Computer Vision

Dominik Kuczkowski
Gradient Science Club 2022

Plan for Today

- What is computer vision?
- Quick history of computer vision
- Convolutions
- Convolutional Neural Networks
- Computer vision tasks
- Transfer learning

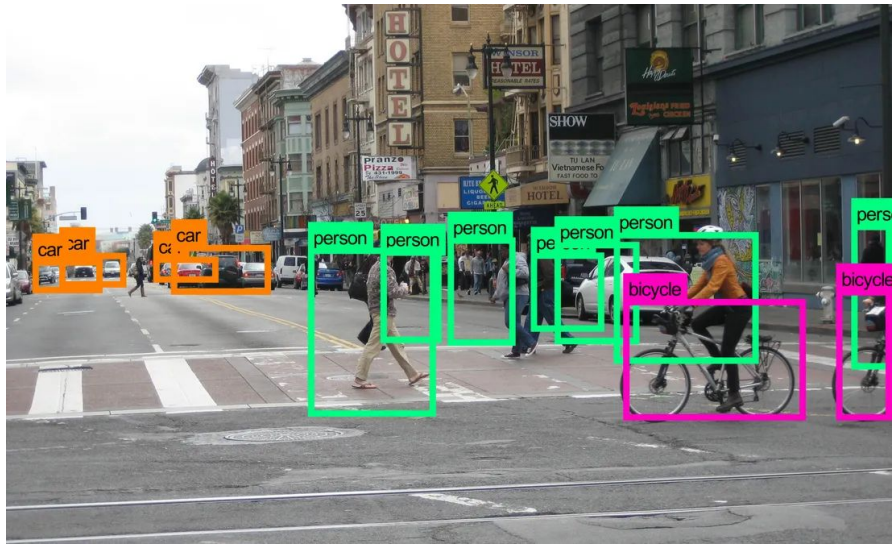
Resources

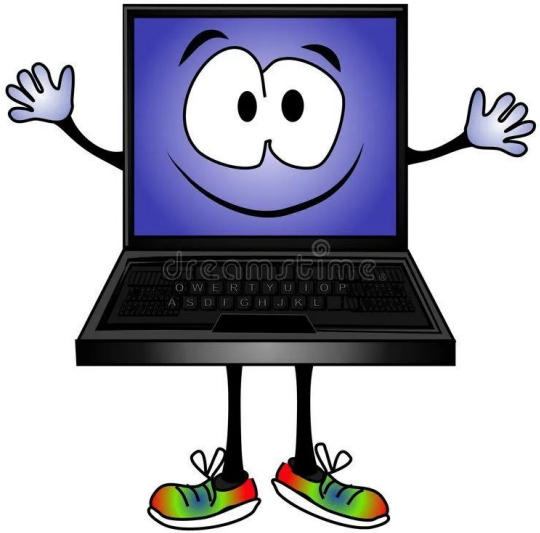
But what are convolutions? - 3blue1brown

CNNs explained - Futurology

What is computer vision?

"If AI enables computers to think, computer vision enables them to see, observe and understand."





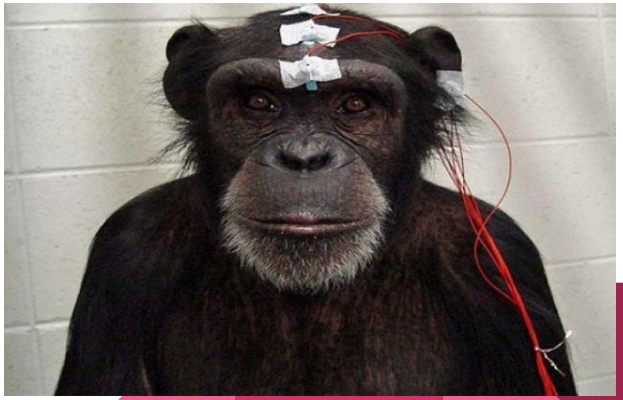
History of computer vision

Research on computer vision started in 1960s. Some researchers believed the problem could be solved in one summer – "The summer vision project".

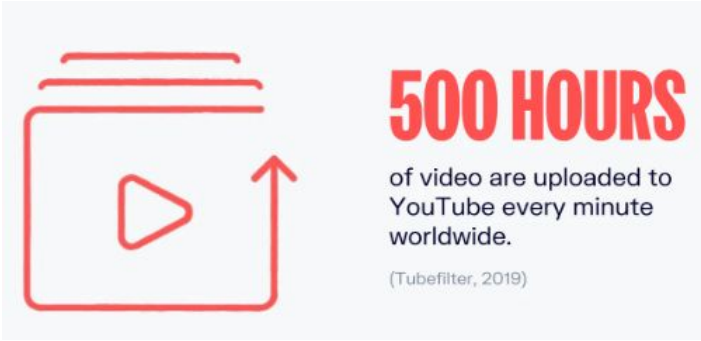
In later decades many algorithms were created for edge detection, optical flow, motion estimation etc.

In 1980s and 1990s neural network based solutions came to life.

Kunihiko Fukushima created – **Neocognitron**, which was inspired by work of Hubel & Wiesel from 1959.



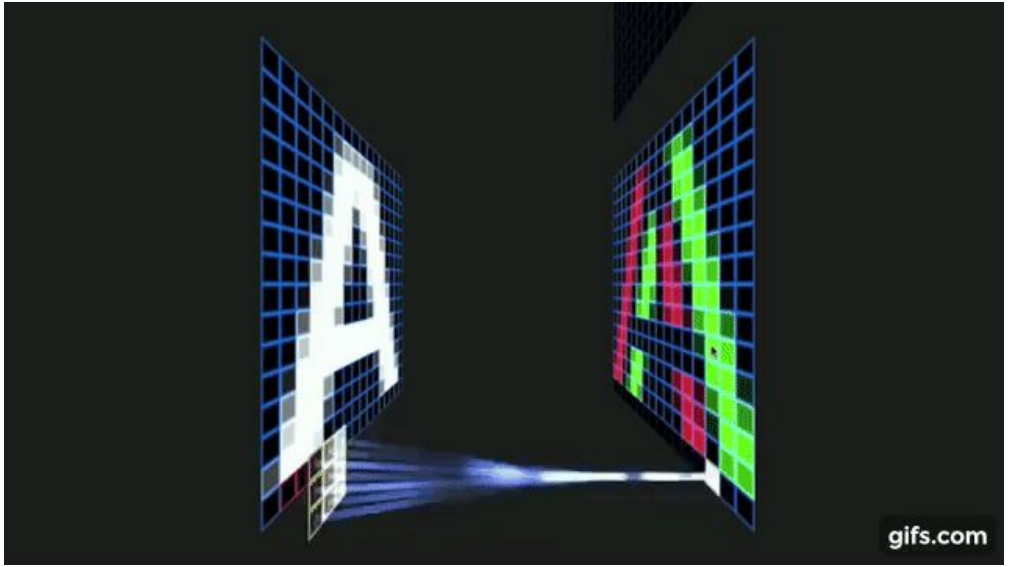
Why do we need computer vision?



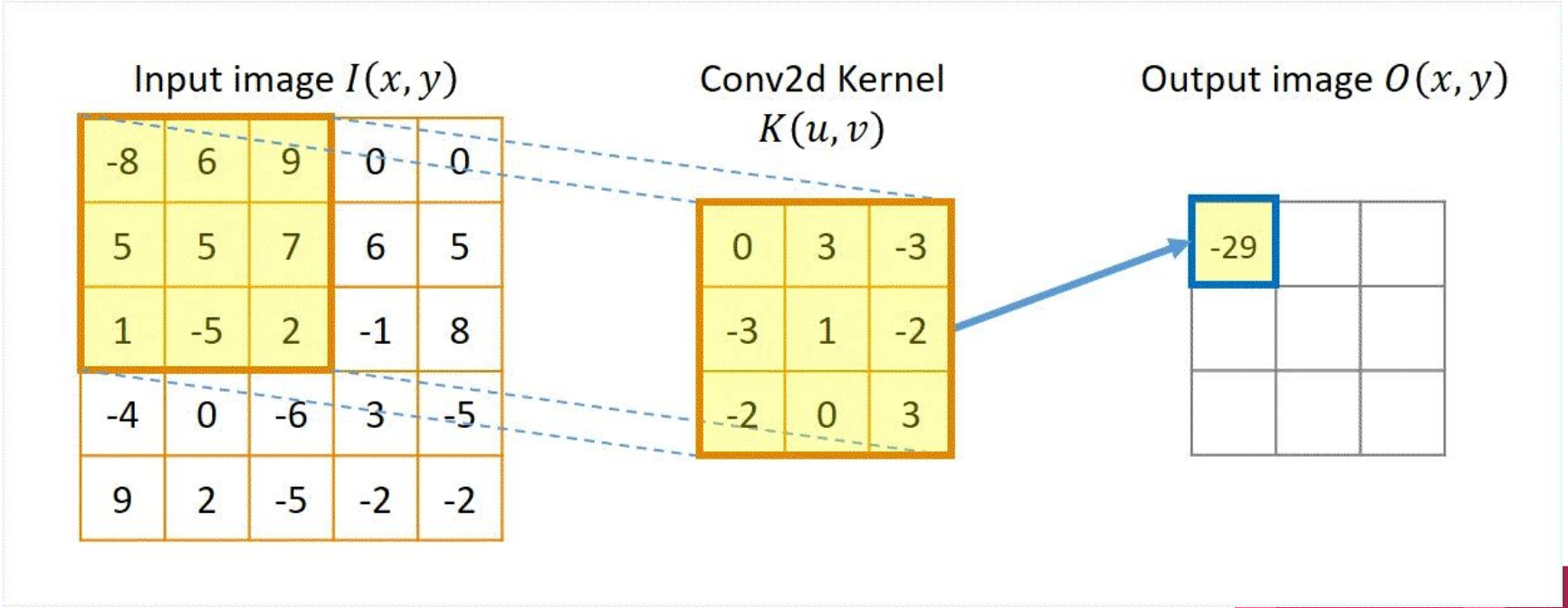
How to analyze all of this data?

The answer is:

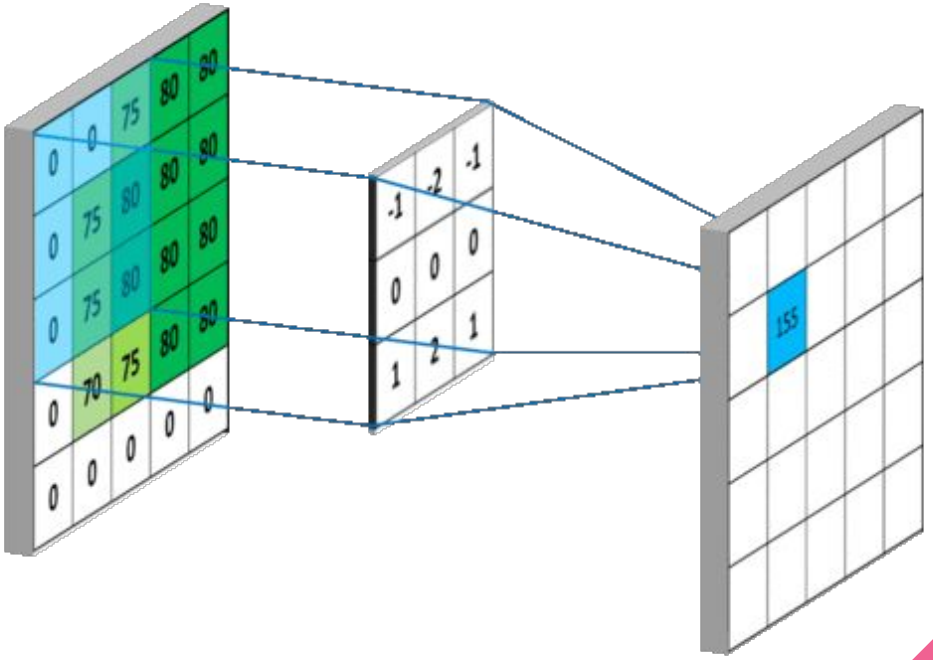
Convolutional Neural Networks.



What is a convolution?



What is a convolution?



Importance of the kernel

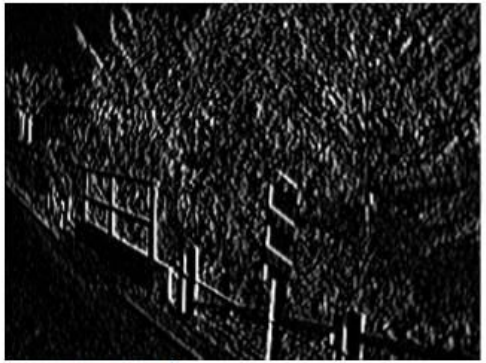
Convolutions are used in classic computer vision algorithms:

- edge detectors
- image blurring
- image sharpening

In Convolutional Neural Networks the parameters of the kernel are learned.

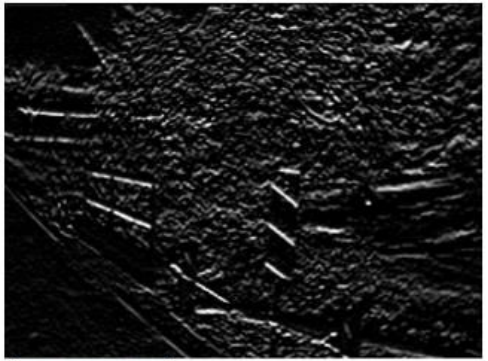
Sobel Edge Detection: Gradient Approximation

Note anisotropy of edge finding



1	0	-1
2	0	-2
1	0	-1

Horizontal diff.



1	2	1
0	0	0
-1	-2	-1

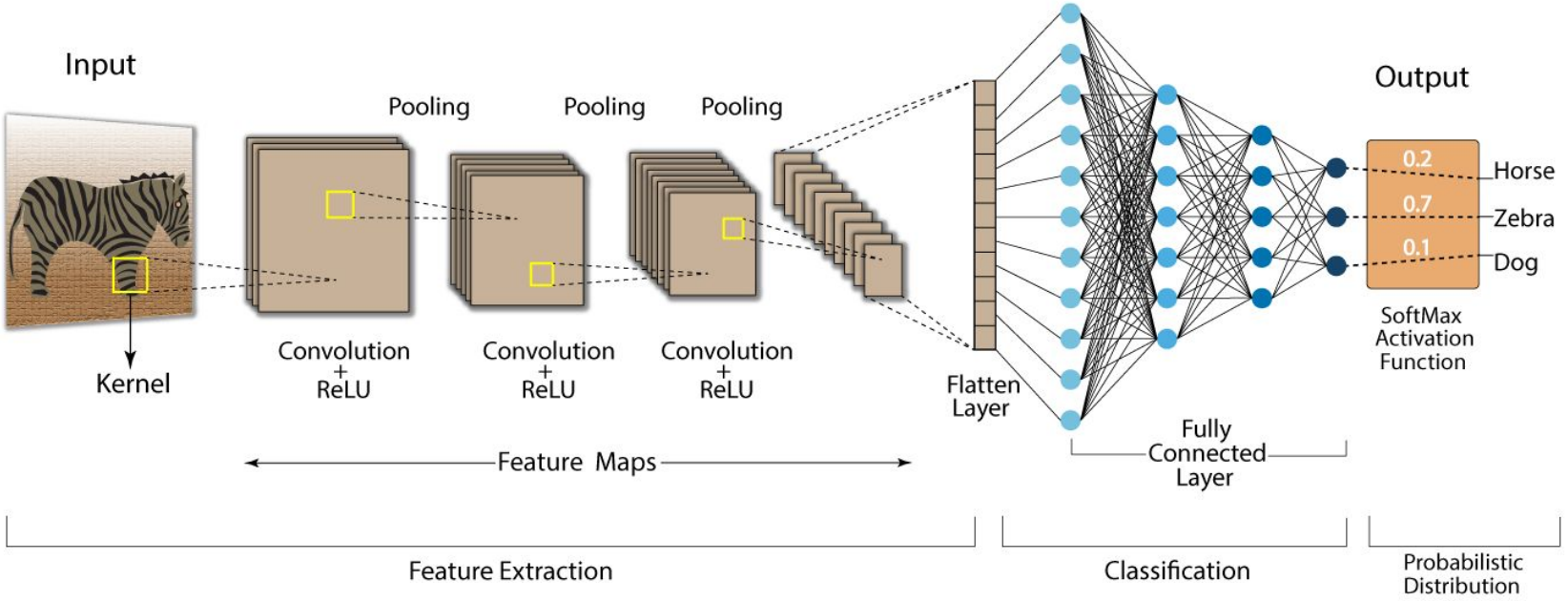
Vertical diff.

Convolutional Neural Networks

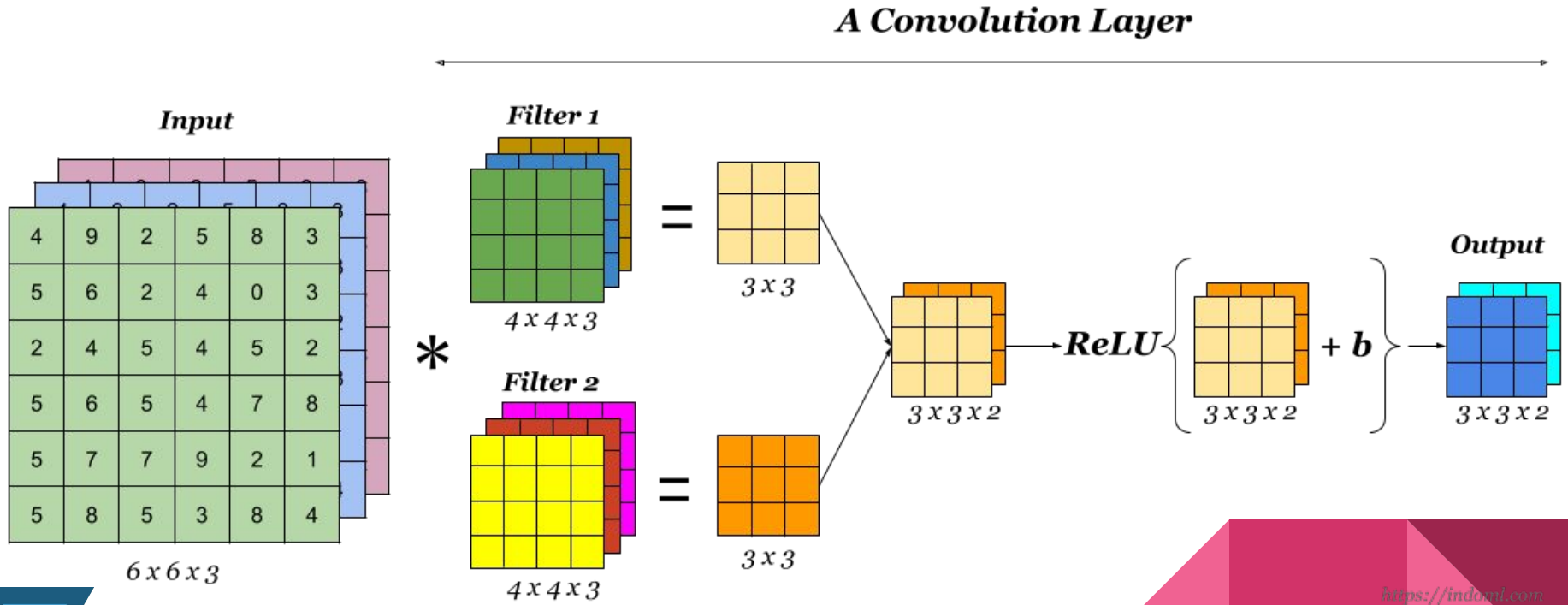
Consist of:

- **convolution layers**
- **non-linear activations**
- **pooling layers**
- **fully connected layers**

Convolution Neural Network (CNN)

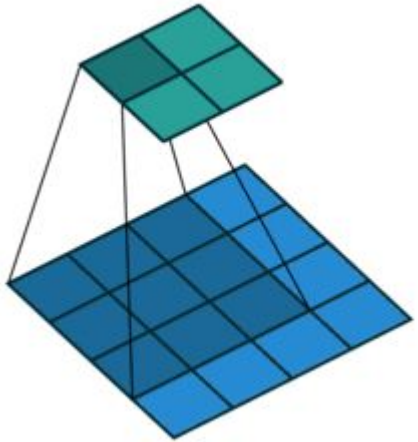


Convolution layers

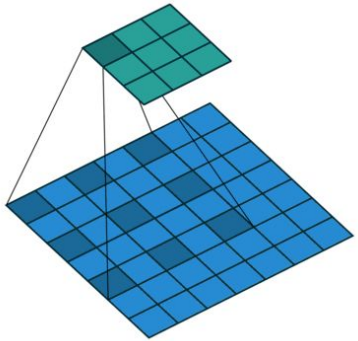
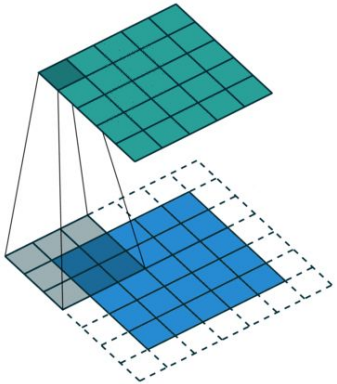
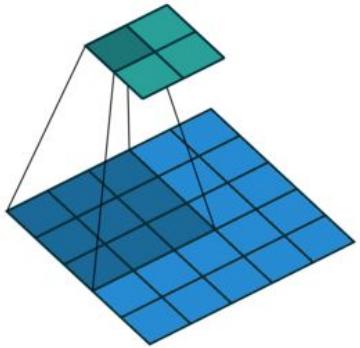
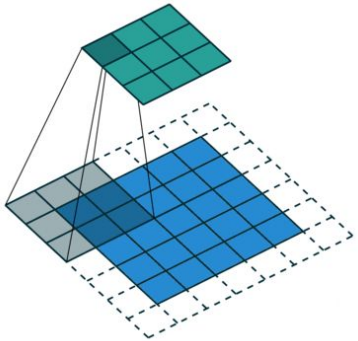


Parameters of convolution

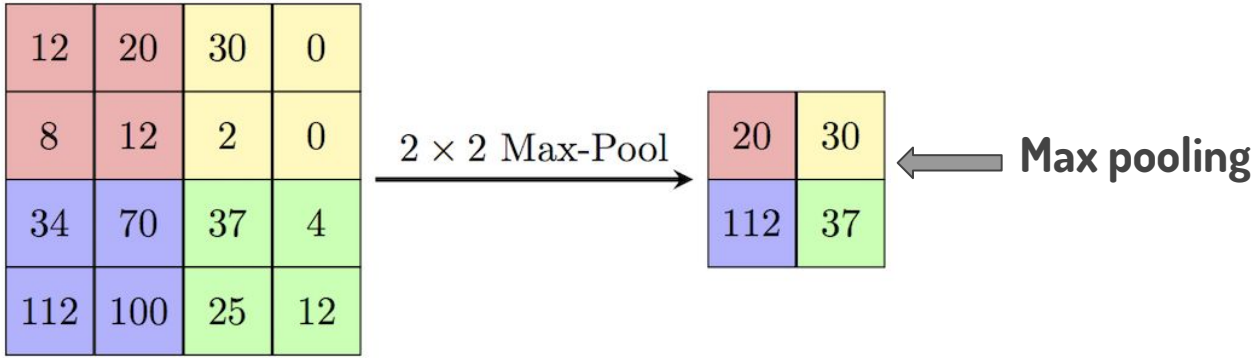
- kernel size
- output channels
- stride
- padding
- dilation

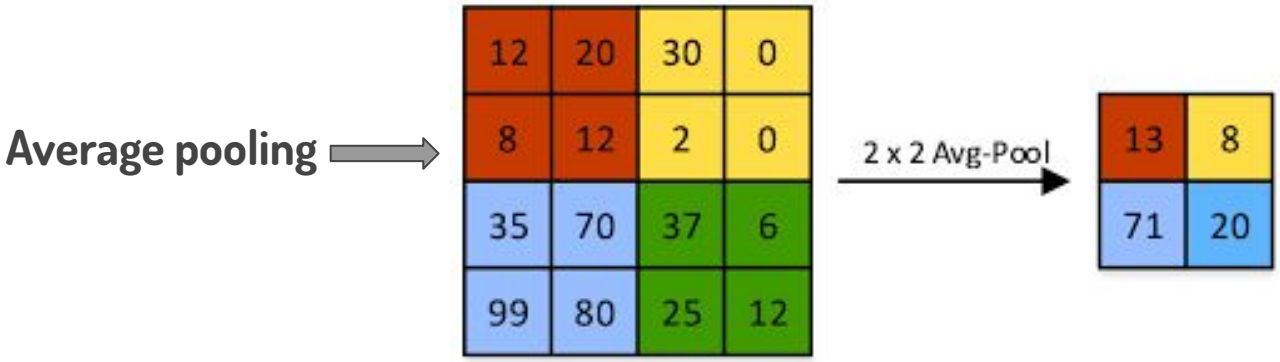


Parameters of convolution

Dilation	Padding = 1, no stride	No padding, stride = 2	Padding = 1, stride = 2
			

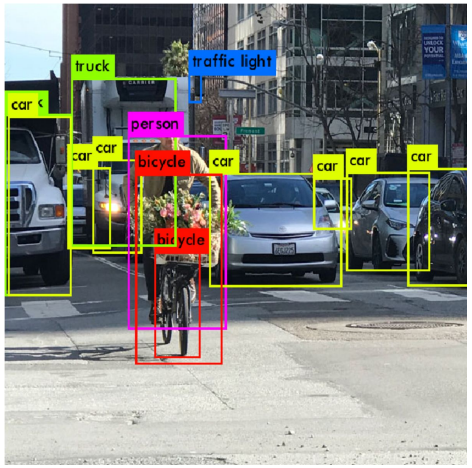
Pooling

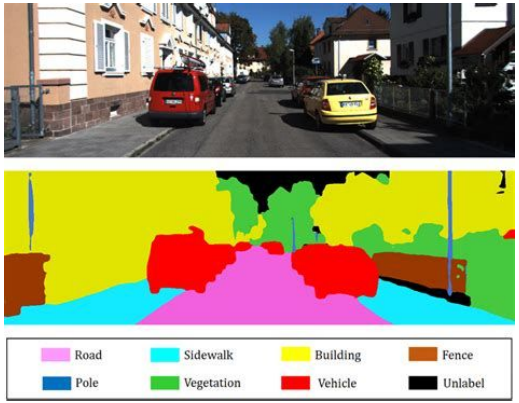

12 20 30 0
8 12 2 0
34 70 37 4
112 100 25 12
2 × 2 Max-Pool
20 30
112 37
Max pooling


Average pooling
12 20 30 0
8 12 2 0
35 70 37 6
99 80 25 12
2 x 2 Avg-Pool
13 8
71 20

Computer vision tasks

- classification
- segmentation
- object detection
- image captioning
- image generation

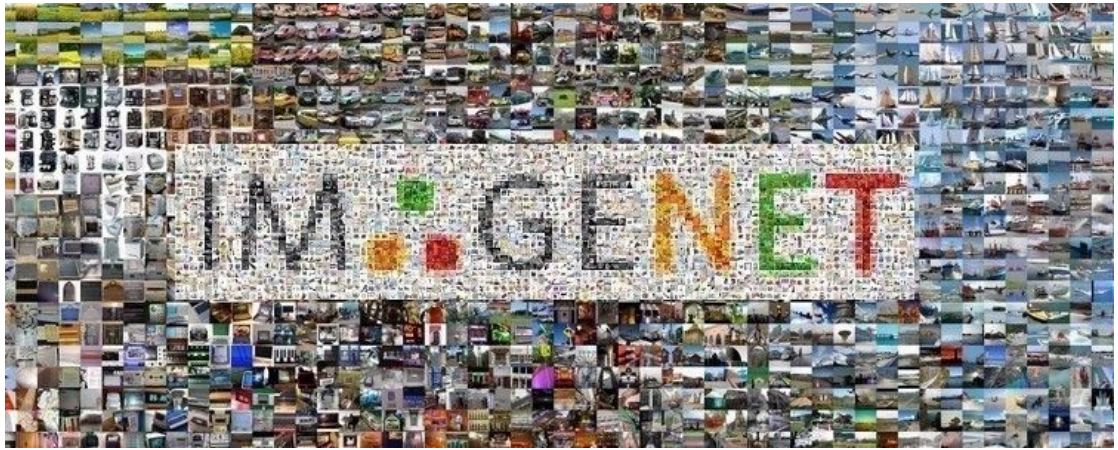




Popular CNN architectures for classification

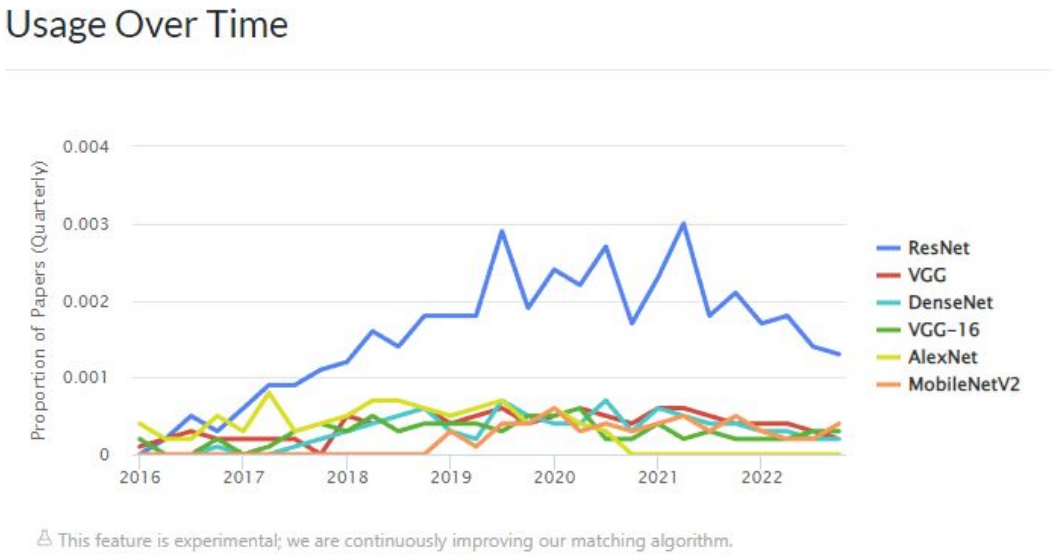
- AlexNet
- ResNet
- EfficientNet
- MobileNet
- Inception

ImageNet

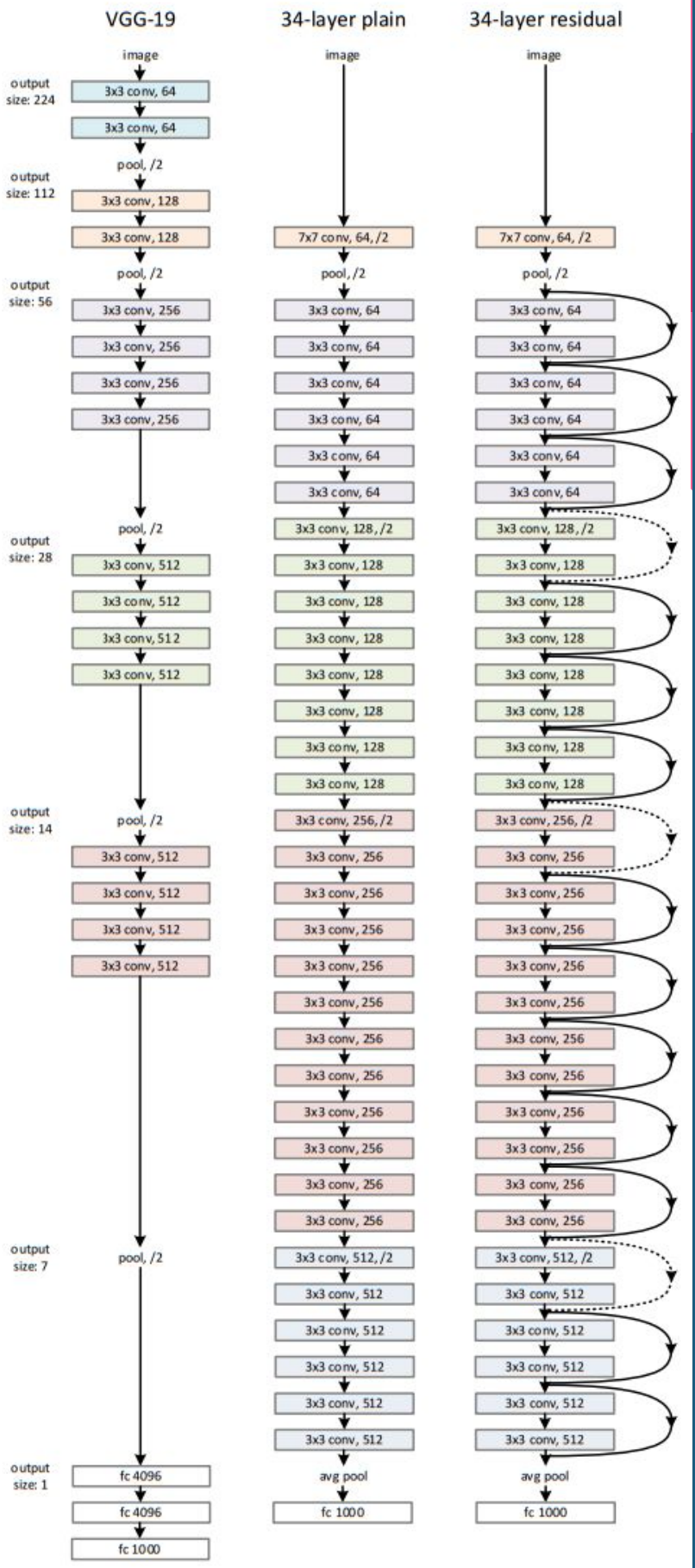


ResNet

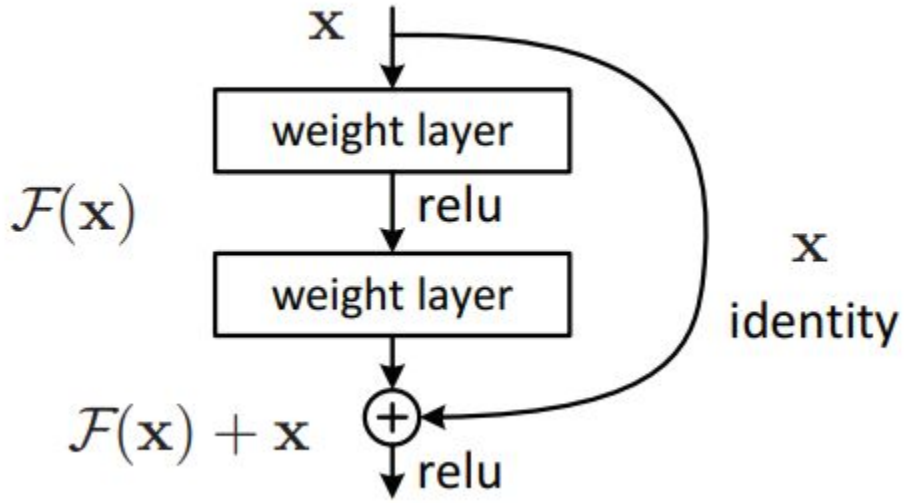
Proposed in 2015 by researchers from Microsoft. Used residual connections to solve the problem of vanishing/exploding gradients. Became very popular since then.



ResNet – architecture

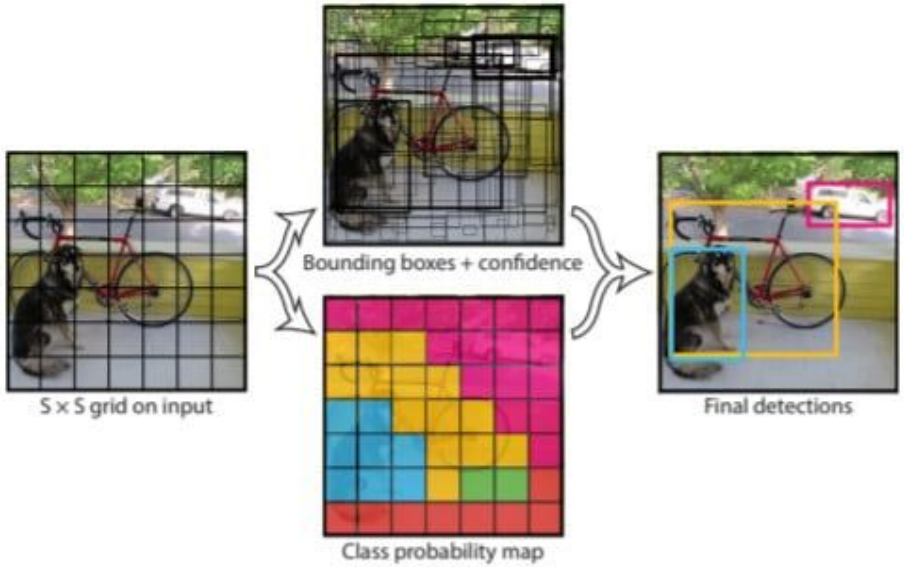


ResNet – residual connection



Object detection - YOLO

YOLO (You Only Look Once) algorithm revolutionised object detection. It is based on CNNs. Provides fast and accurate object detection



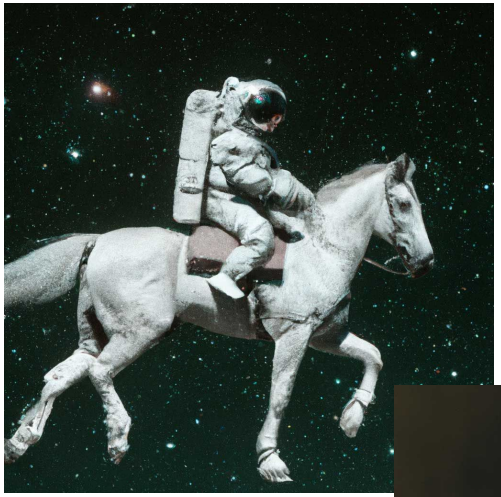
Posture recognition – DeepPose

Uses architecture based on AlexNet, but with different training target. The problem of classification is changed for regression. The goal is to predict x, y coordinates of body joints.

Image and video generation

Field with most recent advances. Very popular text-to-image models. New text-to-video models.

- DALL-E 2
- Stable Diffusion
- Imagen
- Make-a-video



Transfer learning

Using parameters from already trained models to train a network on a small dataset. Very useful approach that enables creation of good quality models.

https://keras.io/api/applications/

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
Xception	88	79.0%	94.5%	22.9M	81	109.4	8.1
VGG16	528	71.3%	90.1%	138.4M	16	69.5	4.2
VGG19	549	71.3%	90.0%	143.7M	19	84.8	4.4
ResNet50	98	74.9%	92.1%	25.6M	107	58.2	4.6
ResNet50V2	98	76.0%	93.0%	25.6M	103	45.6	4.4
ResNet101	171	76.4%	92.8%	44.7M	209	89.6	5.2
ResNet101V2	171	77.2%	93.8%	44.7M	205	72.7	5.4

Hands-on

Computer Vision

All hands-on materials available at
github.com/Gradient-PG/gradient-live-session

Questions
&
Discussion

Thank you!
See you next week on Recurrent Neural Networks.