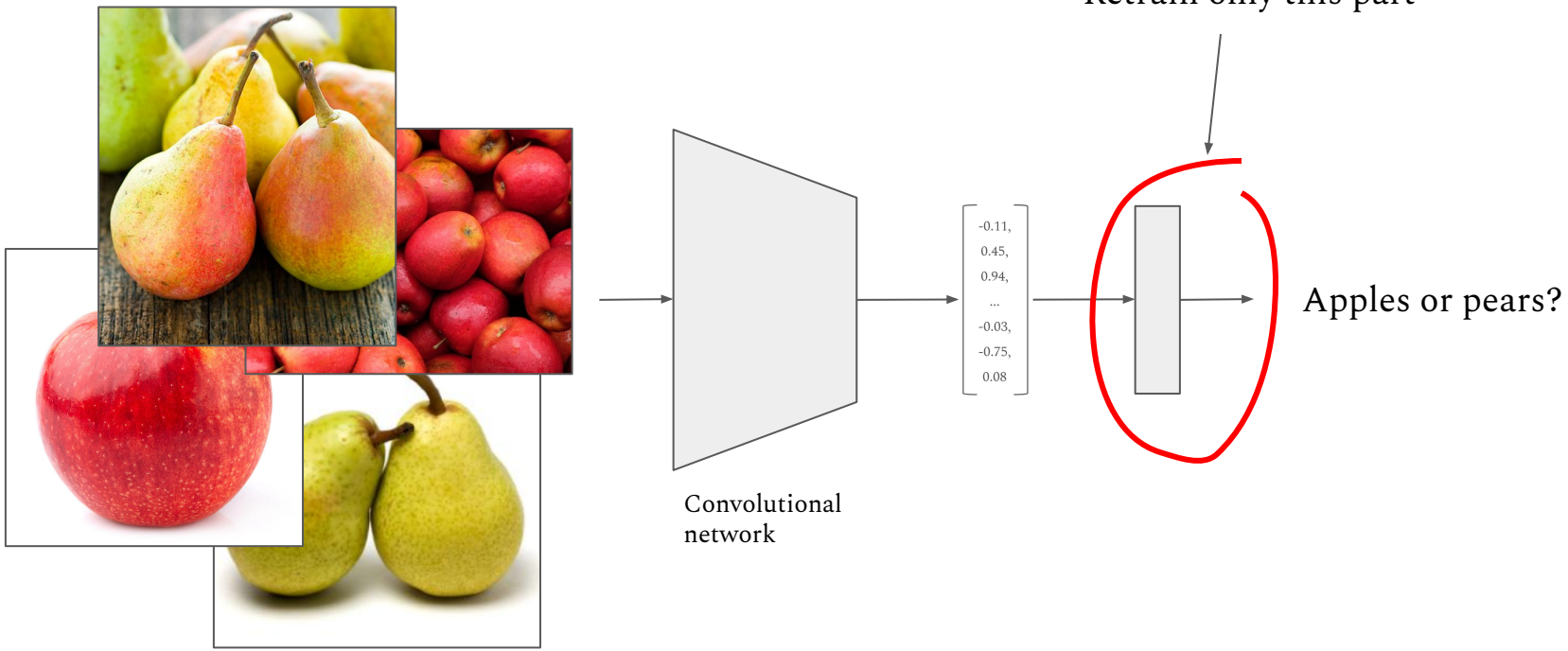
PoseNet model

Pose Confidence Scores vs.
Keypoint Confidence Scores

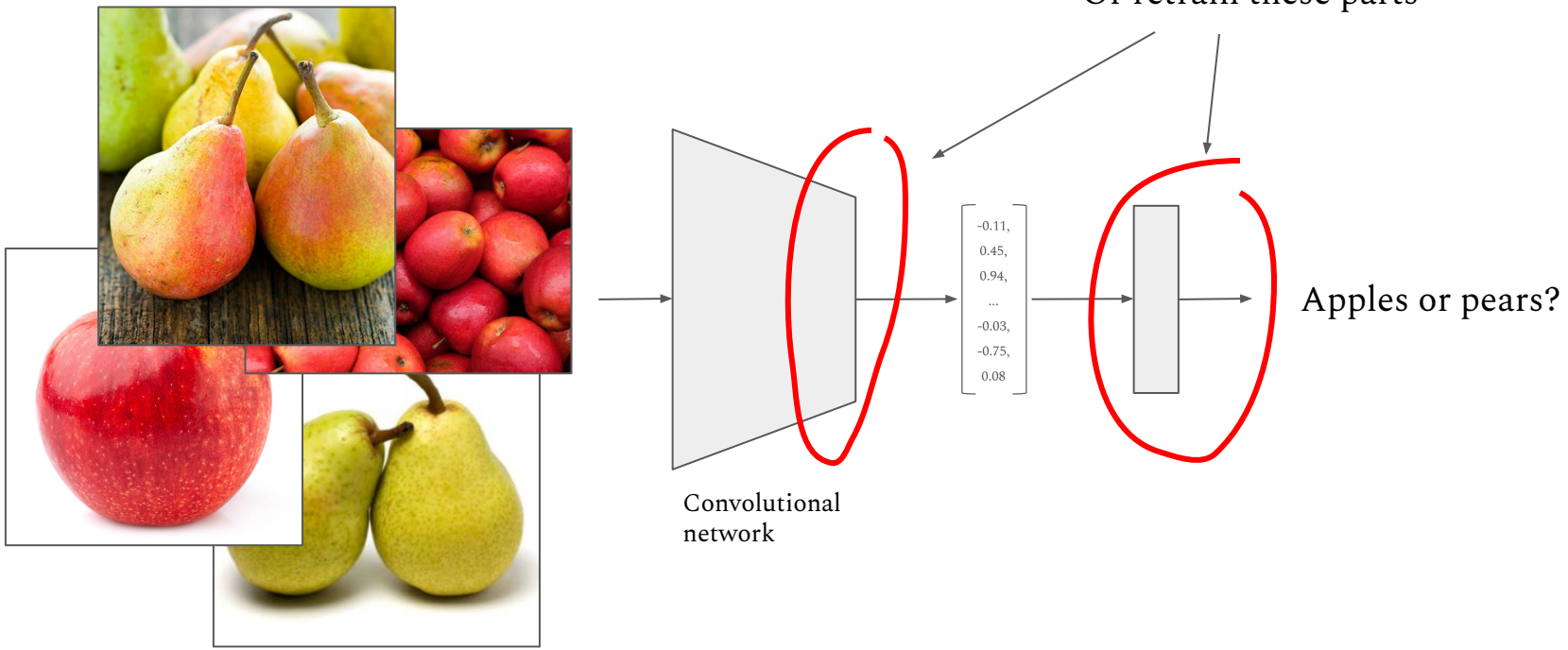
Hands-on

Training models

Retraining a model



Retraining a model



Transfer learning

- Repurposing the model trained on one dataset for a different yet related task
- Advantages:
  - Faster training time
  - Less data needed
  - (Often) more accurate

Teachable Machine

https://teachablemachine.withgoogle.com/

Some tasks

- Try to figure out cases where your classification model fails
  - For images :
    - Change background / room
    - Find edge cases : try to only show parts of the object
- Try to train classification model on very similar objects
  - How much data do you need to train the model?

Try to build something!