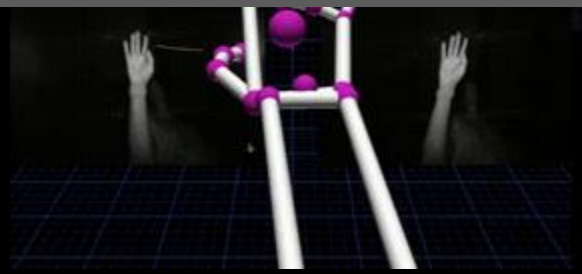
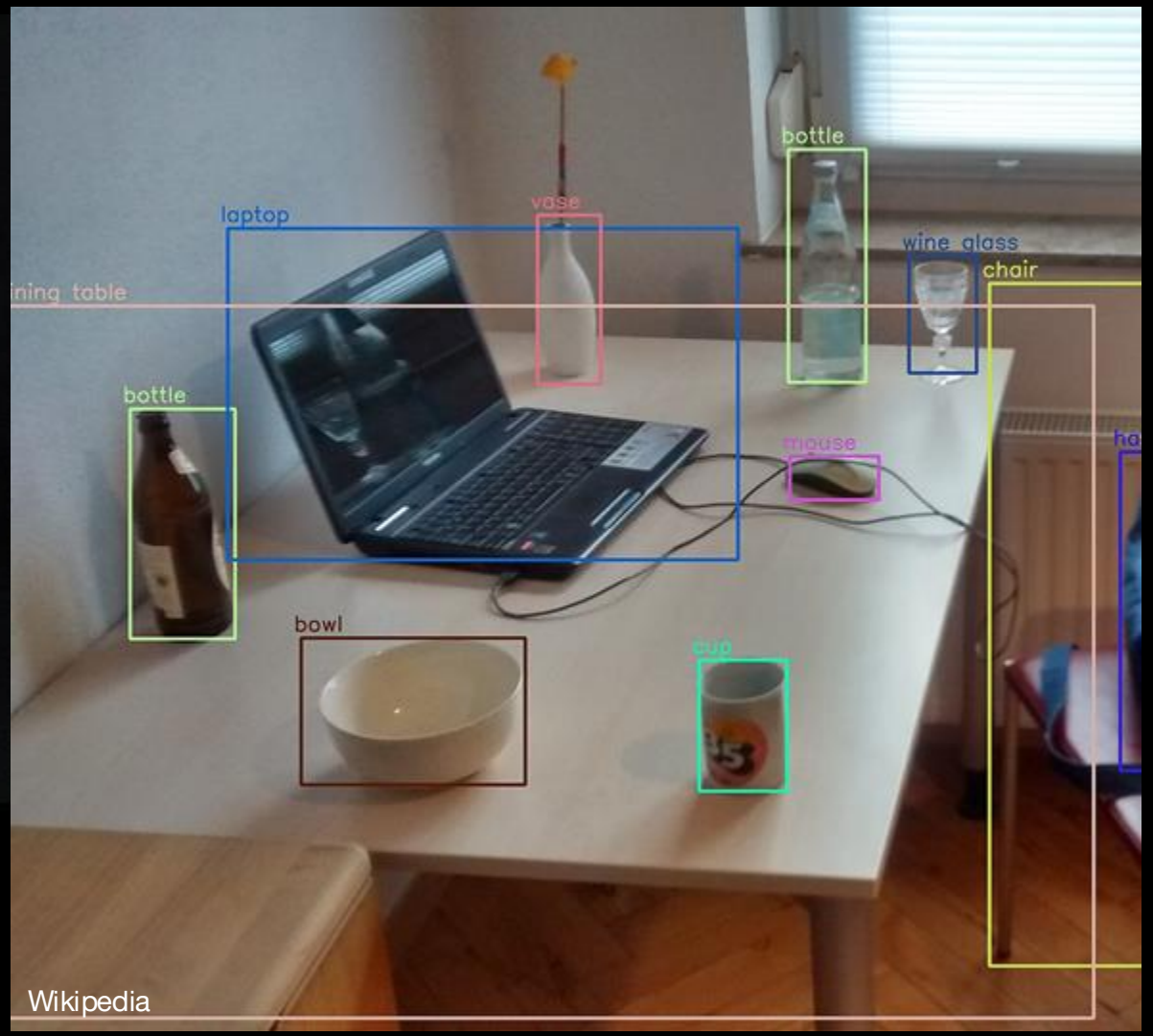
ETH zürich

Teleoperation and Object Detection

Robert Katzschmann

Assistant Professor of Robotics, Soft Robotics Lab

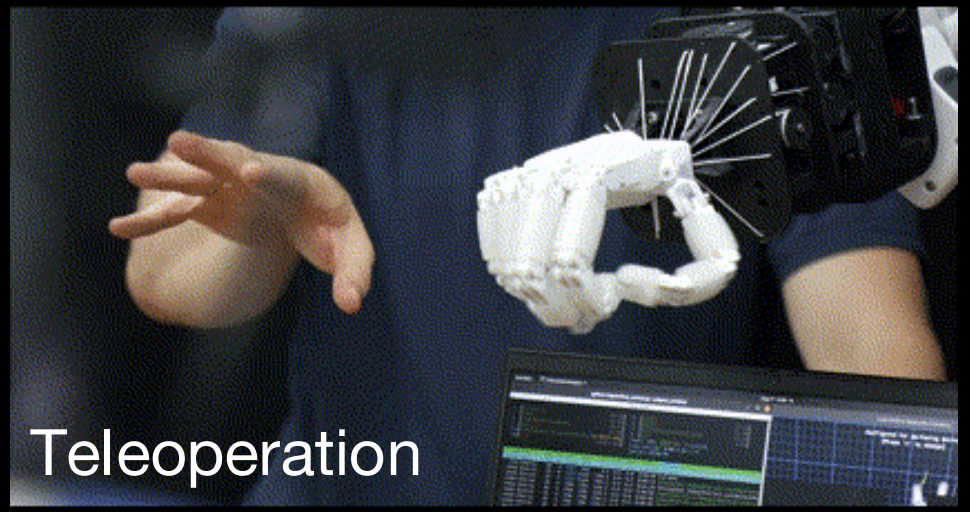




Wikipedia

SoftRobotics Laboratory

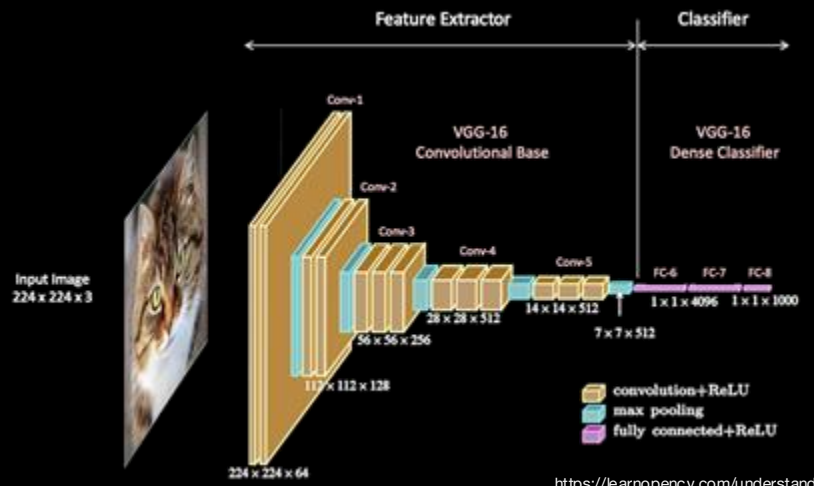
ETH zürich



Teleoperation

Wikipedia

Object detection



https://learnopencv.com/understanding-convolutional-neural-networks-cnn/

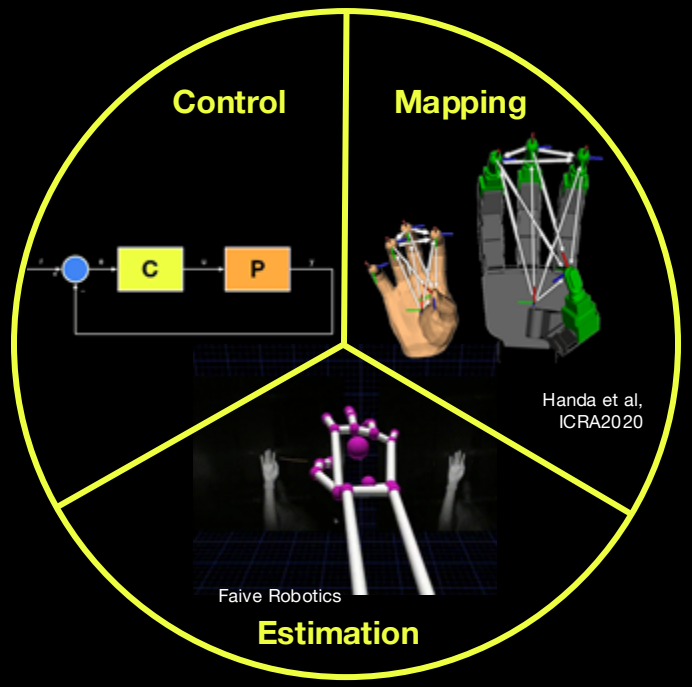
Deep Learning

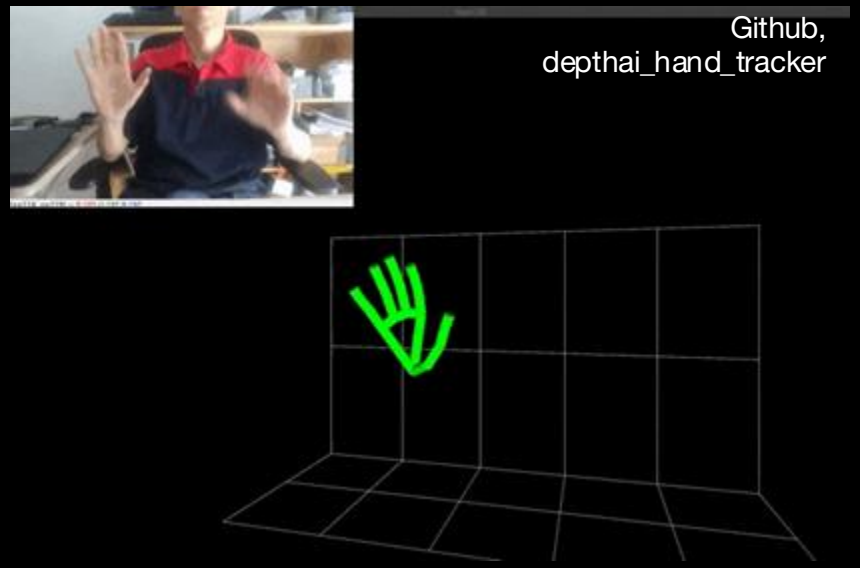
Stereo Vision

Overview

SoftRobotics Laboratory

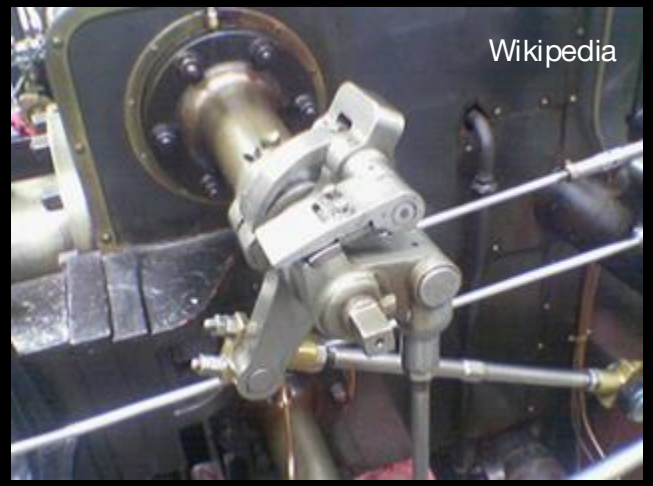
Teleoperation task: Mapping hands

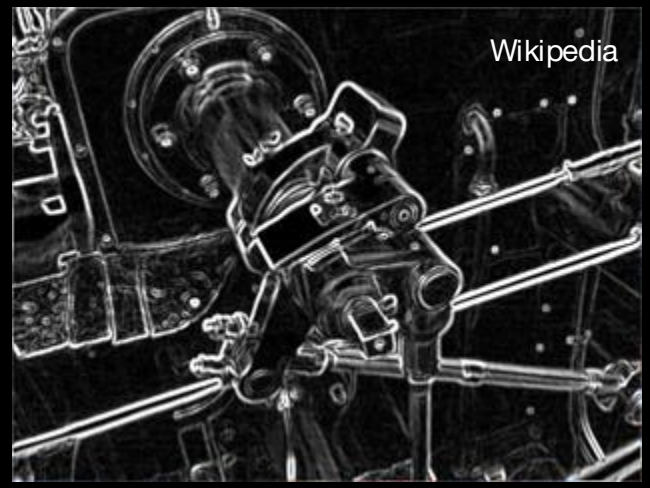




Object Detection using Filters

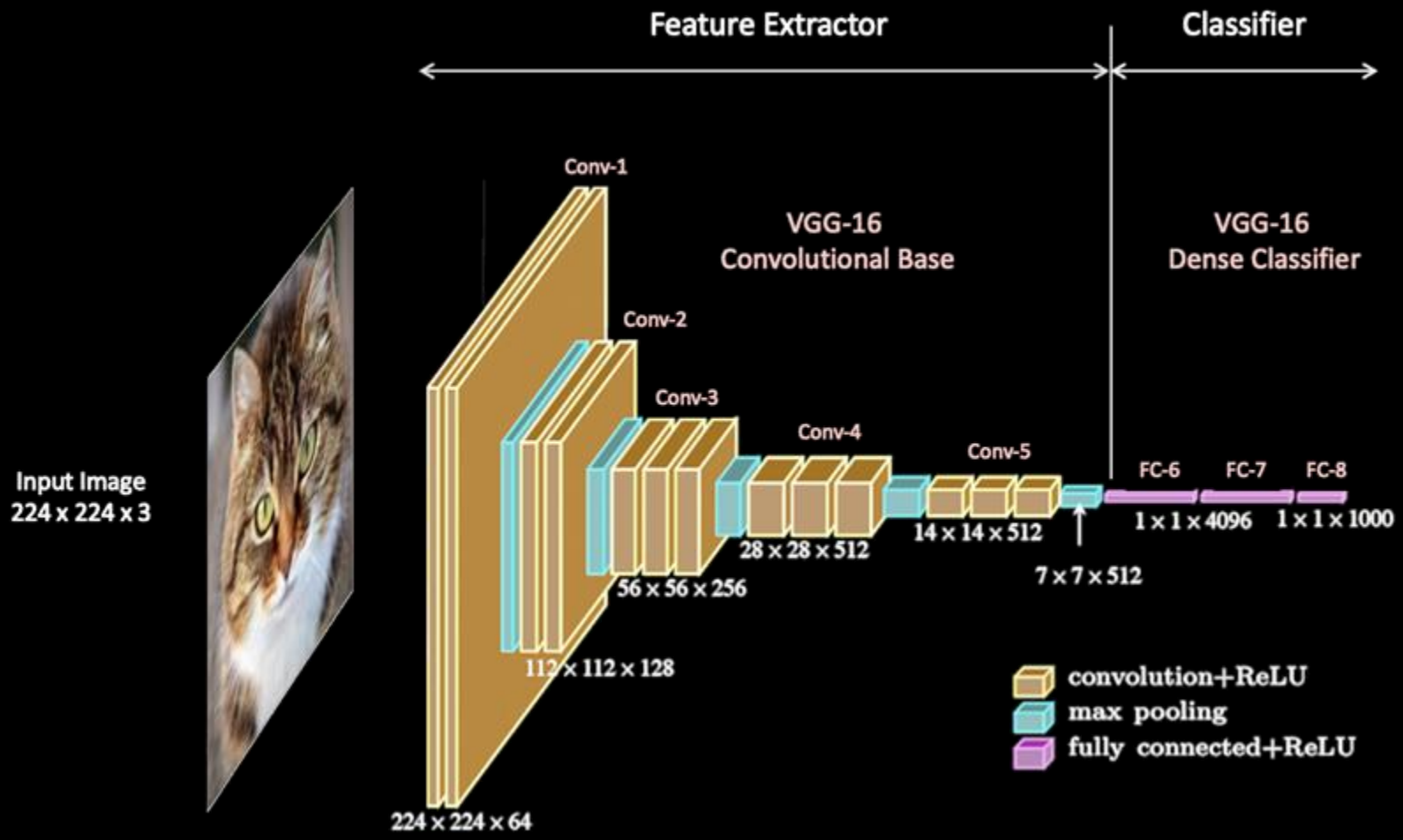
High pass filter





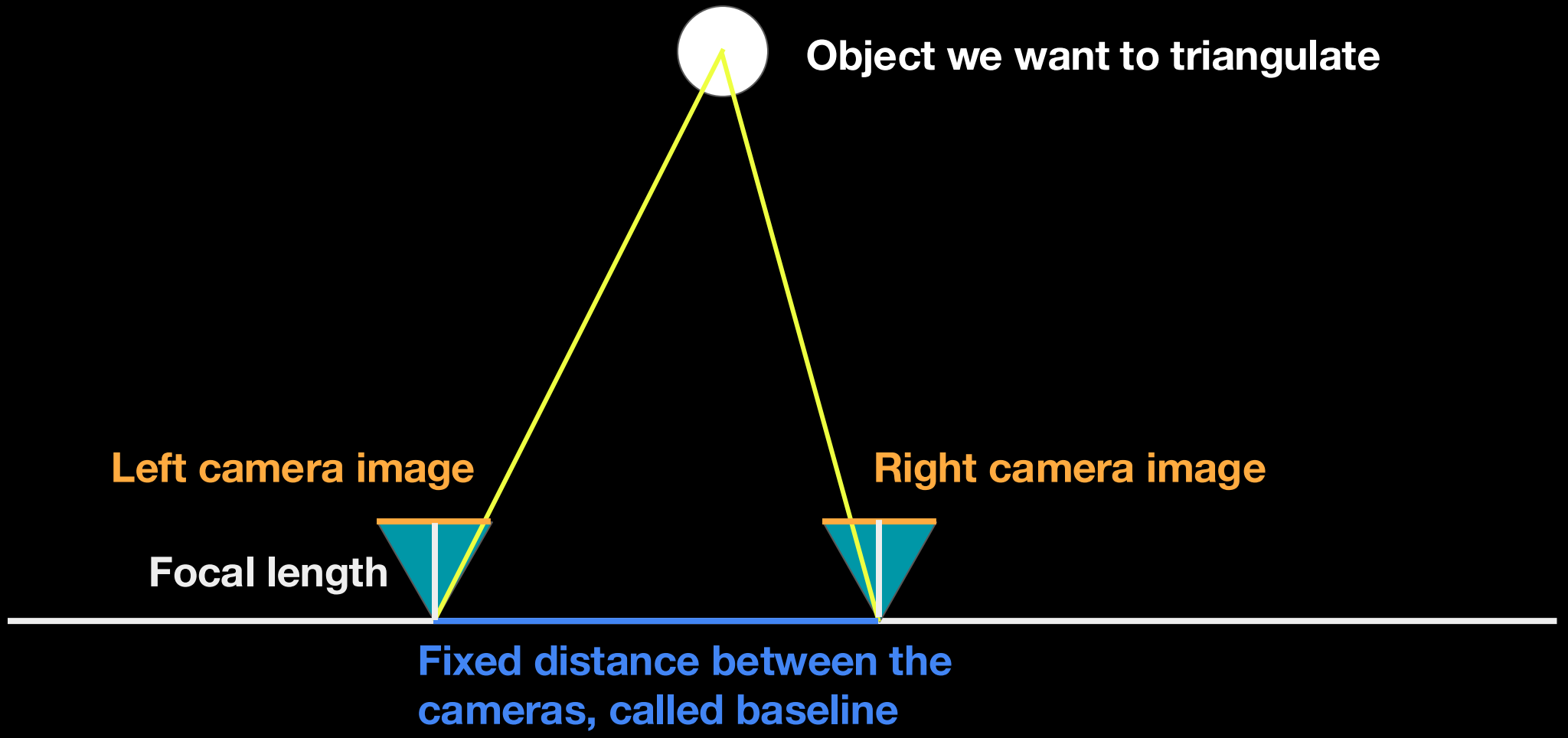
Low pass filter

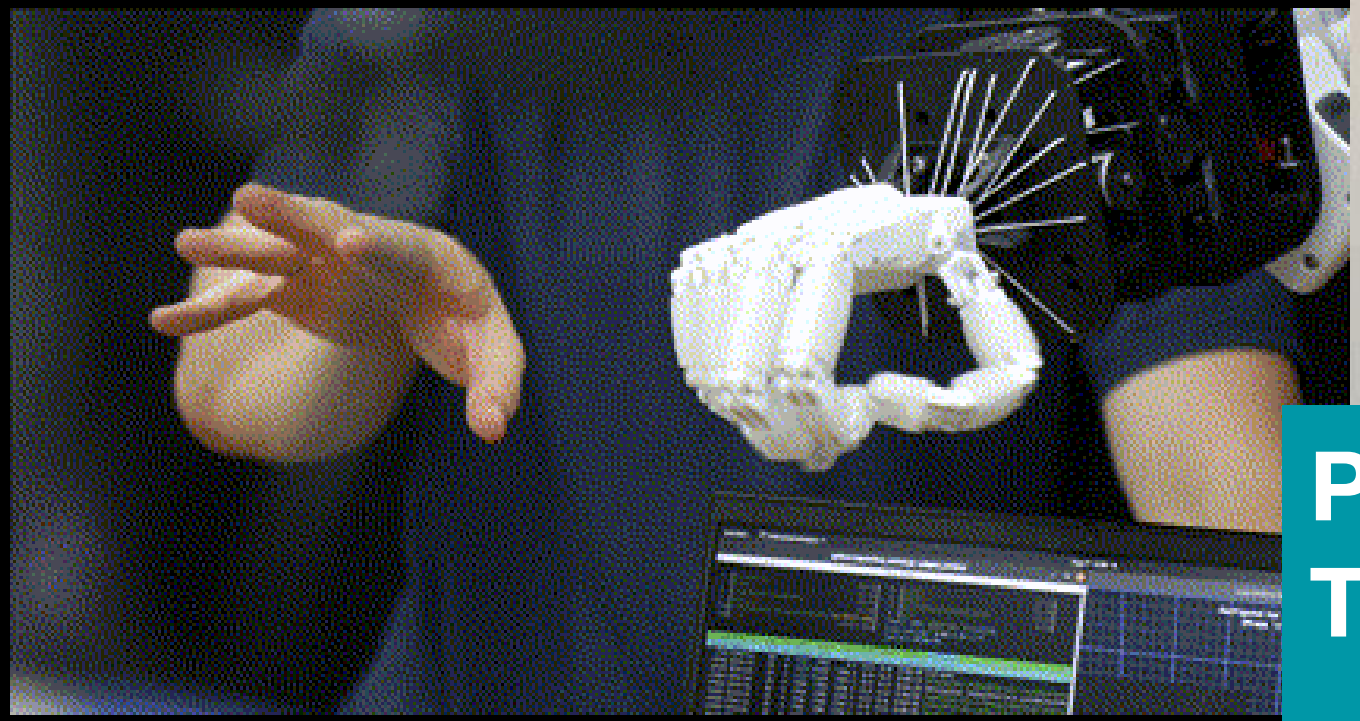
Convolutional Neural Networks



https://learnopencv.com/understanding-convolutional-neural-networks-cnn/

Stereo Vision

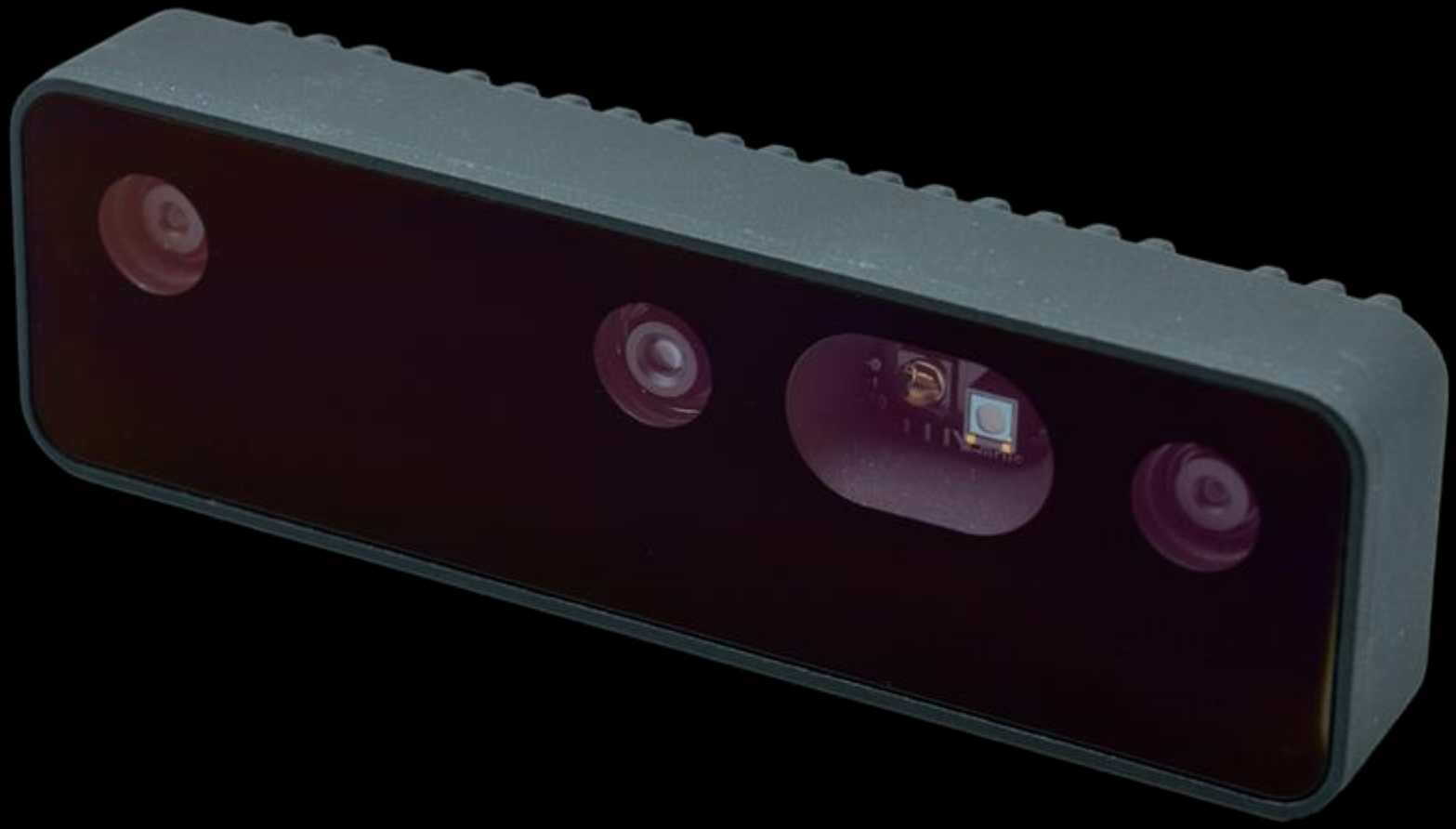




Part 1: Teleoperation

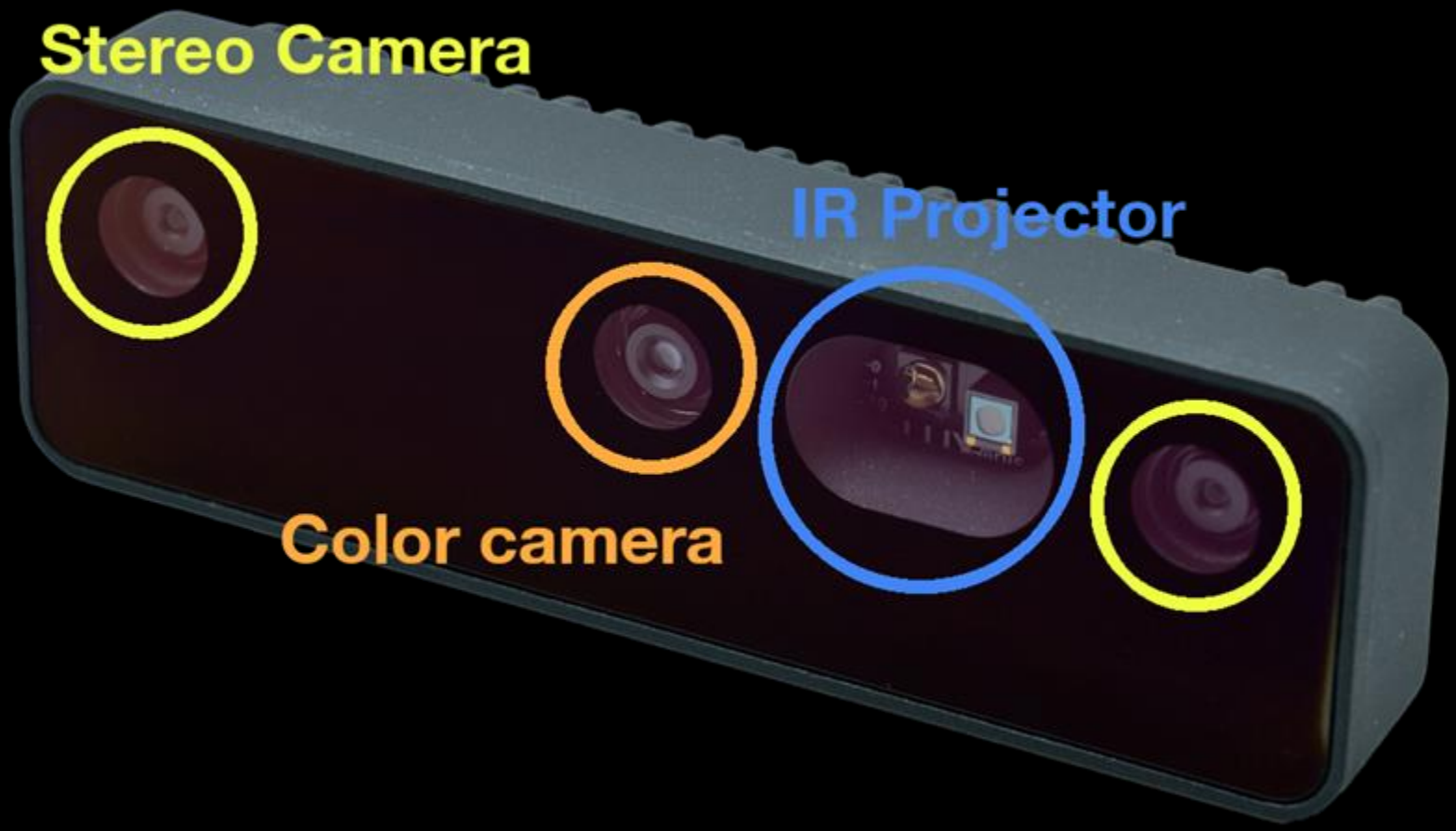
Stereo camera setup

Stereo camera setup



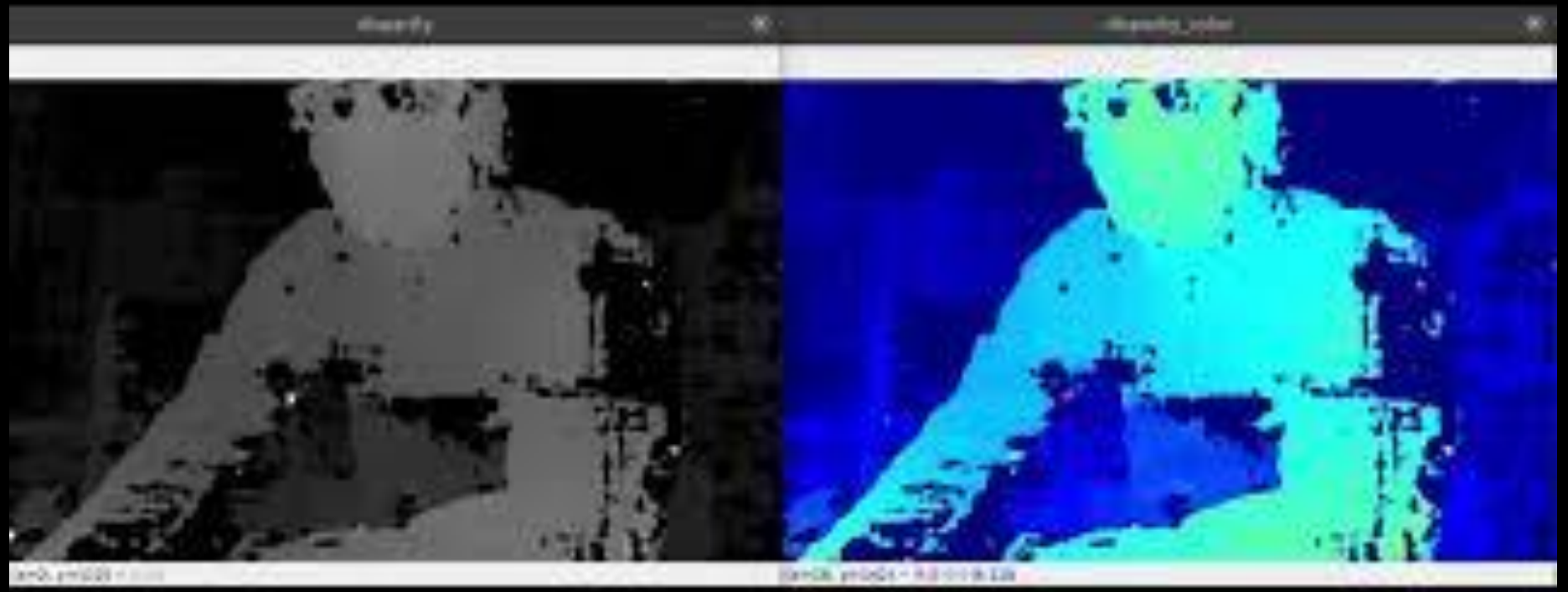
Luxonis

Stereo camera setup



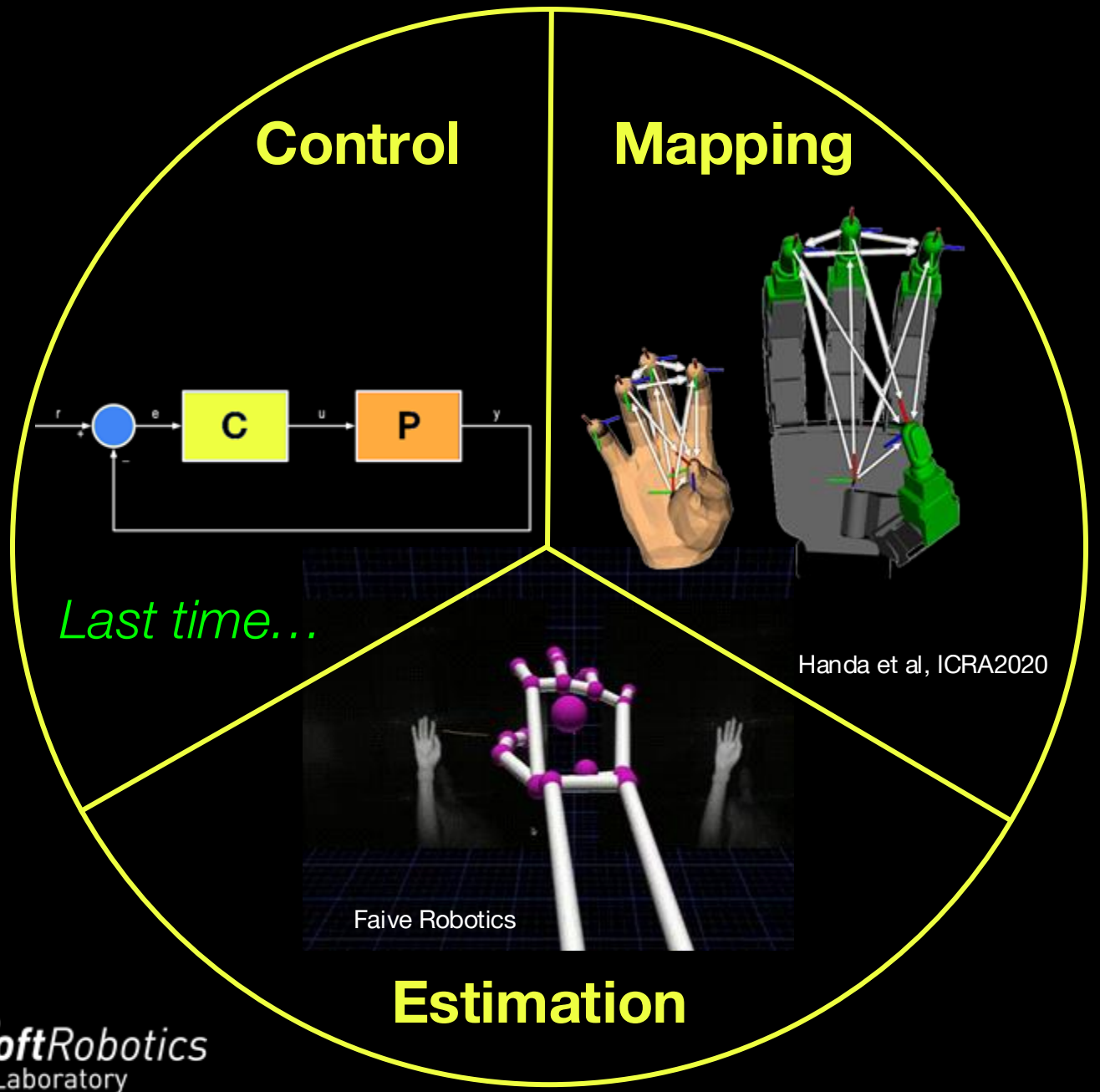
Luxonis

Stereo camera setup



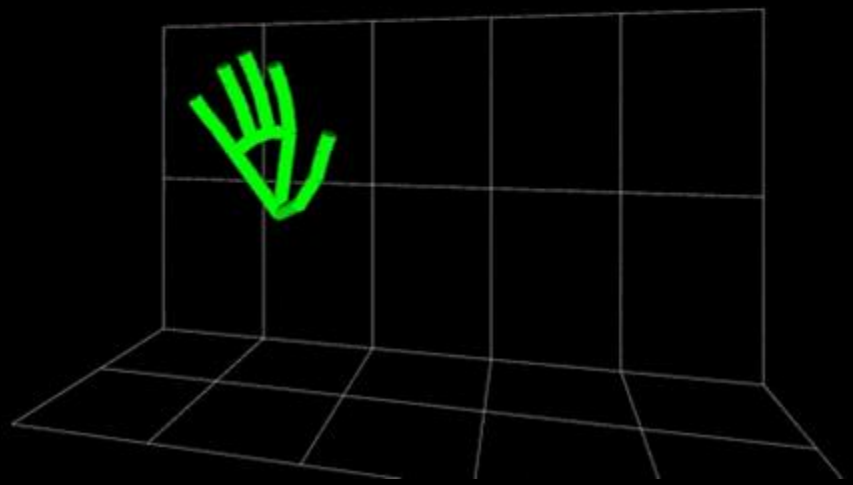
Luxonis

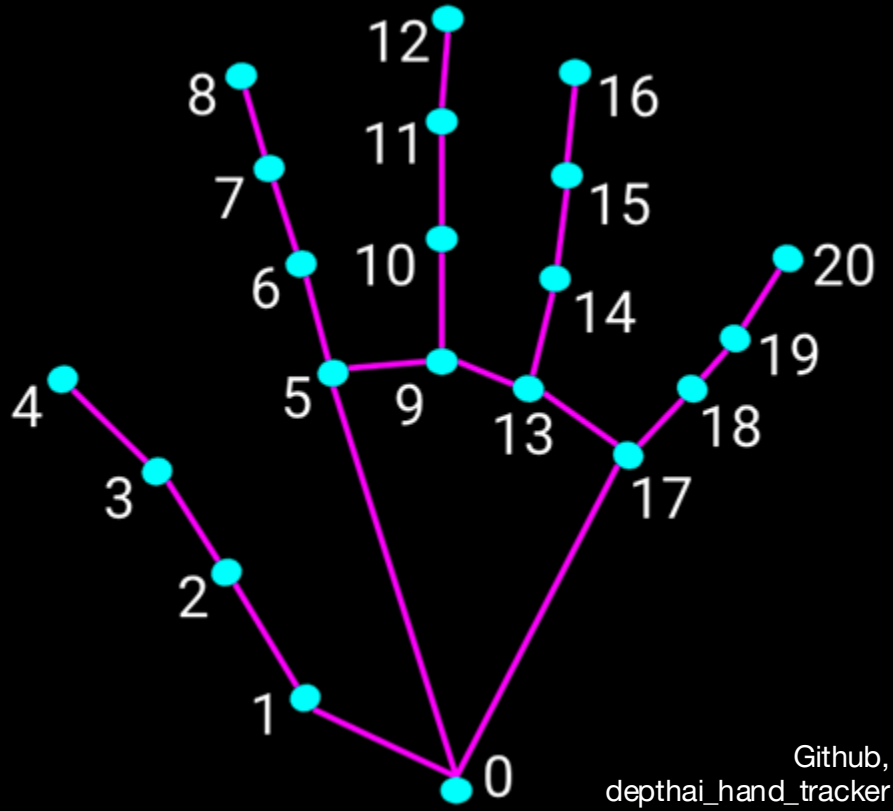
Teleoperation



Teleoperation: Sensing

Github,
depthai_hand_tracker



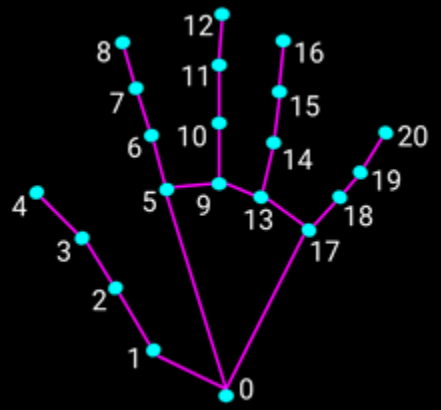


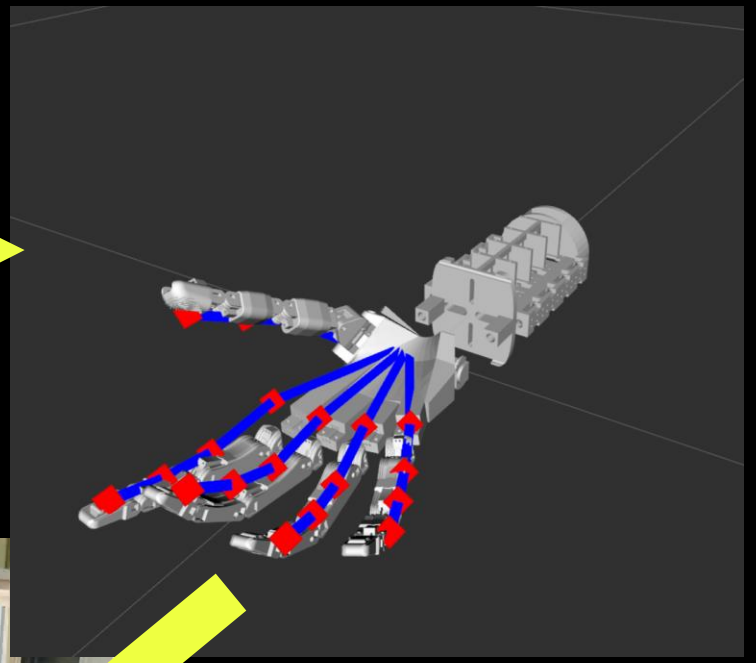
Github,
depthai_hand_tracker

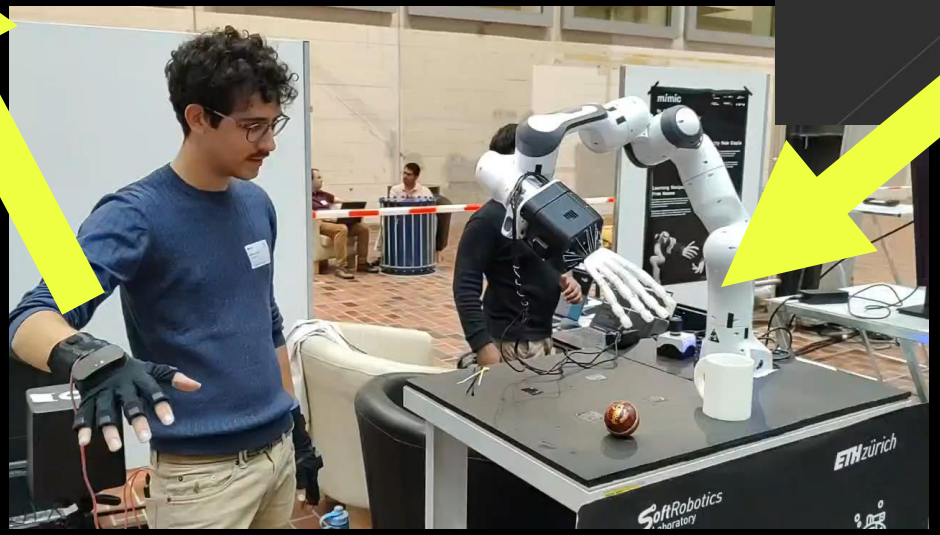
ETH zürich

Soft Robotics Laboratory

Teleoperation: Sensing



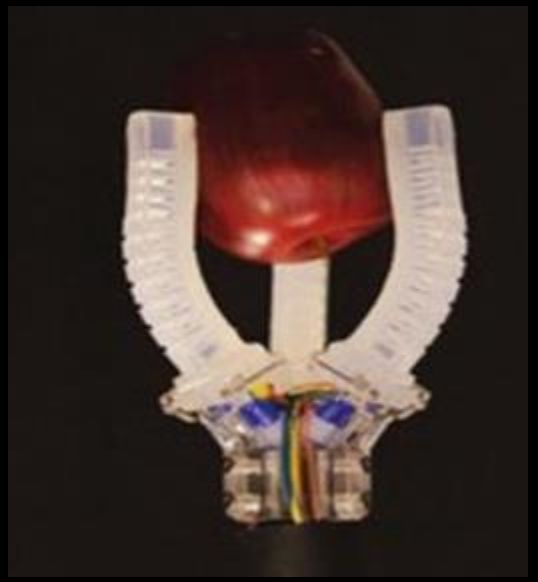




Teleoperation: Mapping

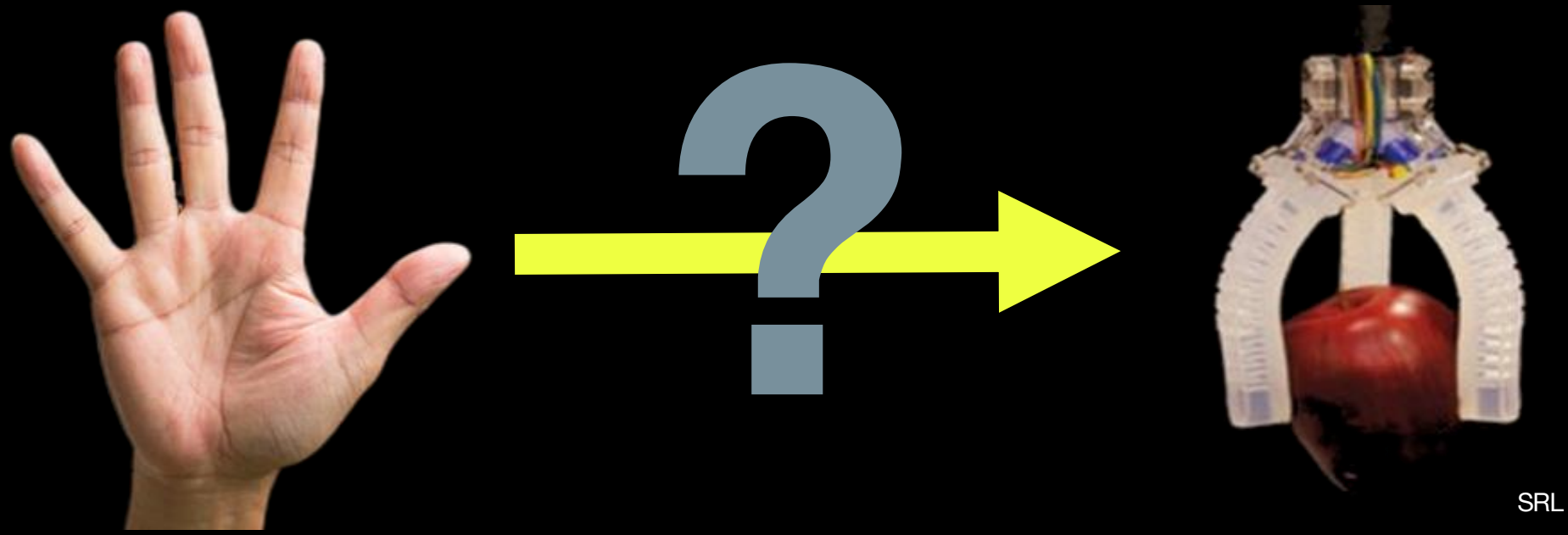
SRL

Guzman et al, JAIHC 2020



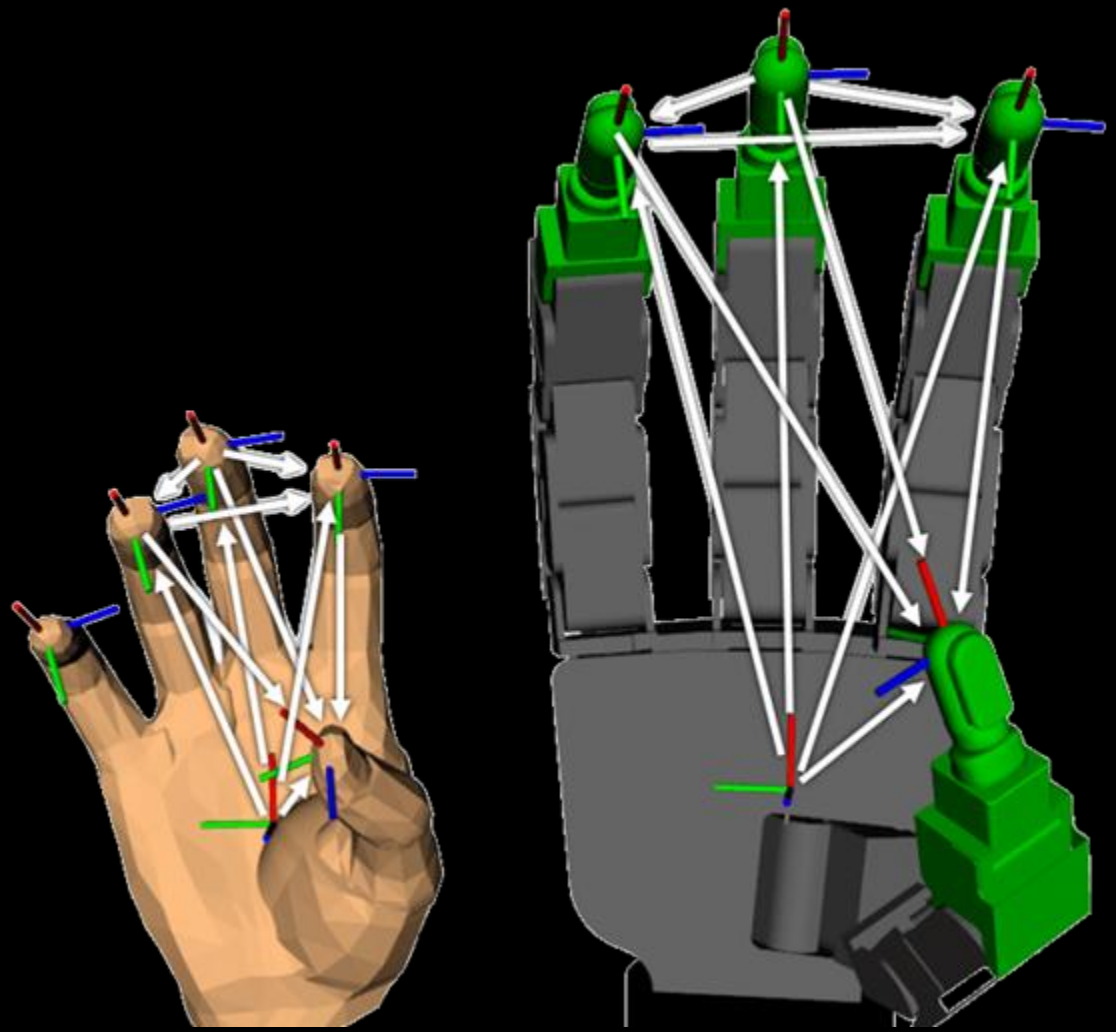
SRL

Teleoperation: Mapping



SRL

Teleoperation: Mapping



Handa et al, ICRA2020

Teleoperation: Mapping

$$E\left(\left(\beta_h, \theta_h\right), q_a\right) = \sum_{i=1}^{10} \left\| \mathbf{v_i^h} - \left(c_i \cdot \mathbf{v_i^a}\right) \right\|_2^2$$

Scaling coefficient

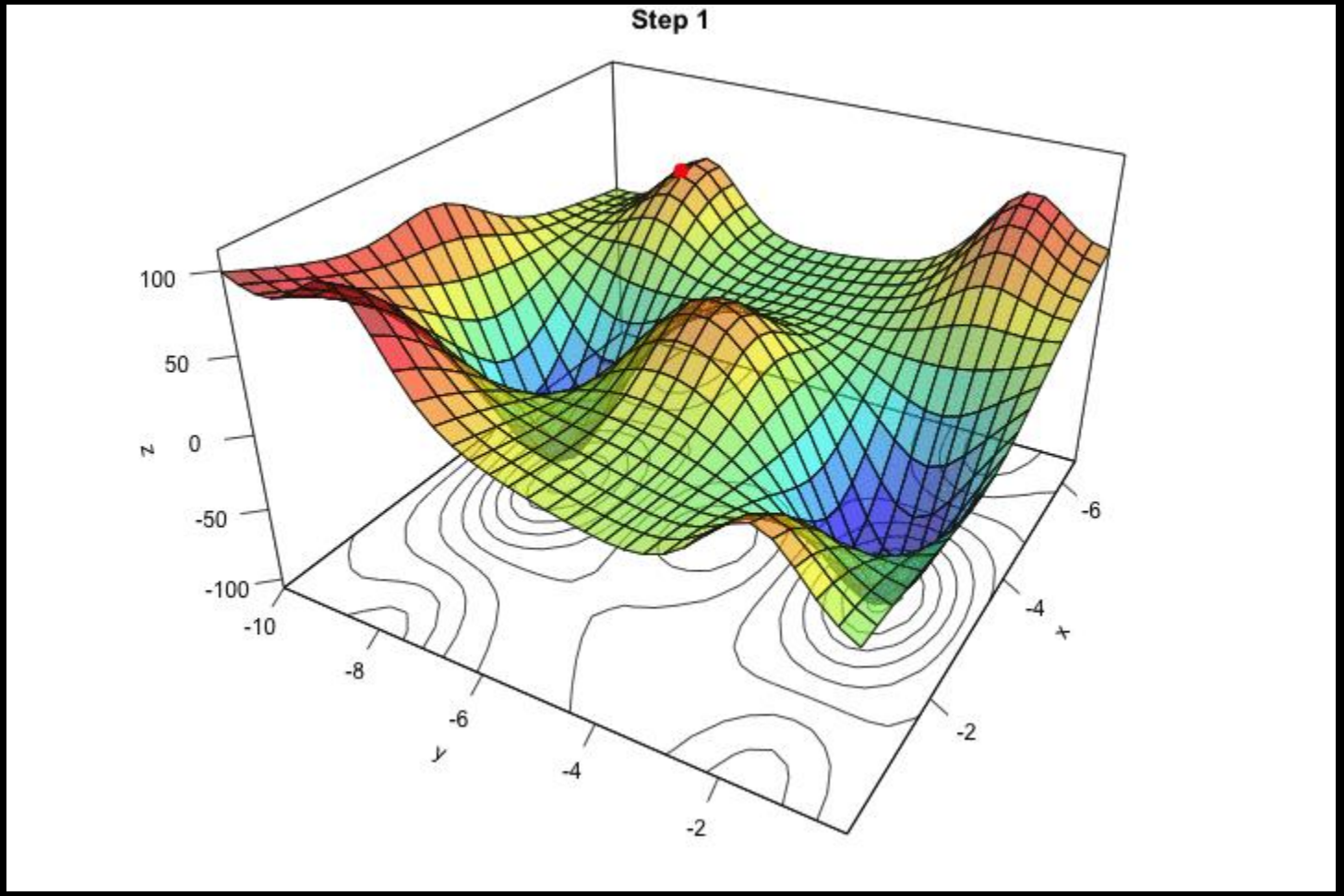
Hand pose

Robot joint angles

Hand keyvector

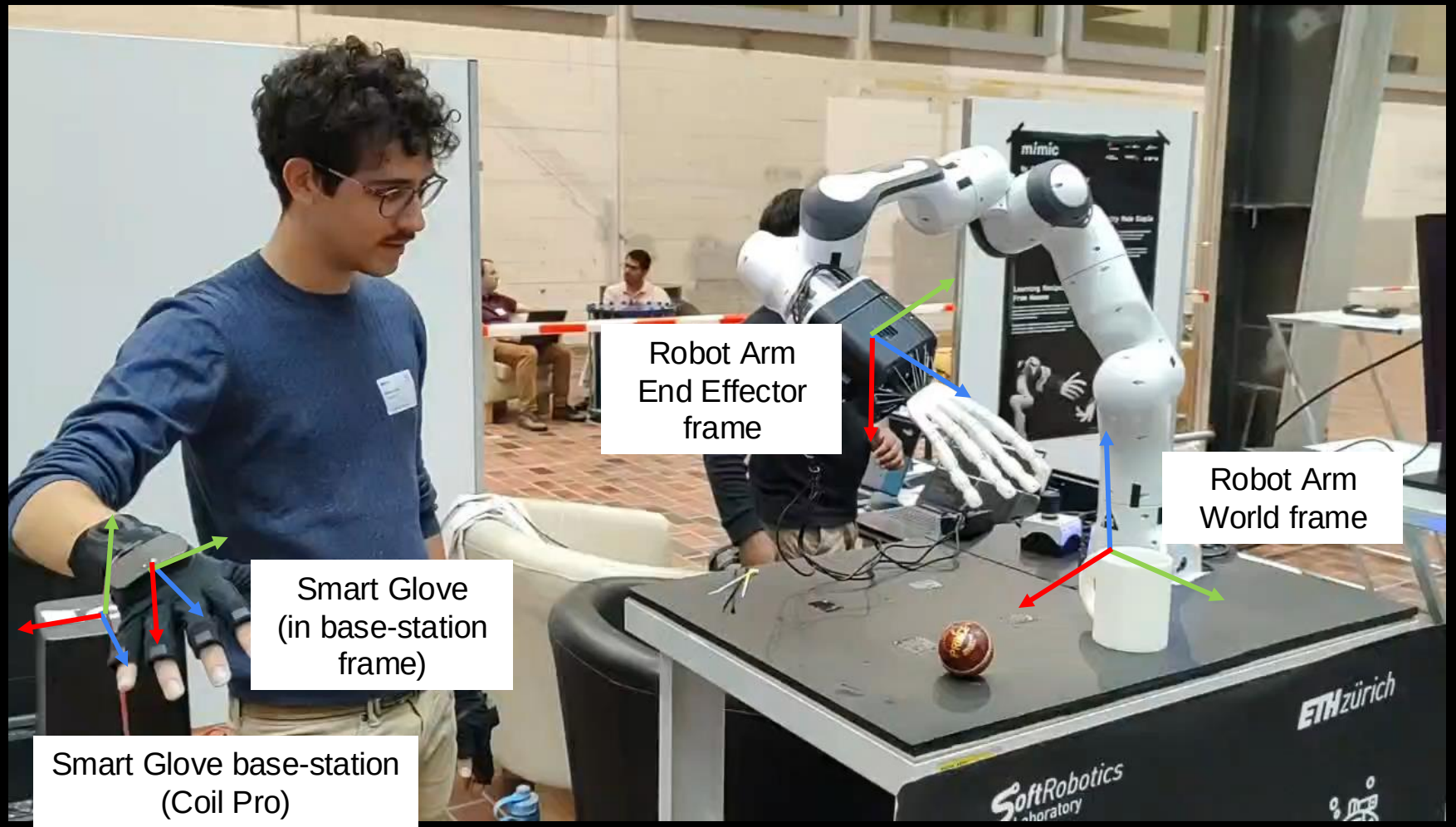
Robot keyvector

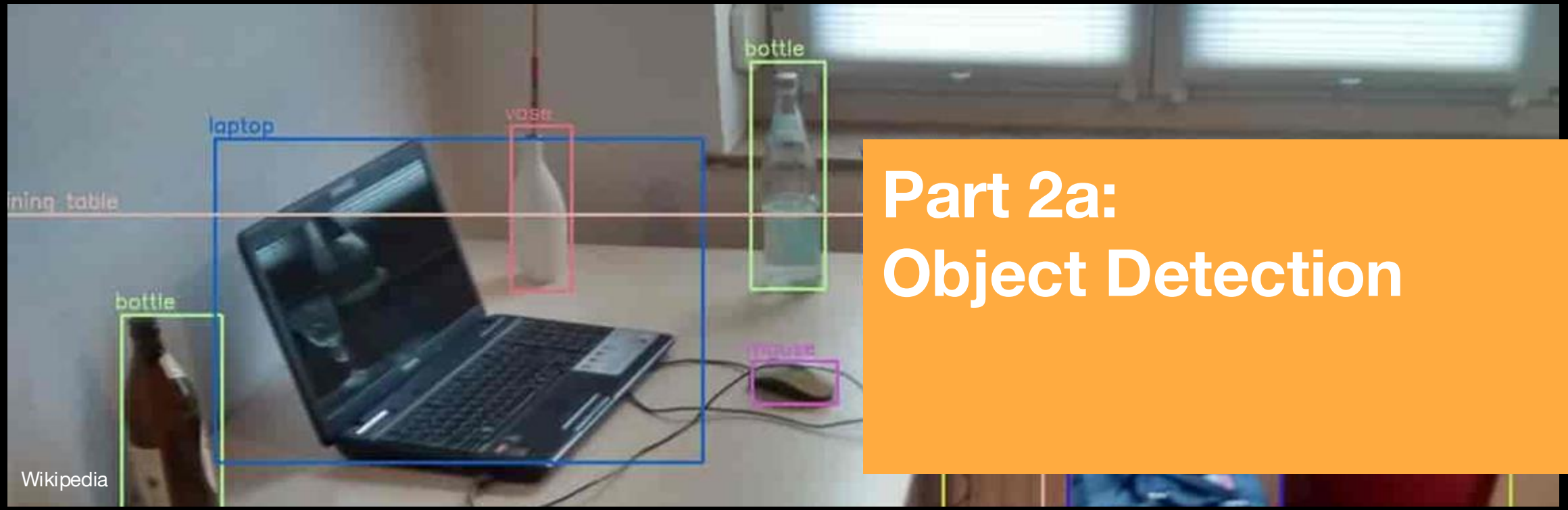
Teleoperation: Mapping



https://egallic.fr/Enseignement/ML/ECB/book/gradient-descent.html

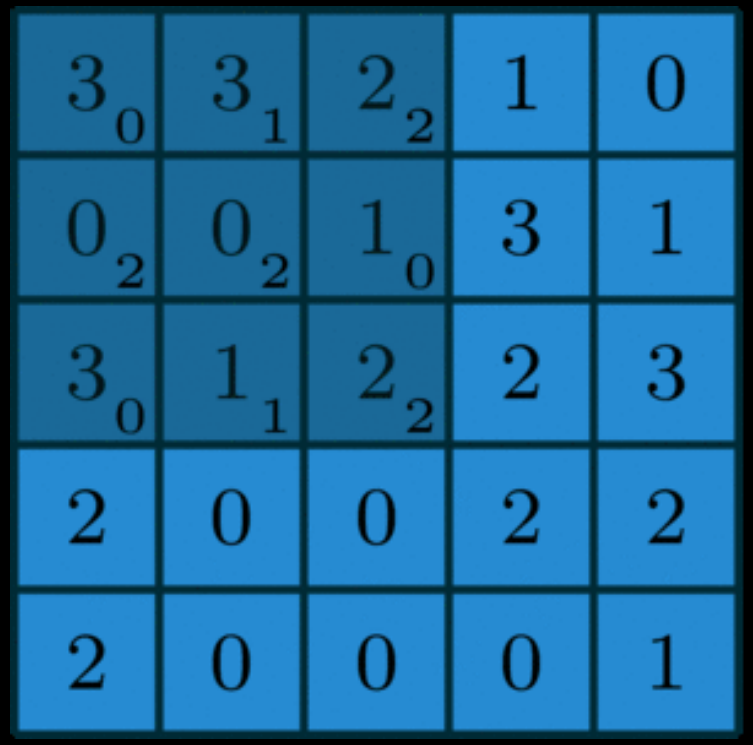
Teleoperation: Robot Arm





Part 2a: Object Detection

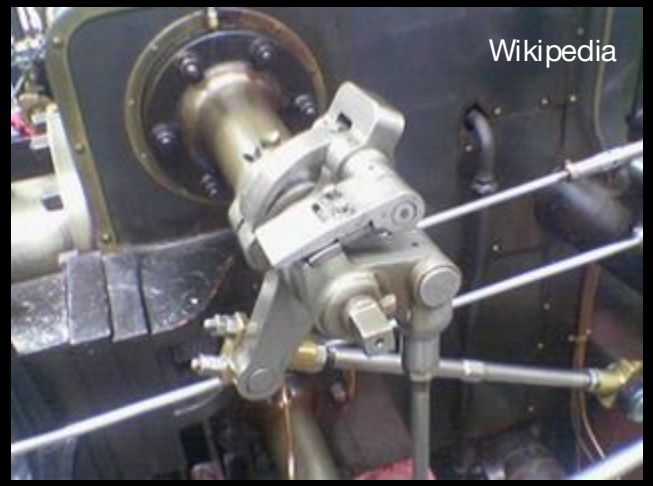
Classical Approaches: Filters

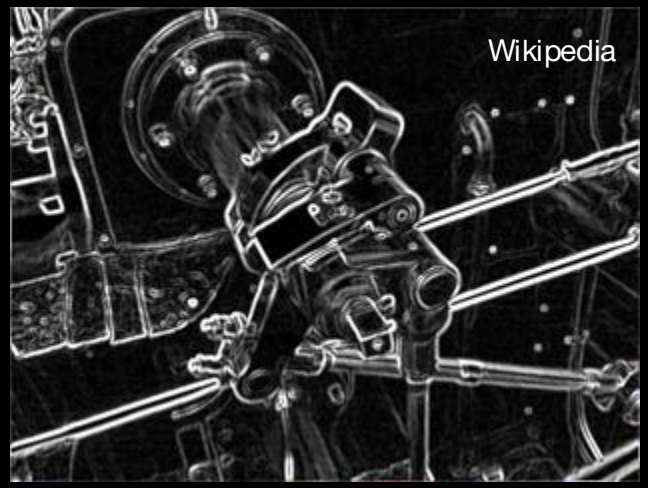


https://www.jie-tao.com/types-of-convolutionstranslation/

Classical Approaches: Filters

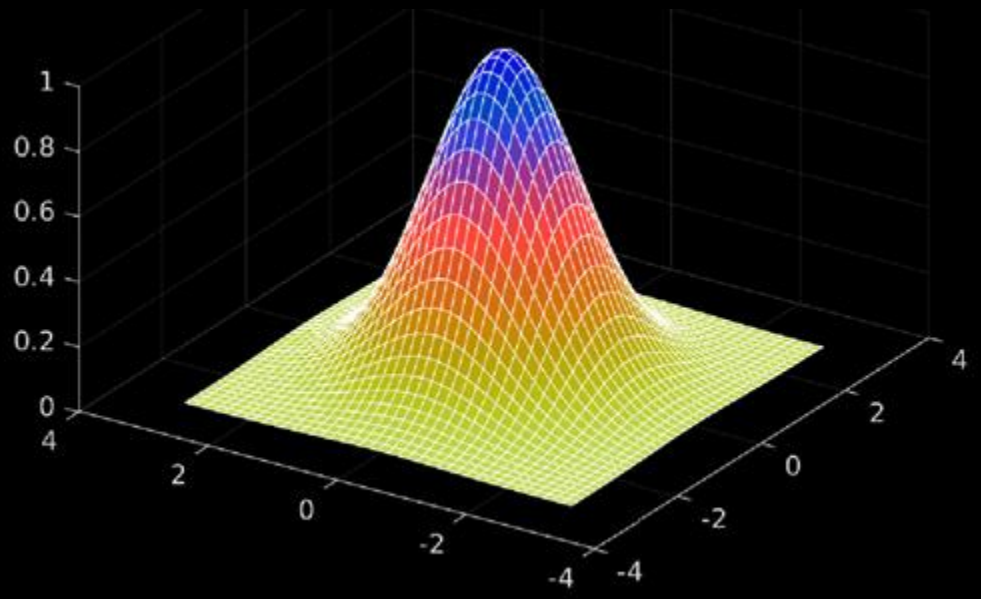
High pass filter



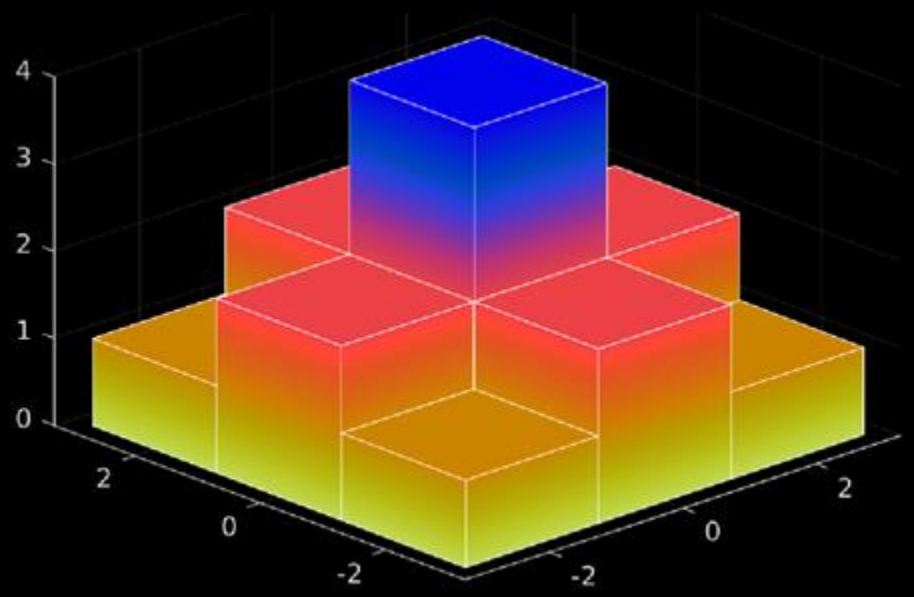


Low pass filter

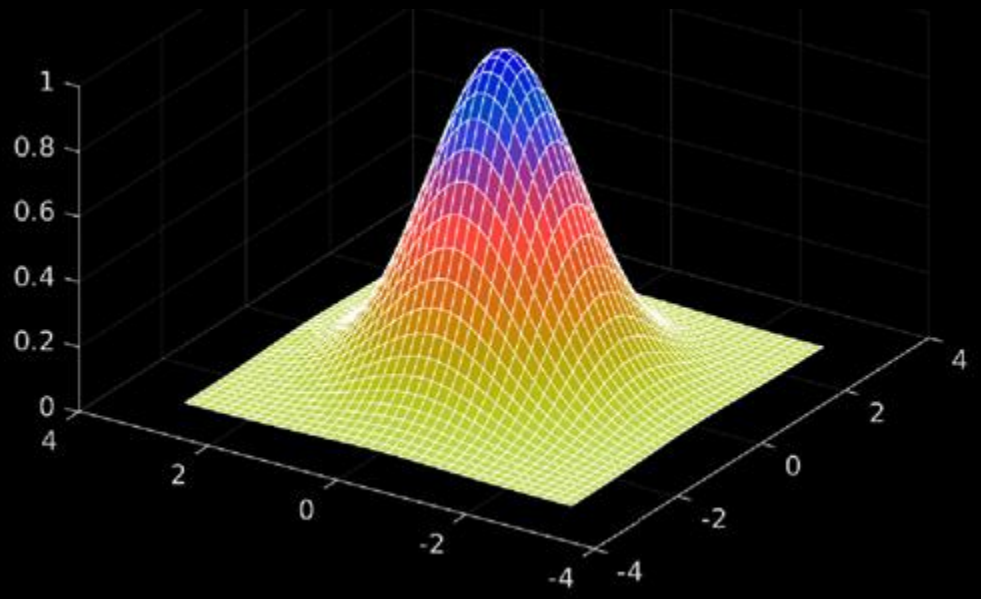
Classical Approaches: Filters



Discretized



Classical Approaches: Filters



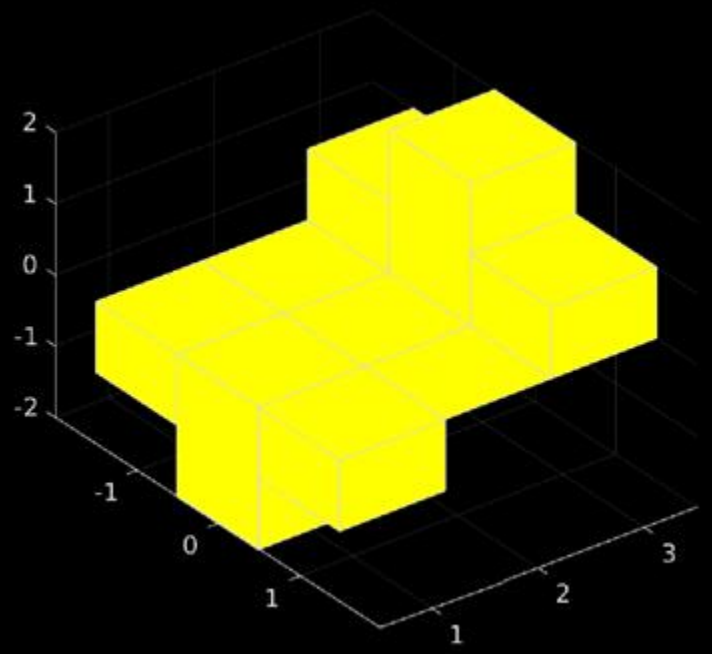
Discretized

1	2	1
2	4	2
1	2	1

Classical Approaches: Filters

No filter 3x3 Gaussian 20x20 Gaussian

Classical Approaches: Filters



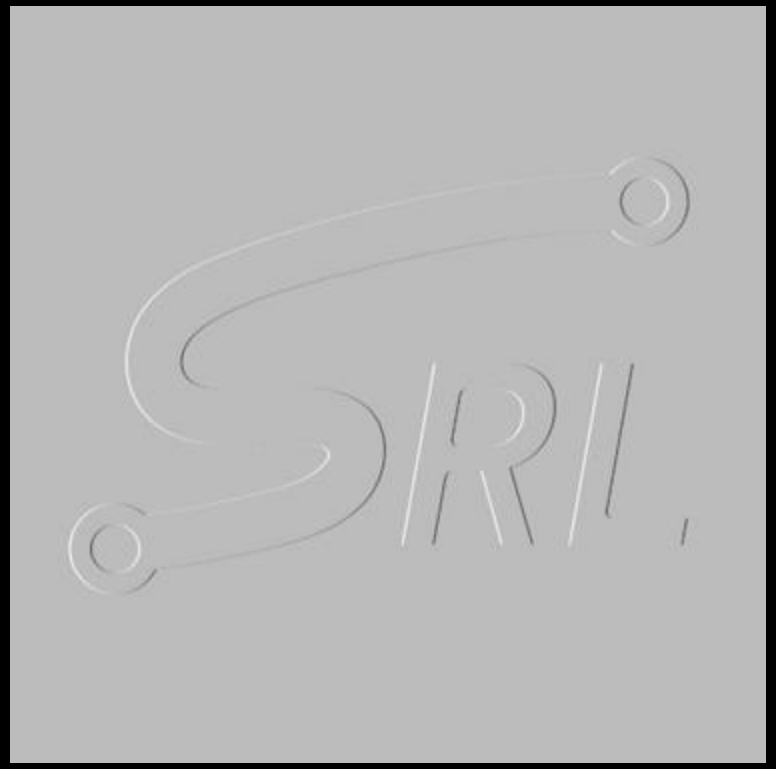
-1	0	1
-2	0	2
-1	0	1

Horizontal derivative

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

Vertical derivative

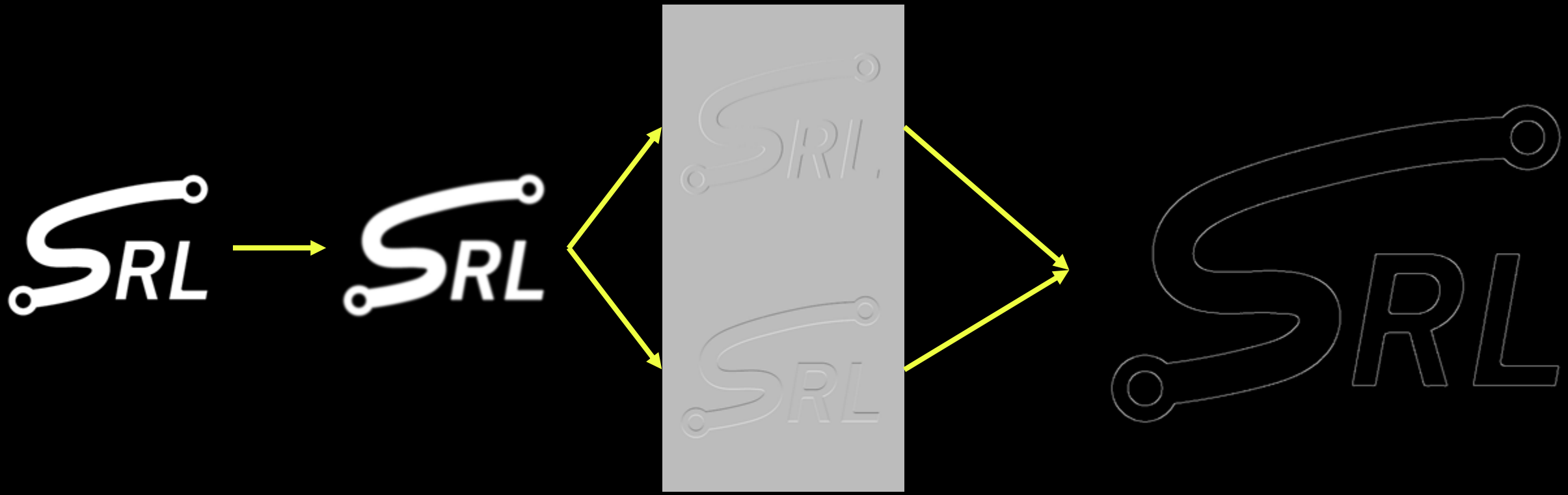
Classical Approaches: Filters



Horizontal Sobel

Vertical Sobel

Classical Approaches: Filters



Classical approaches: Advanced Methods

Features detected

Histogram of oriented gradients

Part 2b: Deep Learning

Deep Learning: Tasks

Deep Learning: Tasks

Classification

Neural network

“Apple”

Deep Learning: Tasks

Object detection

Neural network

Apple

Deep Learning: Tasks

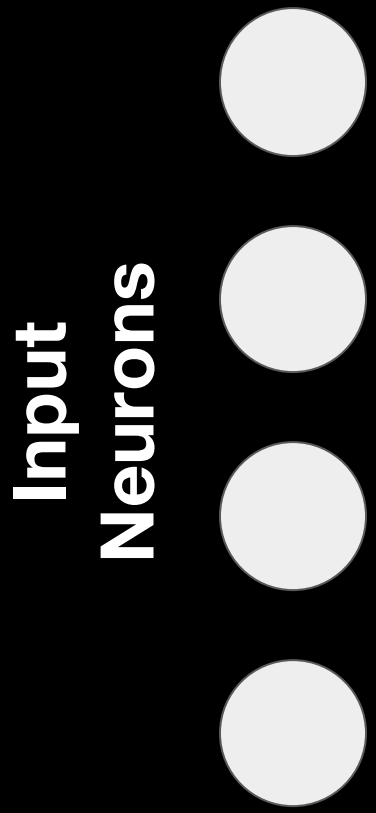
Semantic segmentation

Neural network

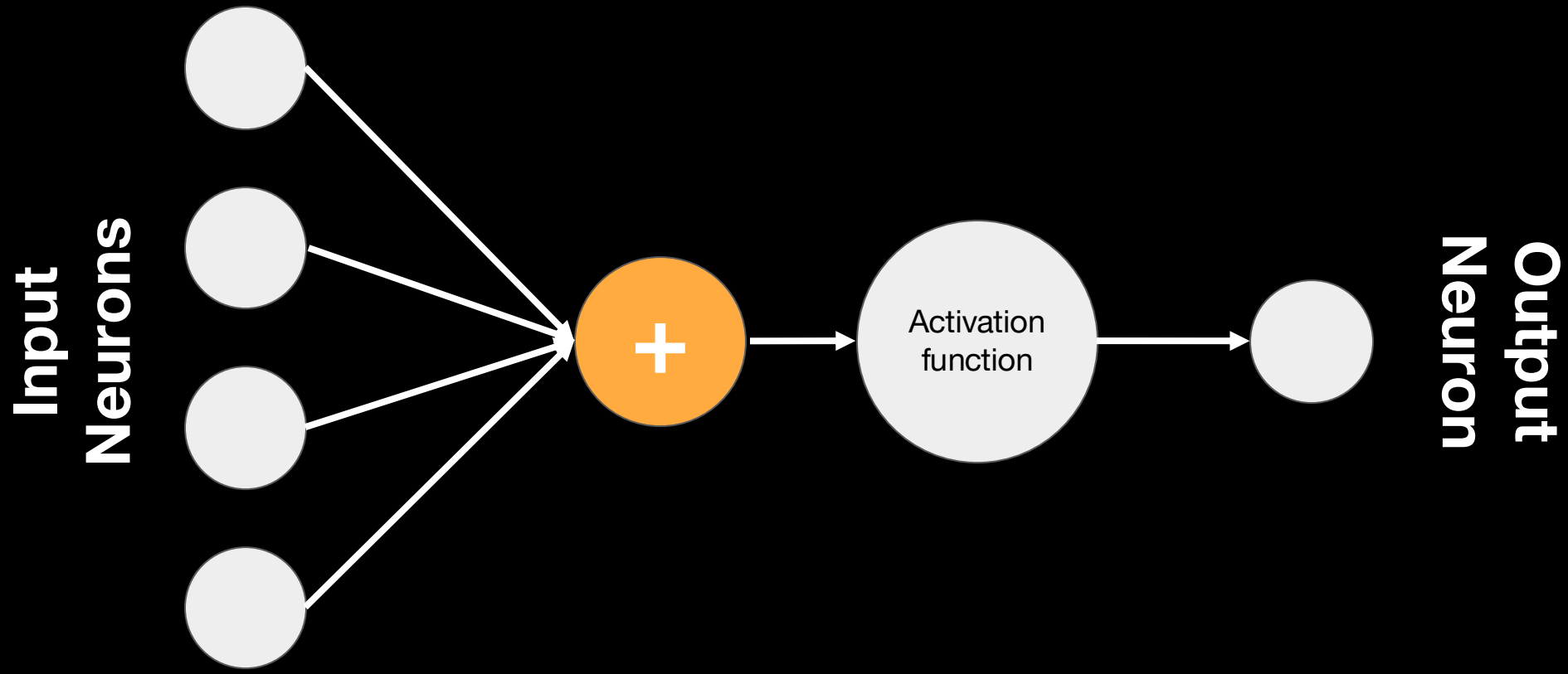
Deep Learning: Tasks

Keypoint detection

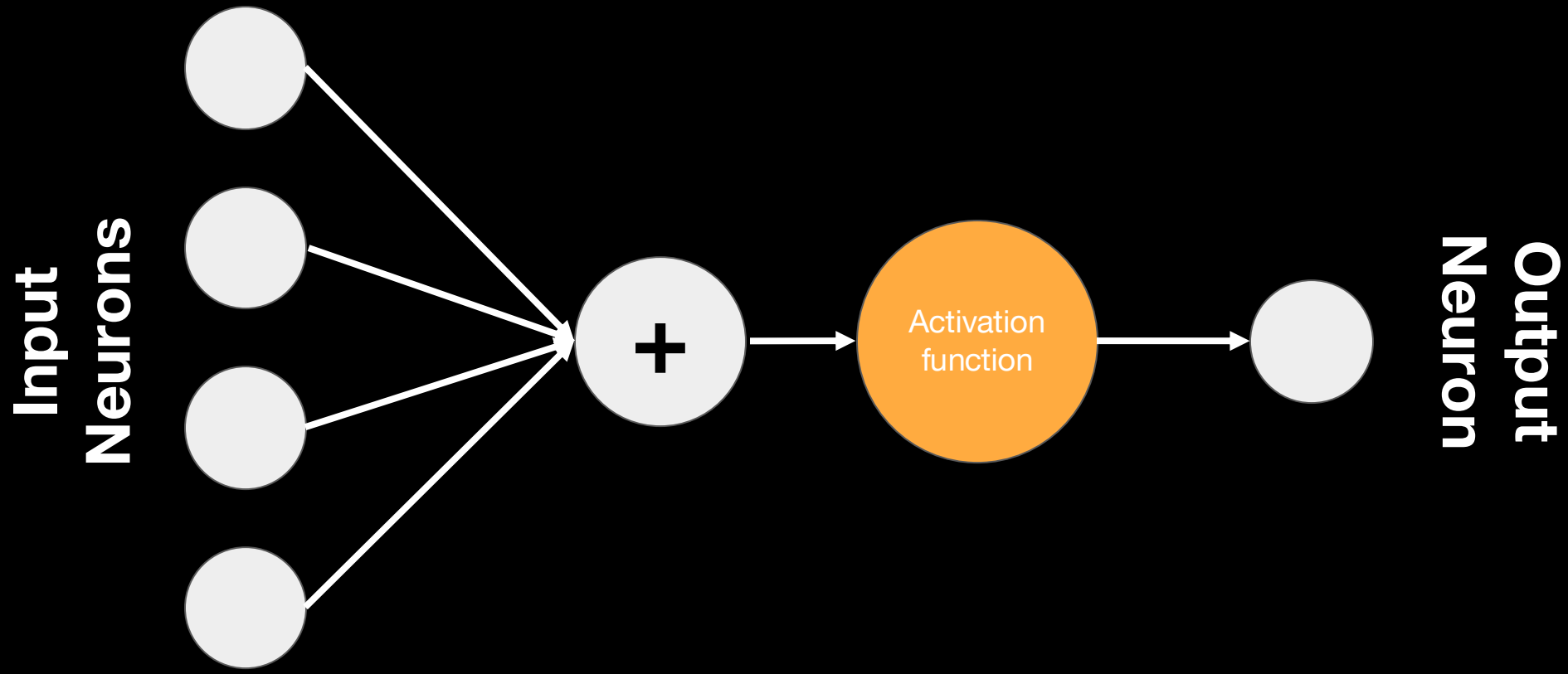
Deep Learning: Neural networks



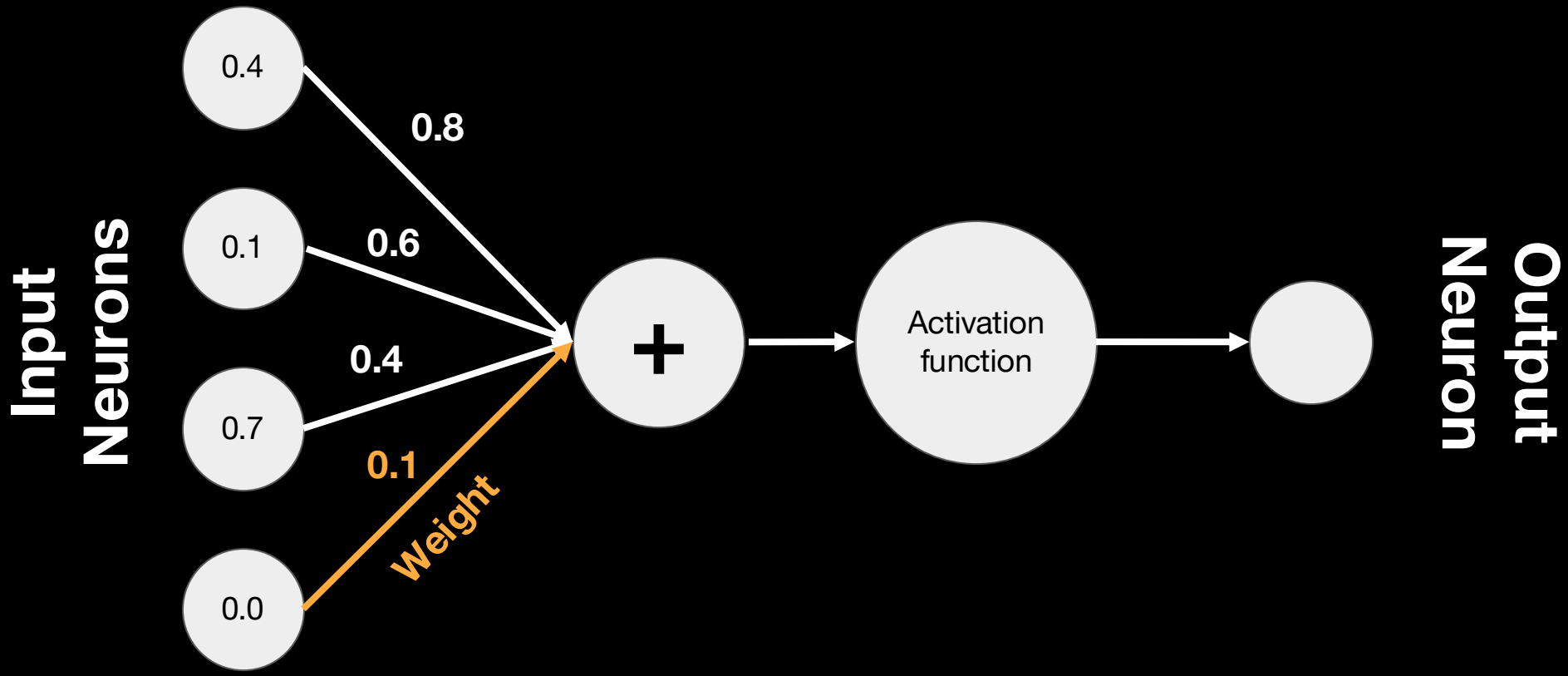
Deep Learning: Neural networks



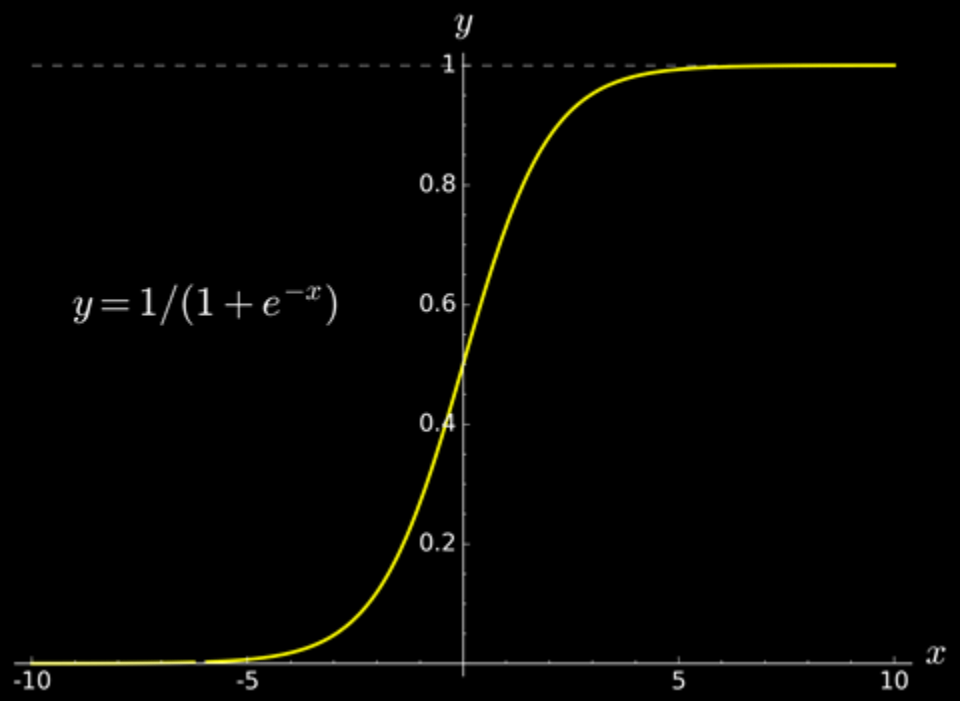
Deep Learning: Neural networks



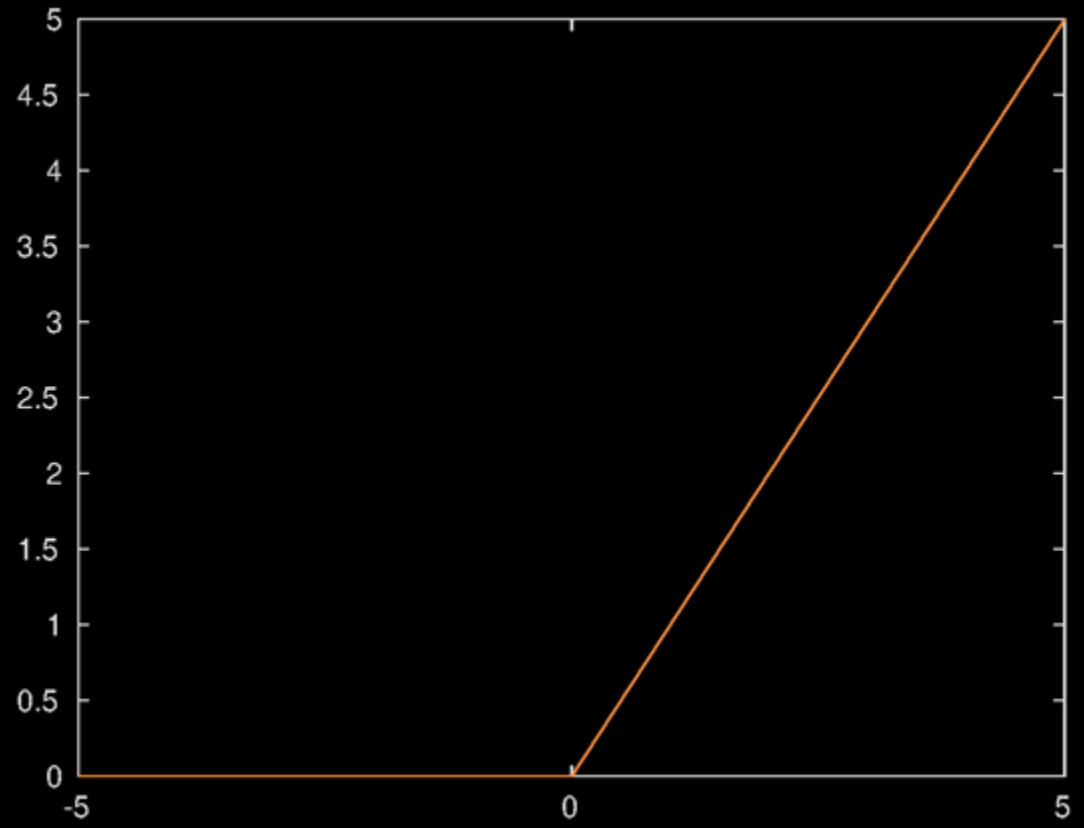
Deep Learning: Neural networks



Deep Learning: Neural networks



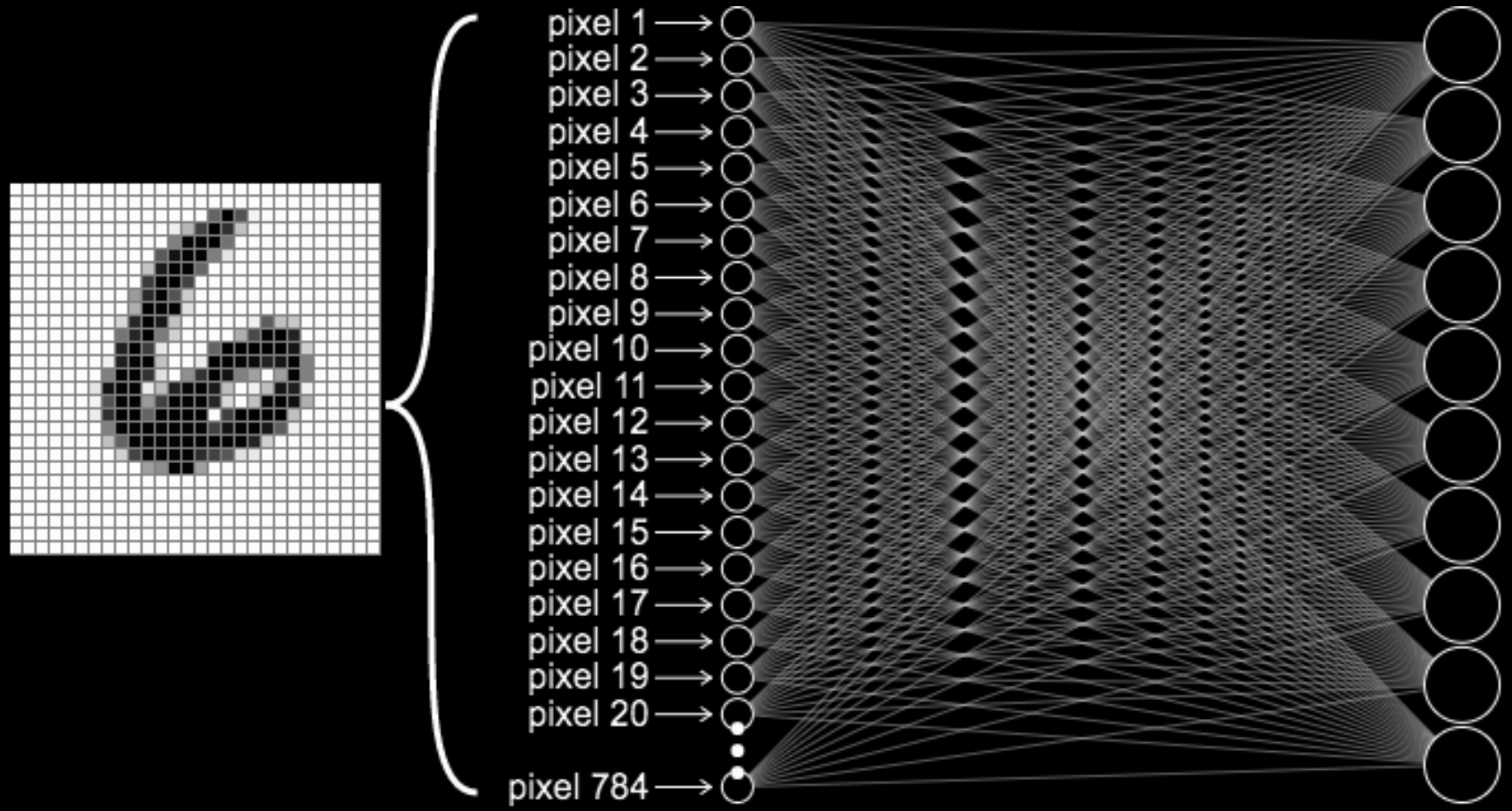
Sigmoid function



ReLu function

ETH zürich SoftRobotics Laboratory

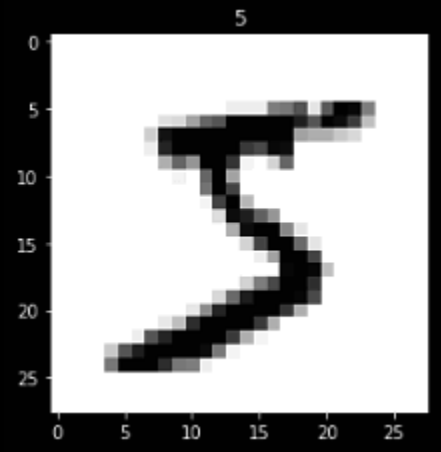
Deep Learning: Example



https://medium.com/dataman-in-ai/module-6-image-recognition-for-insurance-claim-handling-part-i-a338d16c9de0

Deep Learning: Convolutional Neural Networks

MNIST Dataset
28x28 pixels → 784 inputs



MNIST Sample

4K Image
3840x2160 pixels → 8.3 million inputs!!

ETH zürich

SoftRobotics Laboratory

Deep Learning: Convolutional Neural Networks

0	0	0	0	0	0	0	0
0	255	255	255	0	255	255	0
0	0	255	0	0	255	255	0
0	255	255	255	0	0	255	0
0	0	255	0	0	255	0	0
0	0	255	0	0	255	0	0
0	0	255	0	0	255	0	0
0	0	0	0	0	0	0	0

*

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

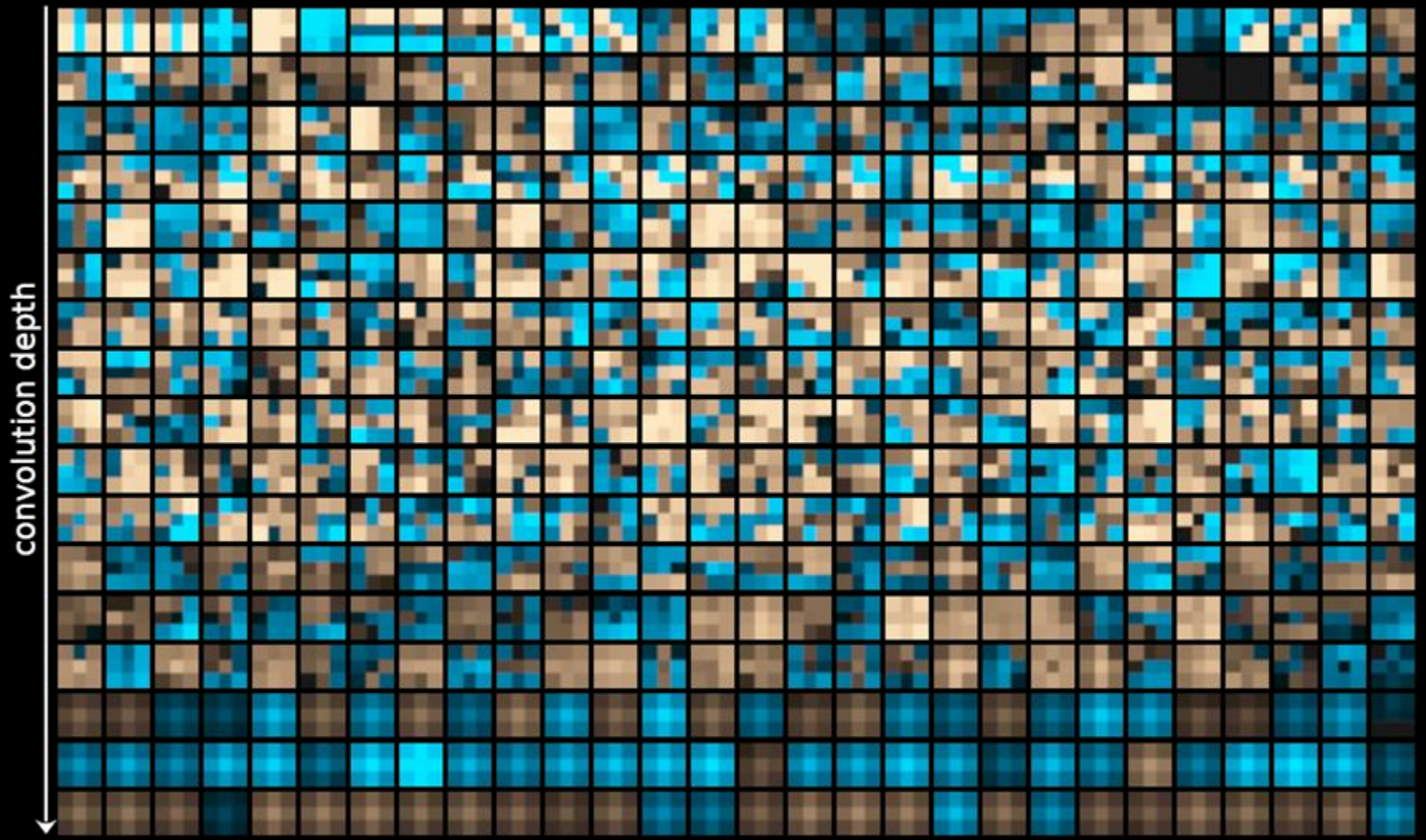
Hand crafted filter

?	?	?
?	?	?
?	?	?

Learned filter

ETH zürich Soft Robotics Laboratory

Deep Learning: Convolutional Neural Networks



ETH zürich

SoftRobotics Laboratory

Gavrikov et al

Deep Learning: Convolutional Neural Networks

3	4	8	1
2	7	9	6
4	5	0	4
7	4	1	3

Max Pooling

2x2

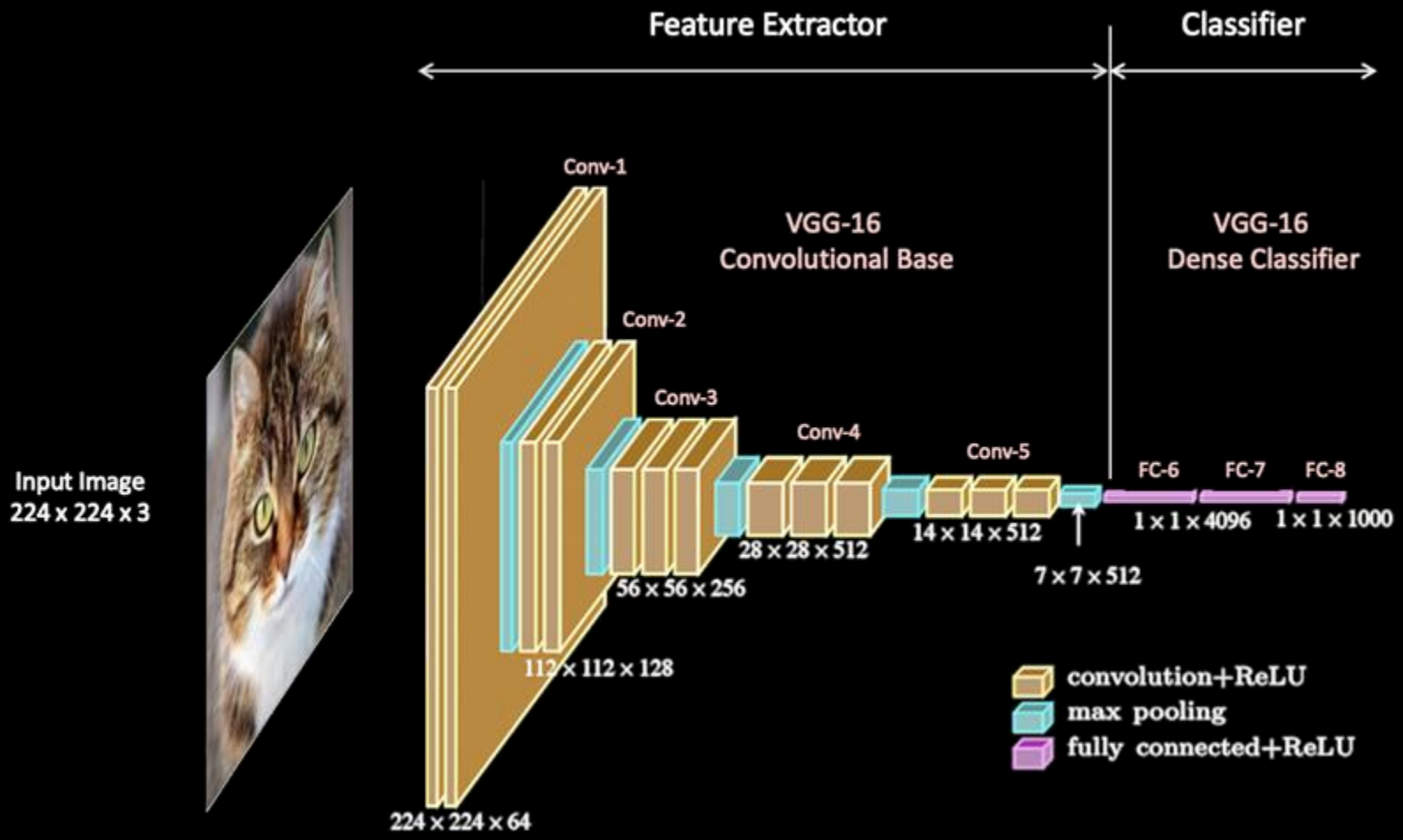
7	9
7	4

Average Pooling

2x2

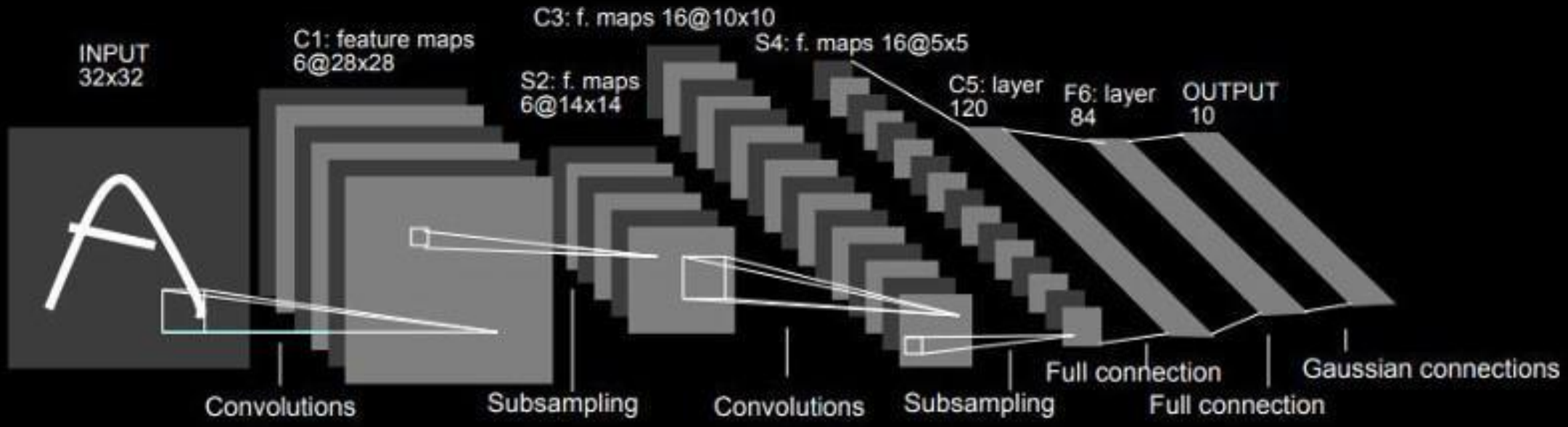
4	6
5	2

Deep Learning: Convolutional Neural Networks



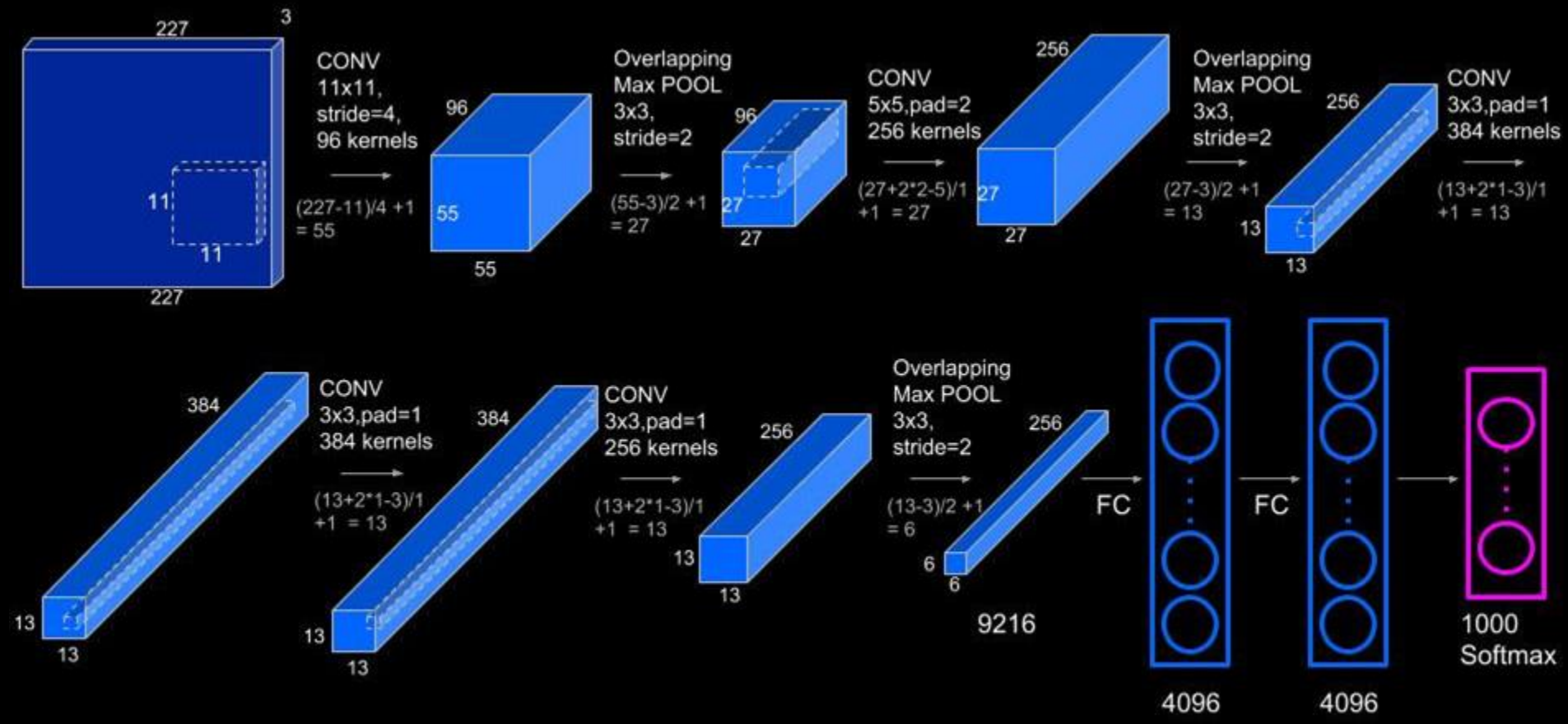
https://learnopencv.com/understanding-convolutional-neural-networks-cnn/

Deep Learning: Convolutional Neural Networks Arch



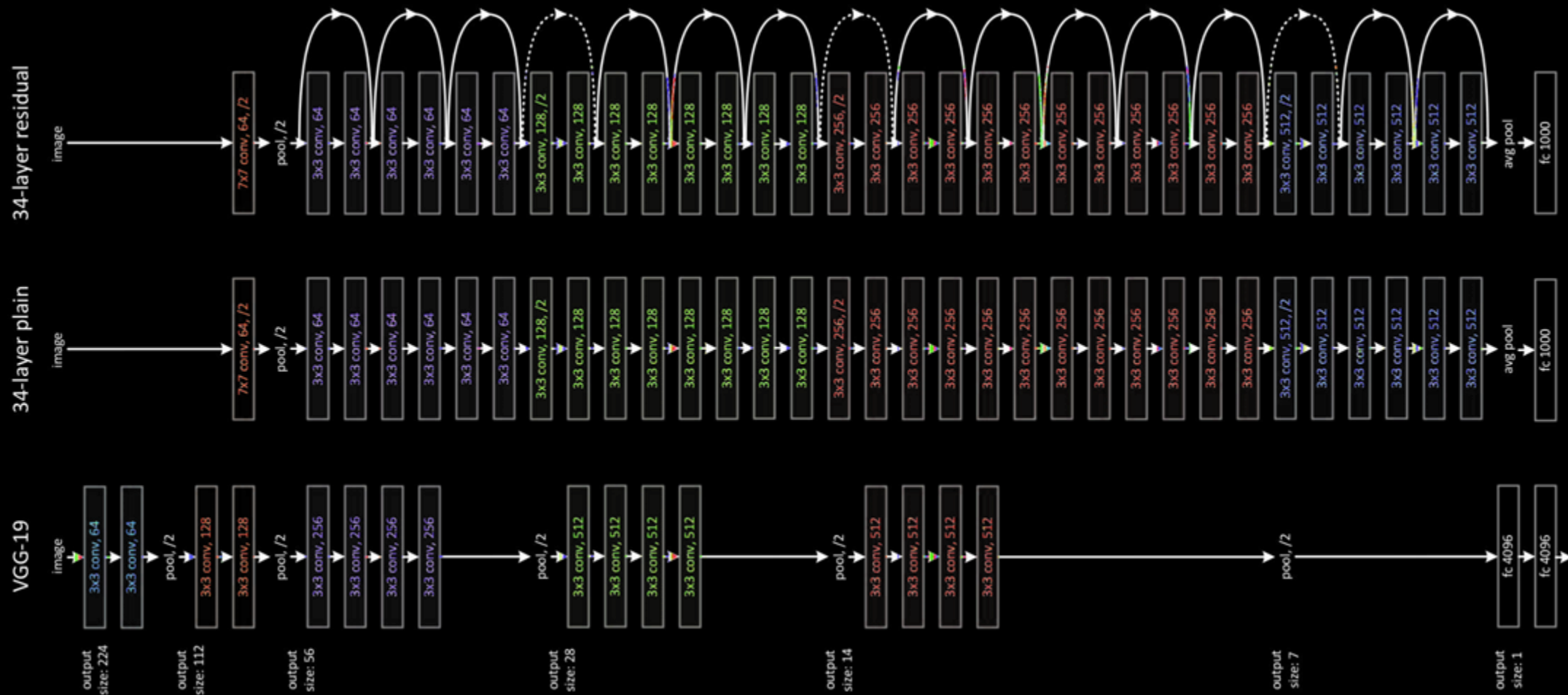
Lecun et al

Deep Learning: Convolutional Neural Networks Arch



https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/

Deep Learning: Convolutional Neural Networks Arch



He et al, 2015

Deep Learning: Outro

Tasks

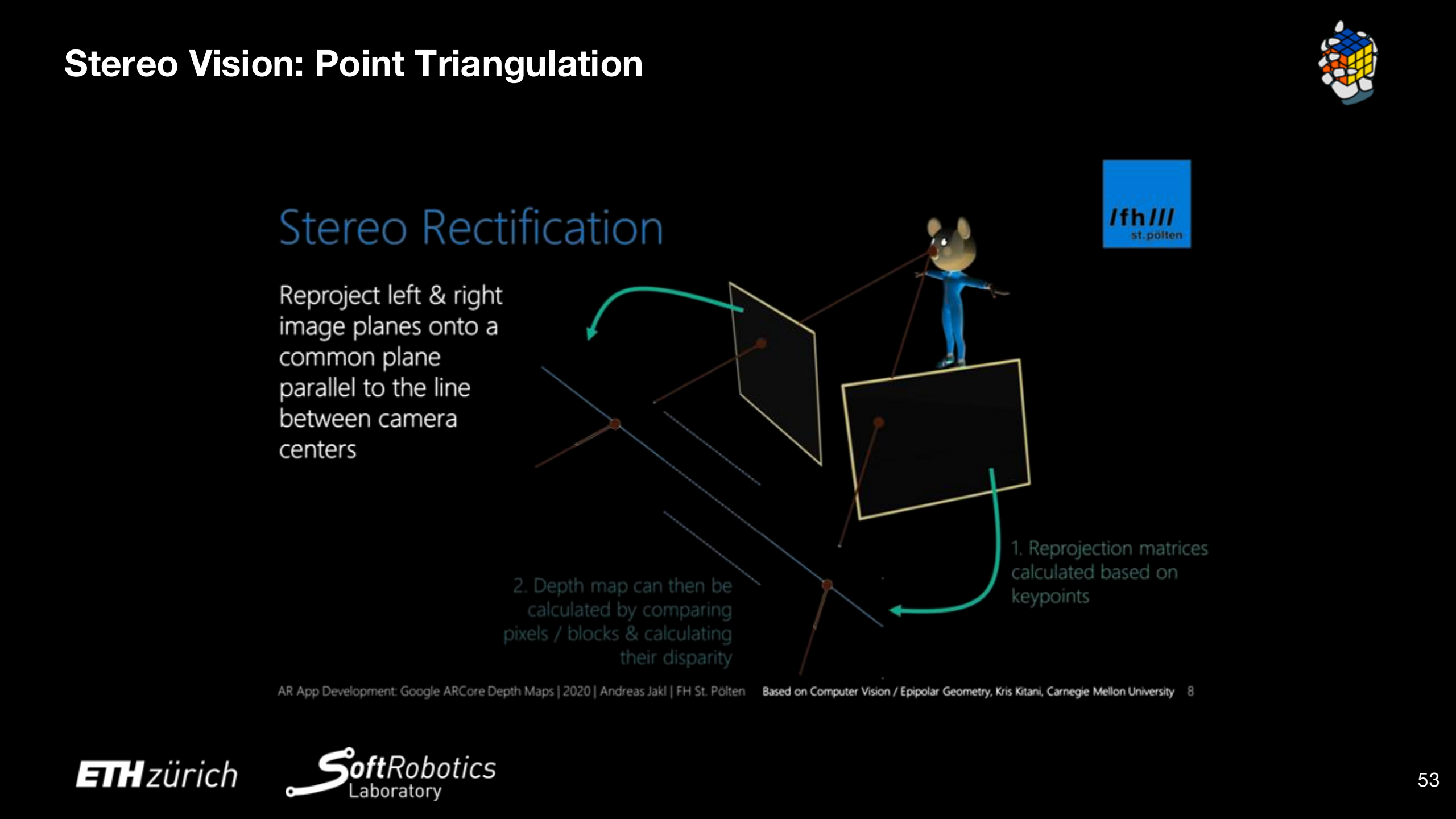
Convolutional
neural networks

Neural networks
and how they work

Common
architectures

Part 2c:
Stereo vision

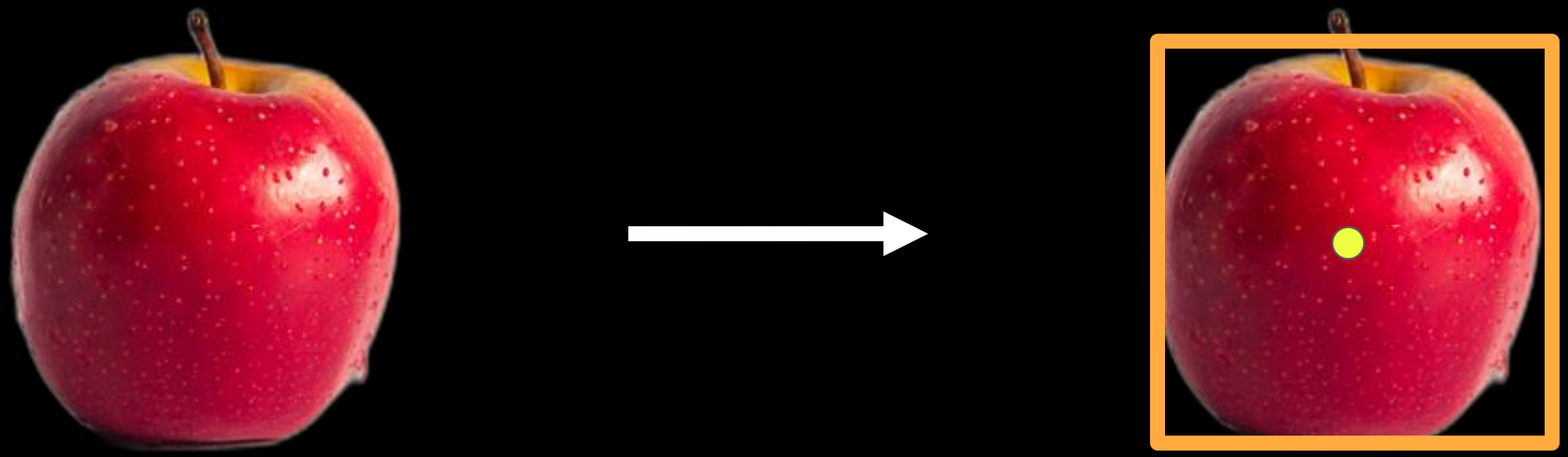
Stereo Vision: Point Triangulation



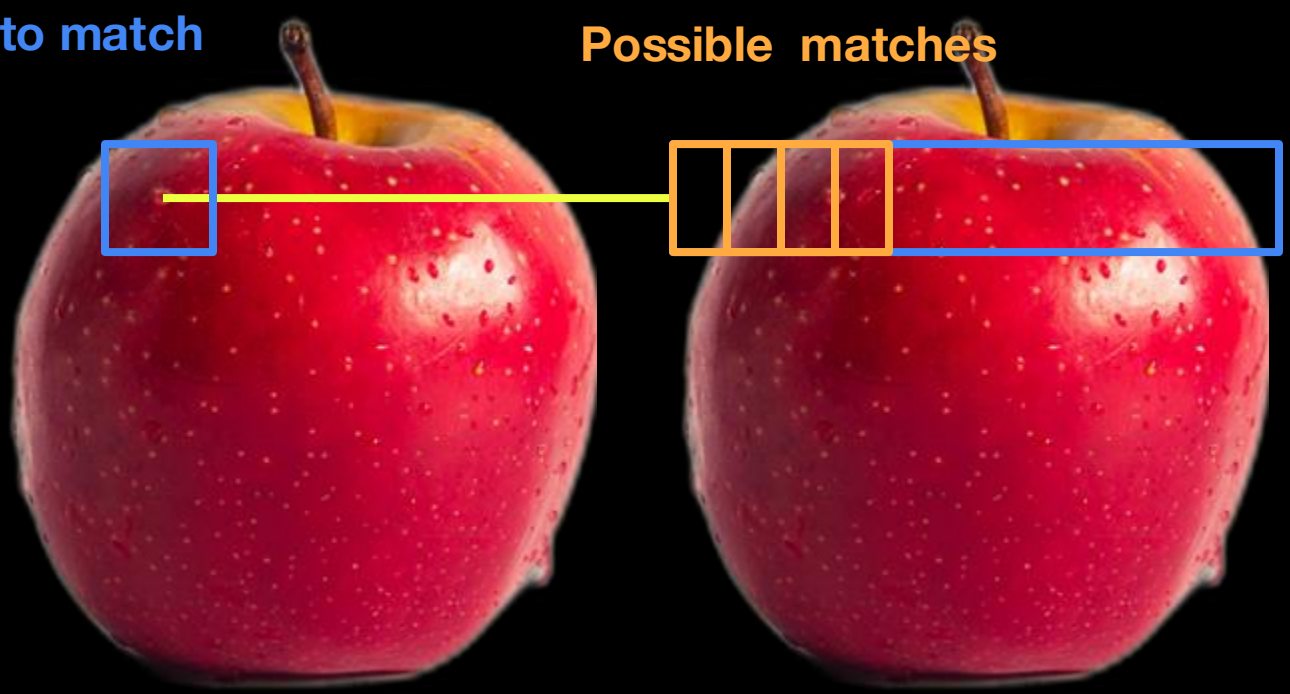
Stereo Vision: Point Triangulation

Stereo Vision: Point Triangulation

Stereo Vision: Point Triangulation



Stereo Vision: Point Matching



Stereo Vision: Point Matching

$$s = \sum_{(u,v) \in \mathbf{I}} (\mathbf{I}_1[u,v] - \mathbf{I}_2[u,v])^2$$

Sum of Squared Differences

$$s = \sum_{(u,v) \in \mathbf{I}} |\mathbf{I}_1[u,v] - \mathbf{I}_2[u,v]|$$

Sum of Absolute Differences

Stereo Vision: Point Matching