Deep Learning for Computer Vision

Why

What is Computer Vision?

- First defined in the 60s in artificial intelligence groups
- "Mimic the human visual system"
- Center block of robotic intelligence

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

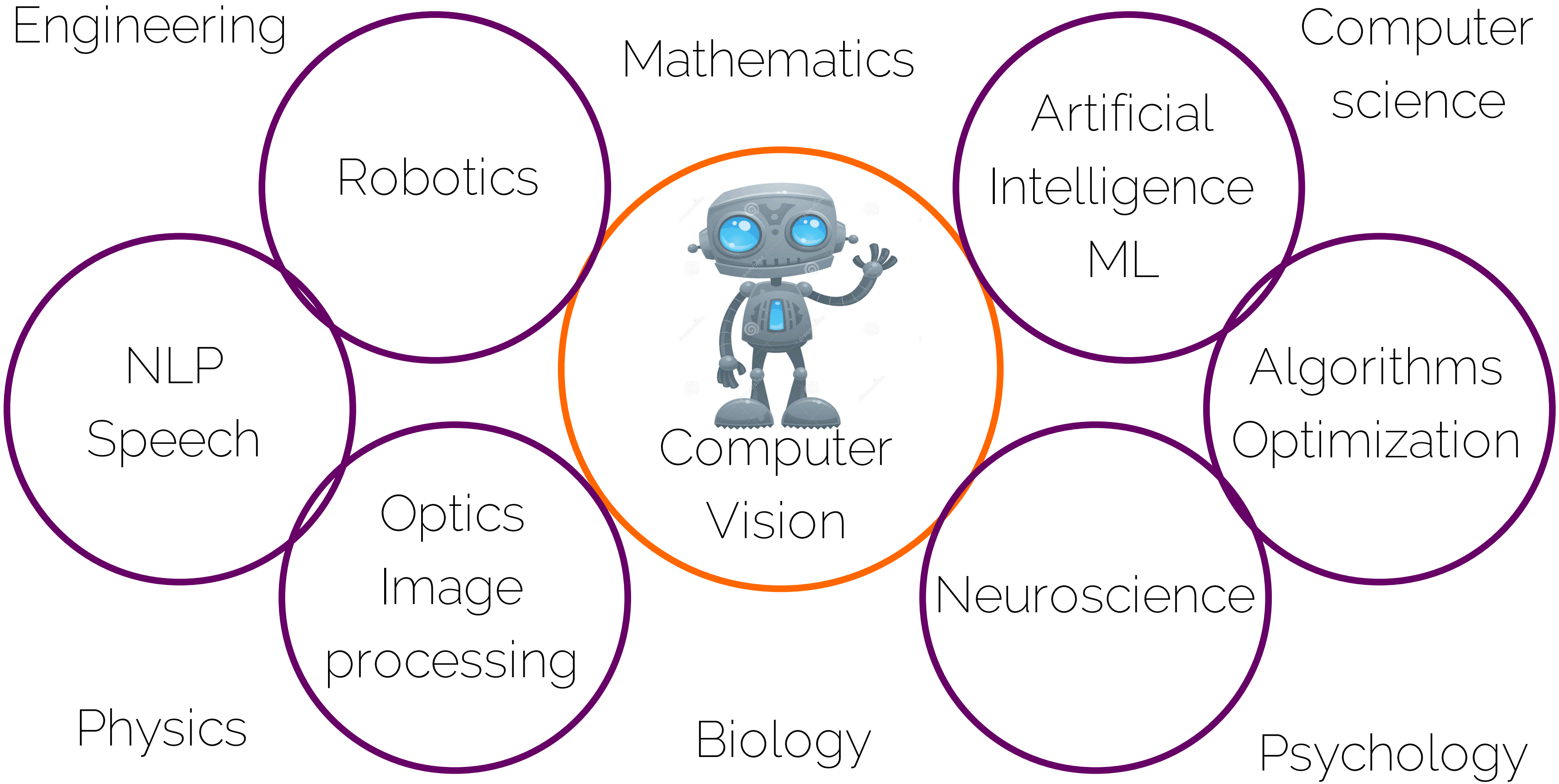
THE SUMMER VISION PROJECT

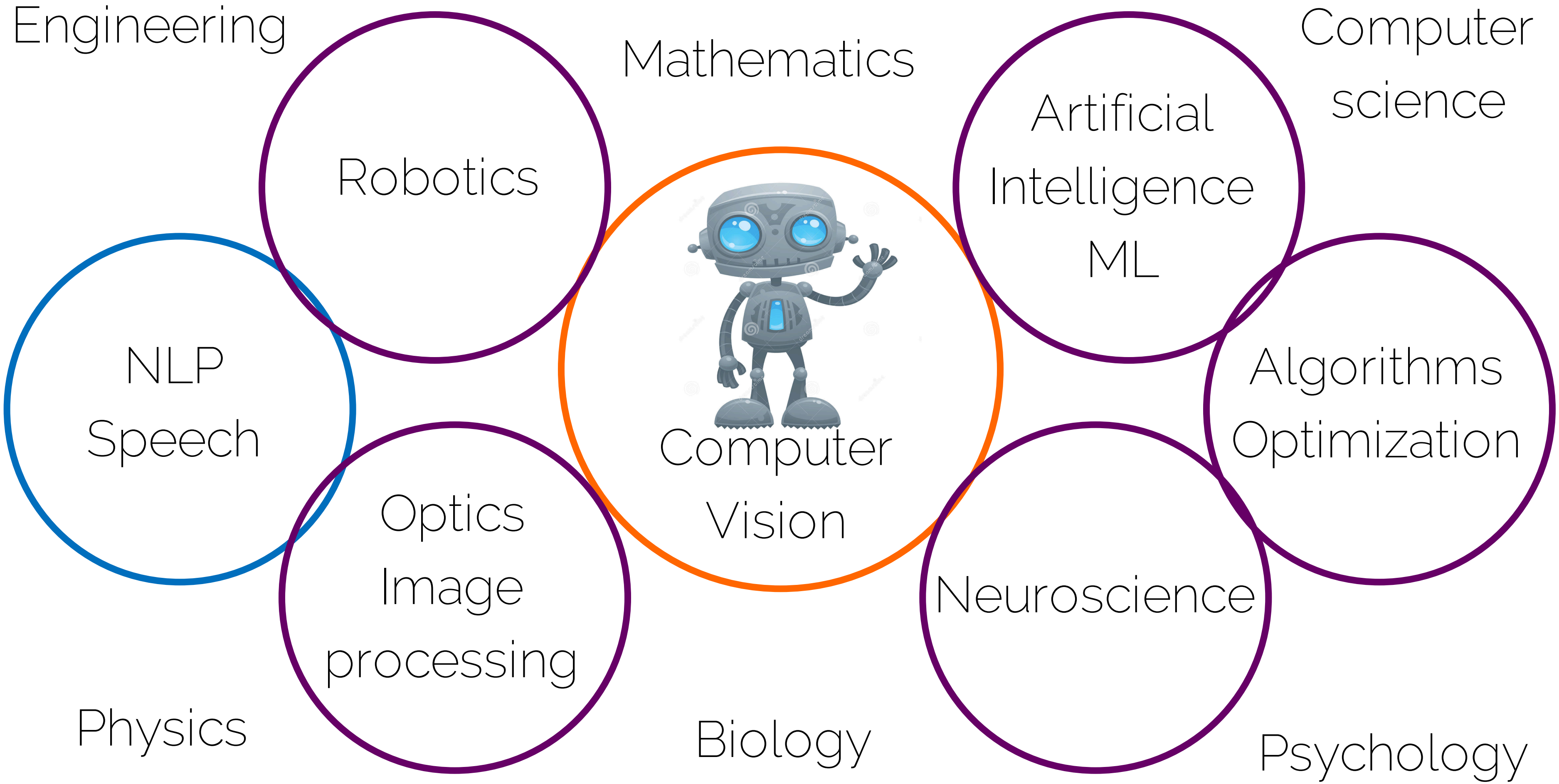
Seymour Papert

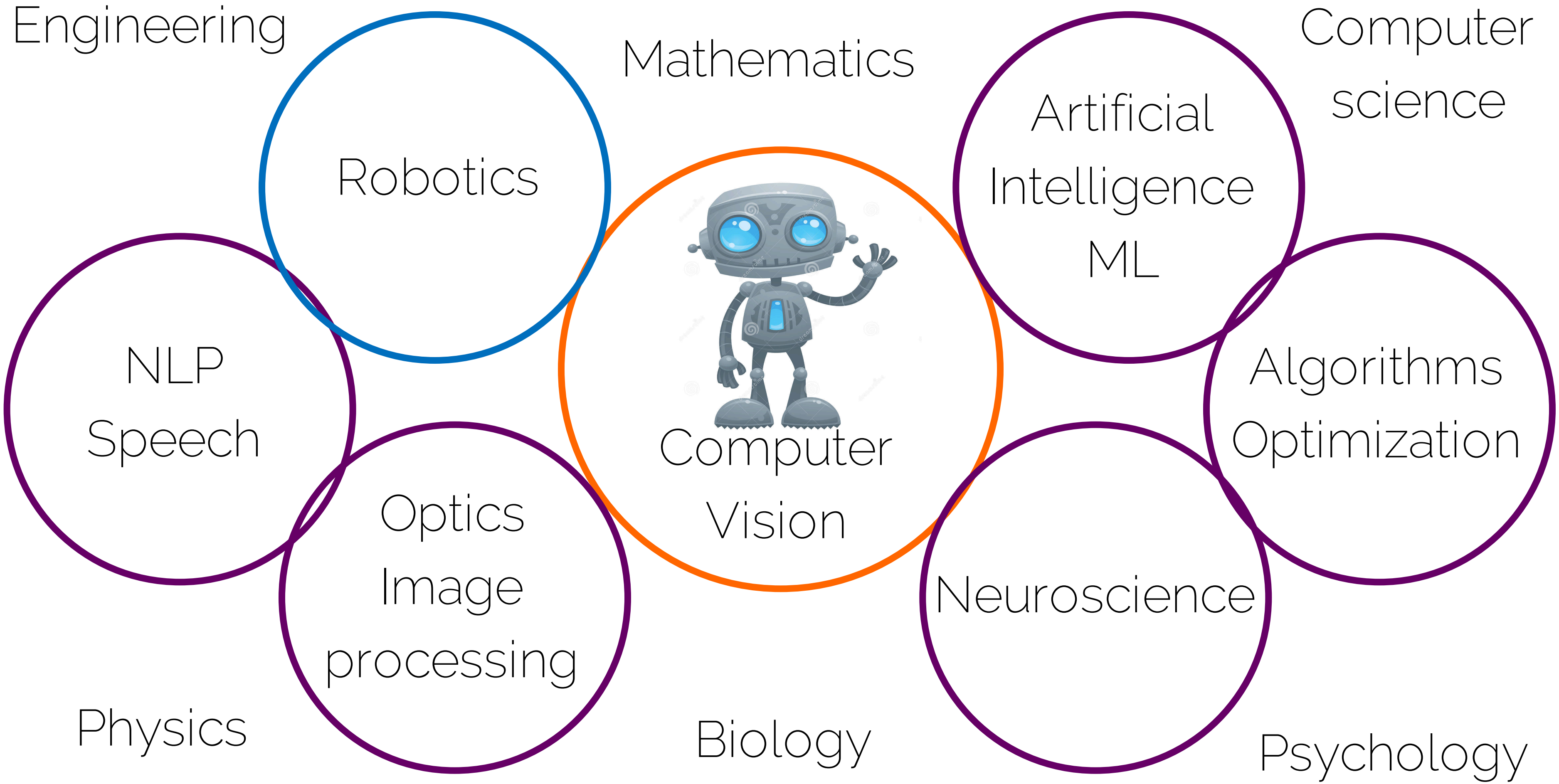
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

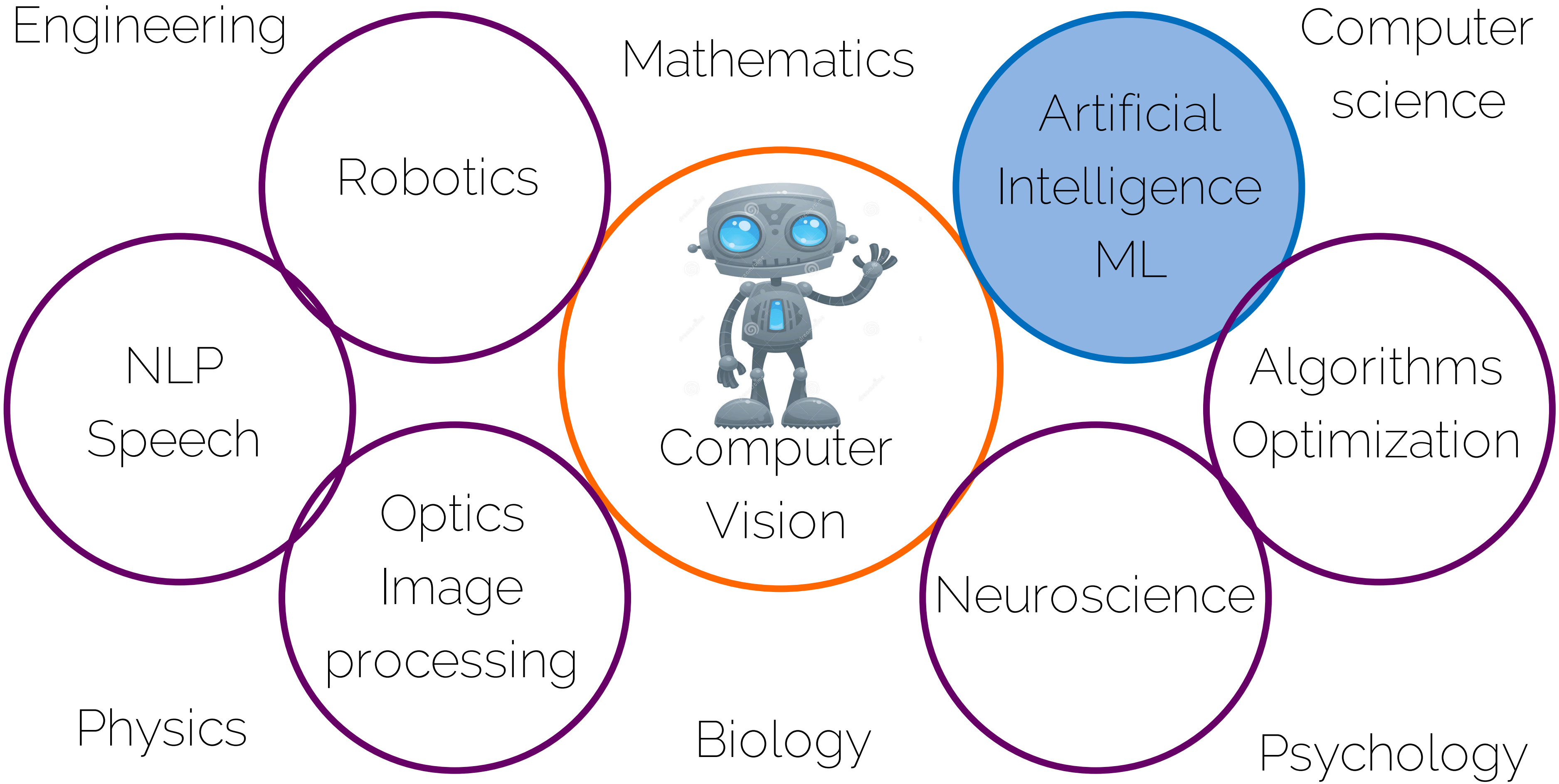
Some decades later…

Computer
Vision

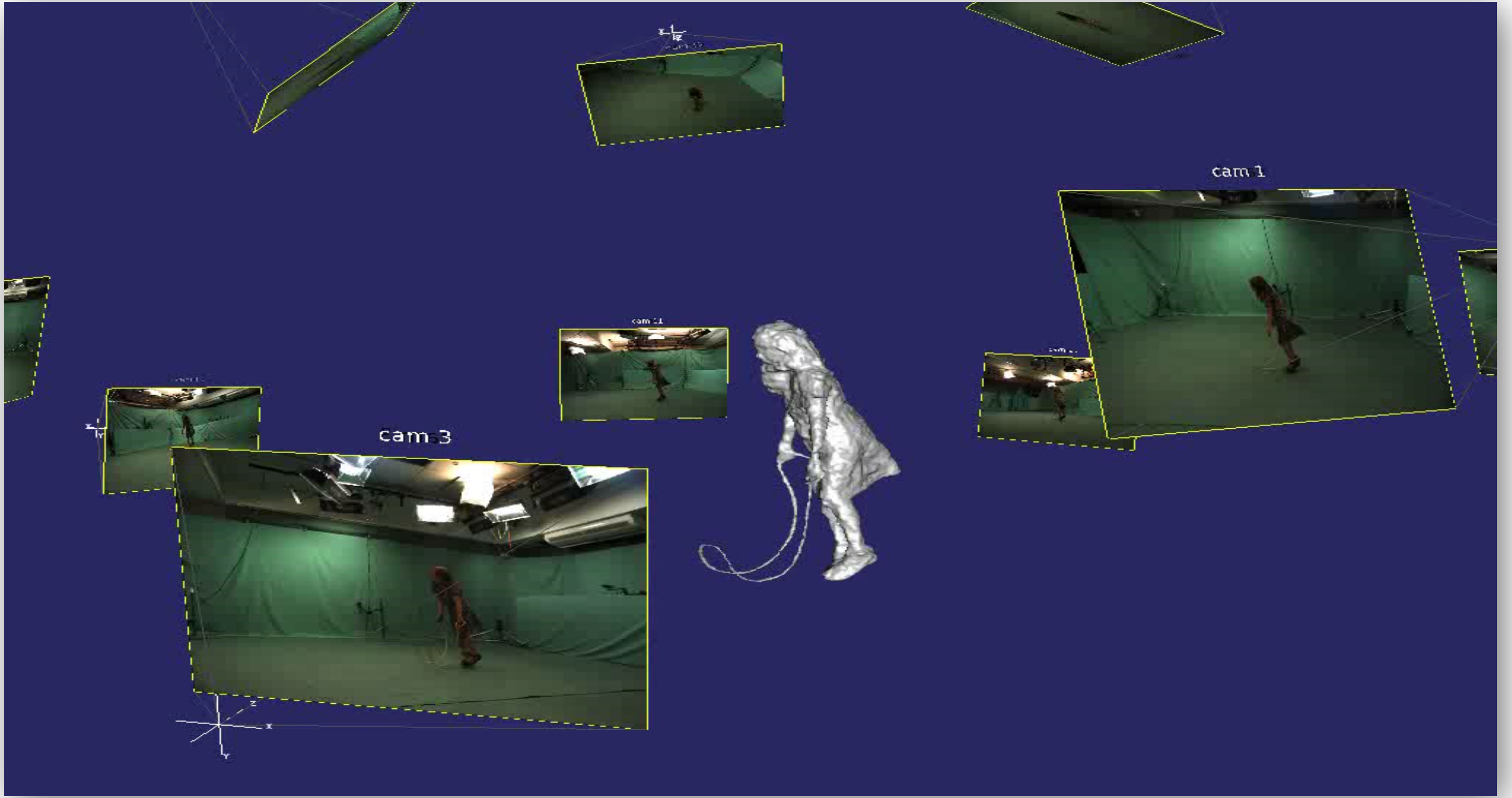

Engineering
Mathematics
Computer science
Robotics
Artificial Intelligence
ML
NLP
Speech
Computer Vision
Algorithms
Optimization
Optics
Image processing
Neuroscience
Physics
Biology
Psychology


Engineering
Mathematics
Computer science
Robotics
Artificial Intelligence
ML
NLP
Speech
Computer Vision
Algorithms
Optimization
Optics
Image processing
Neuroscience
Physics
Biology
Psychology


Engineering
Mathematics
Computer science
Robotics
Artificial Intelligence
ML
NLP
Speech
Computer Vision
Algorithms
Optimization
Optics
Image
processing
Neuroscience
Physics
Biology
Psychology

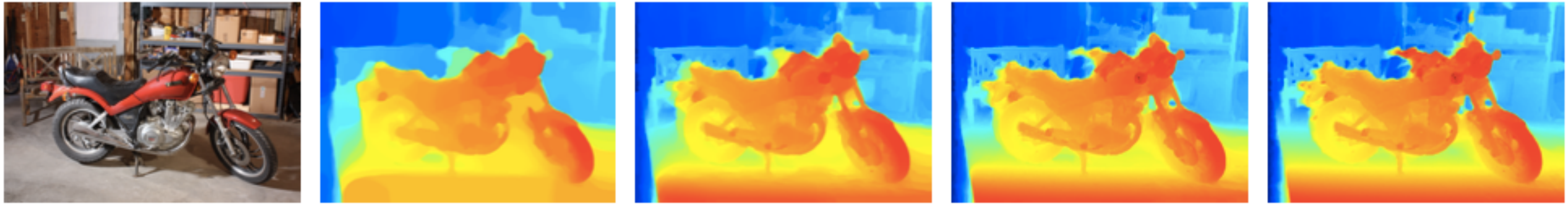

Engineering
Mathematics
Computer science
Robotics
Artificial Intelligence ML
NLP Speech
Computer Vision
Algorithms Optimization
Optics Image processing
Neuroscience
Physics
Biology
Psychology

Computer Vision at TUM

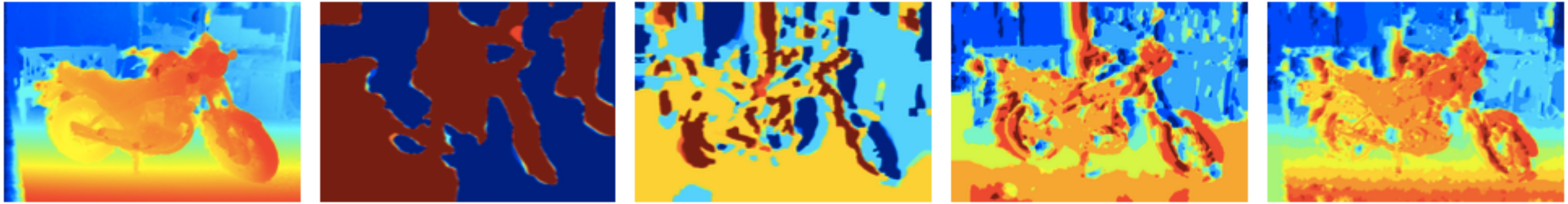


Oswald, Stühmer, Cremers., *ECCV 2014.*

Computer Vision at TUM



One of the input images | Proposed ($L = 2$) | Proposed ($L = 4$) | Proposed ($L = 8$) | Proposed ($L = 16$)



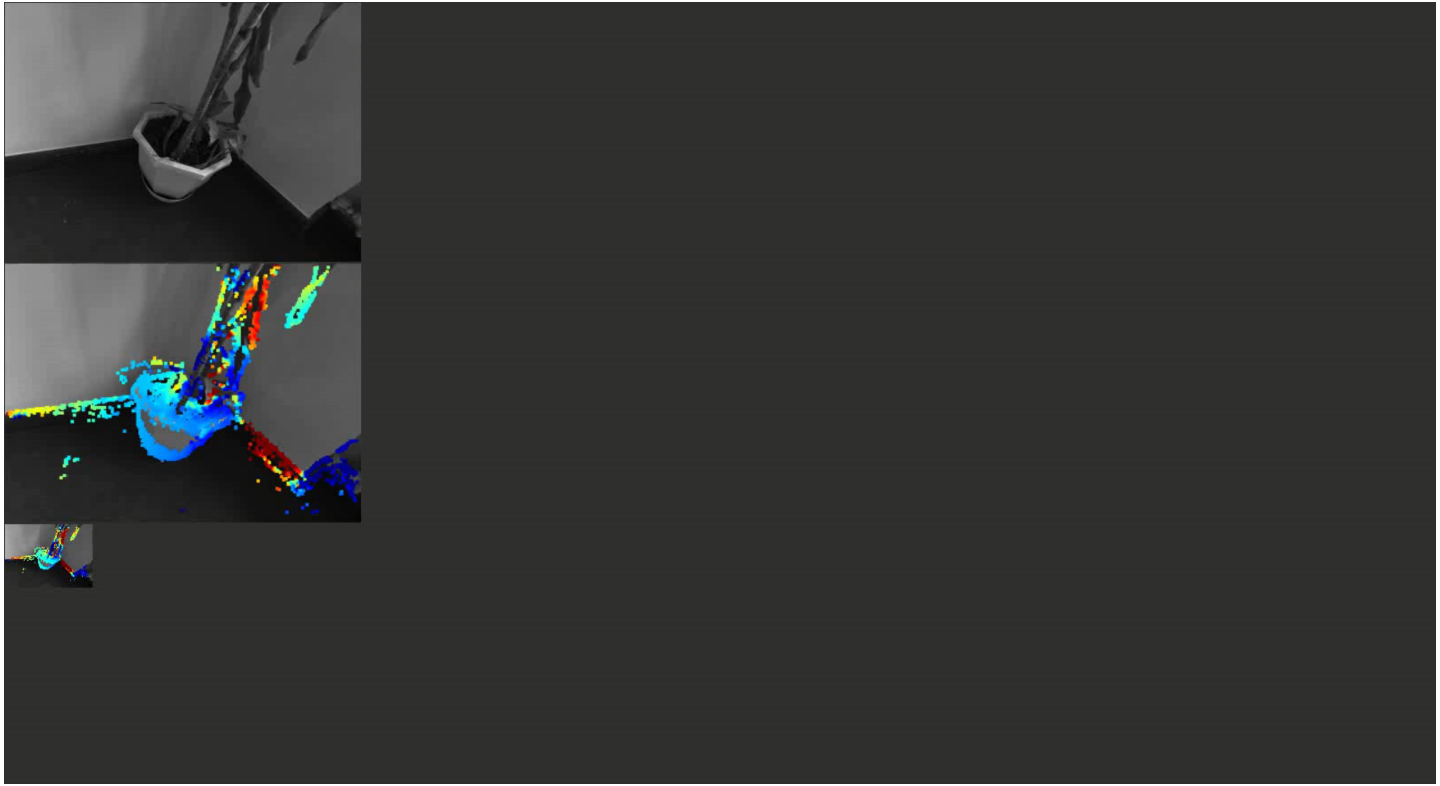
Baseline ($L = 270$) | Baseline ($L = 2$) | Baseline ($L = 4$) | Baseline ($L = 8$) | Baseline ($L = 16$)

Möllenhof, Laude, Moeller, Lellmann, Cremers. *CVPR 2016.*

Computer Vision at TUM

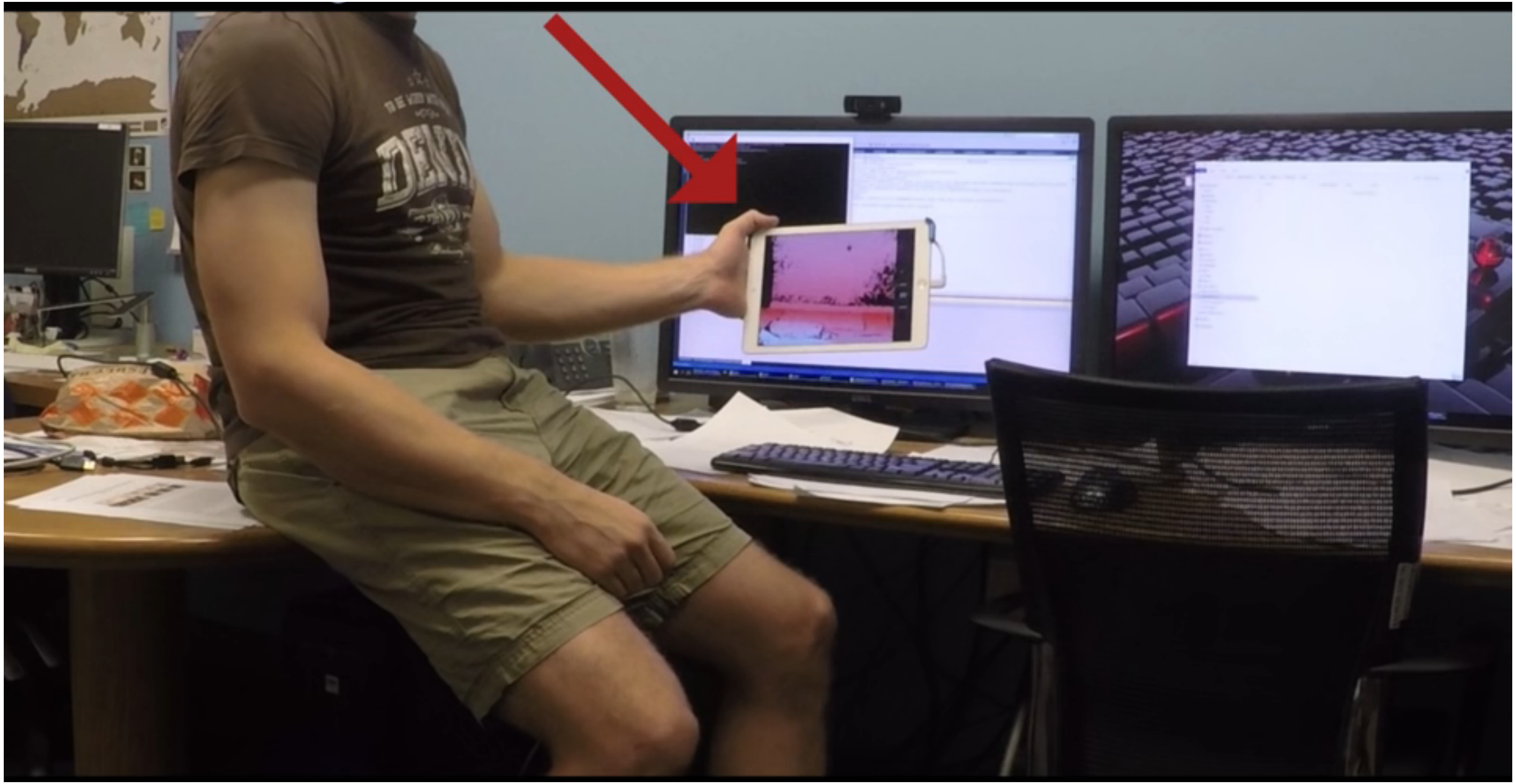
Engel, Schöps, Cremers., *ECCV 2014.*

Computer Vision at TUM



Engel, Koltun, Cremers., *PAMI 2017.*

Computer Vision at TUM



BundleFusion: Dai, Niessner, Zollhoefer, Izadi, Theobalt, ToG *2017.*

Computer Vision at TUM

Face2Face: Thies, Zollhoefer, Stamminger, Theobalt, Niessner., CVPR *2016.*

CV lectures at TUM

- Machine Learning for Robotics and Computer Vision
- Computer Vision 1: Variational Methods
- Computer Vision 2: Multiple View Geometry
- Convex optimization for Machine Learning and Computer Vision
- Probabilistic Graphical Models in Computer Vision
- Analysis of Three-Dimensional Shapes

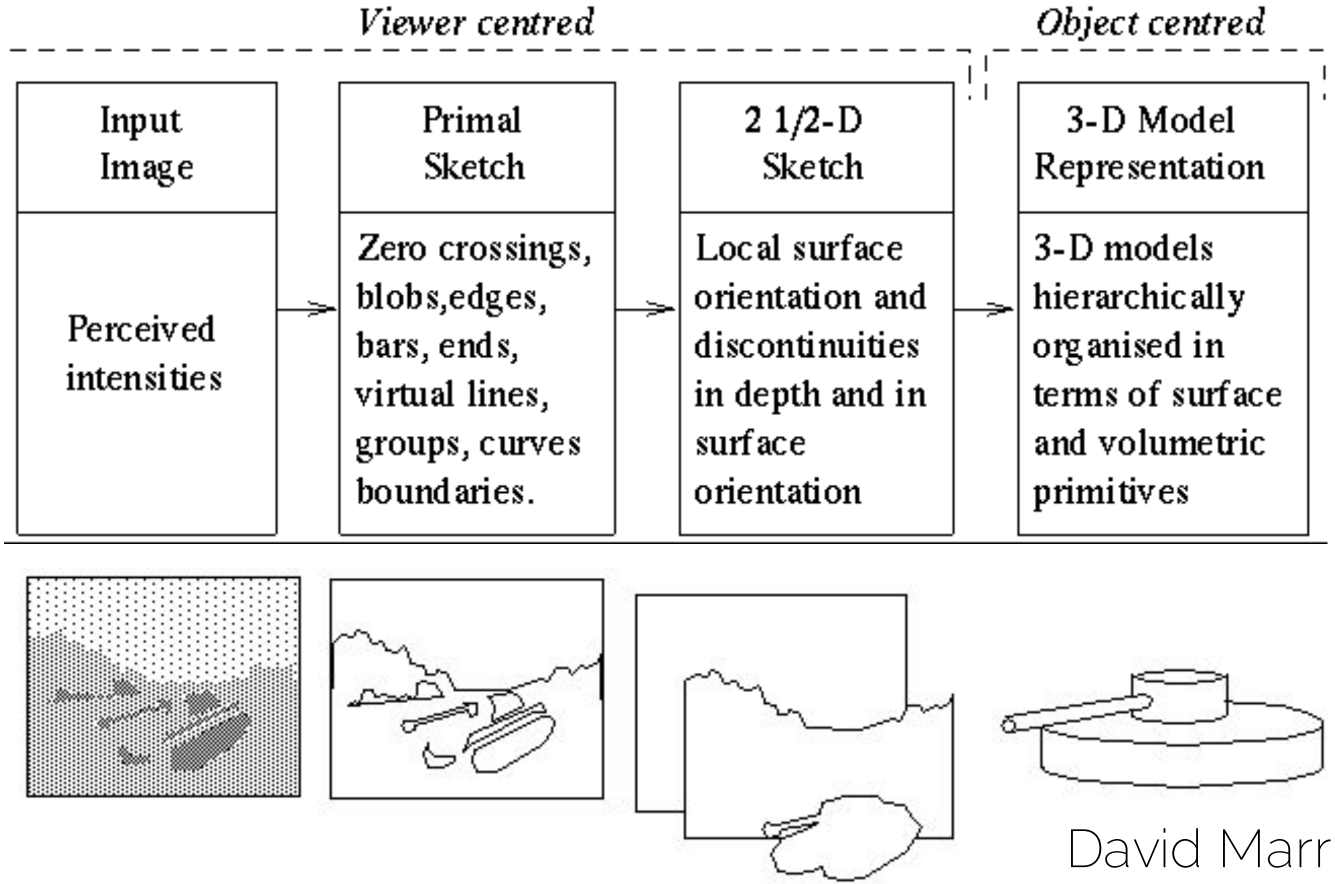
Problem solving

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1–4. The three levels at which any machine carrying out an information-processing task must be understood.

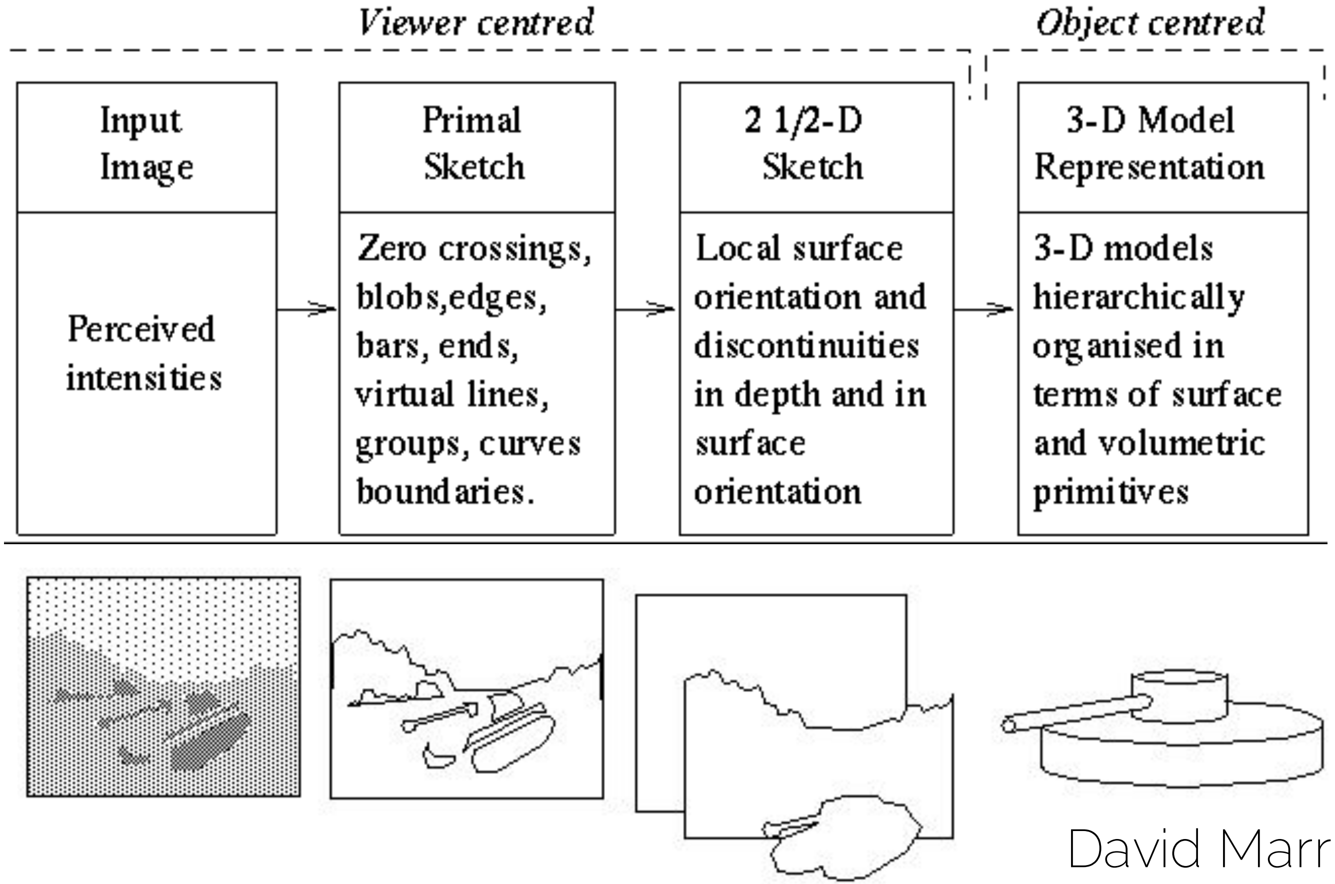
David Marr.

Stages of vision



David Marr. Vision. 1982

Stages of vision



David Marr. Vision. 1982

Image classification

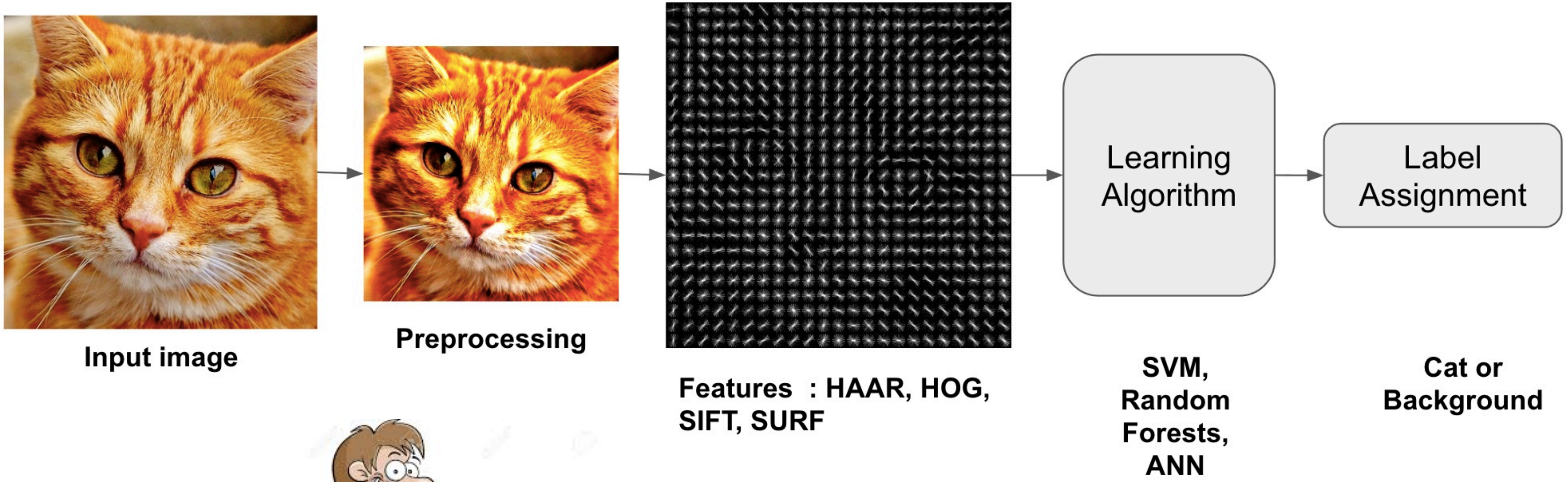
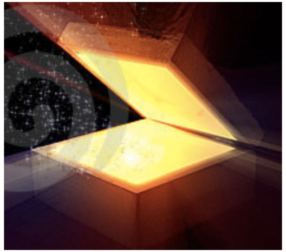


Image classification

Input image

→ Awesome magic box → Label Assignment

Cat or Background



Open the box

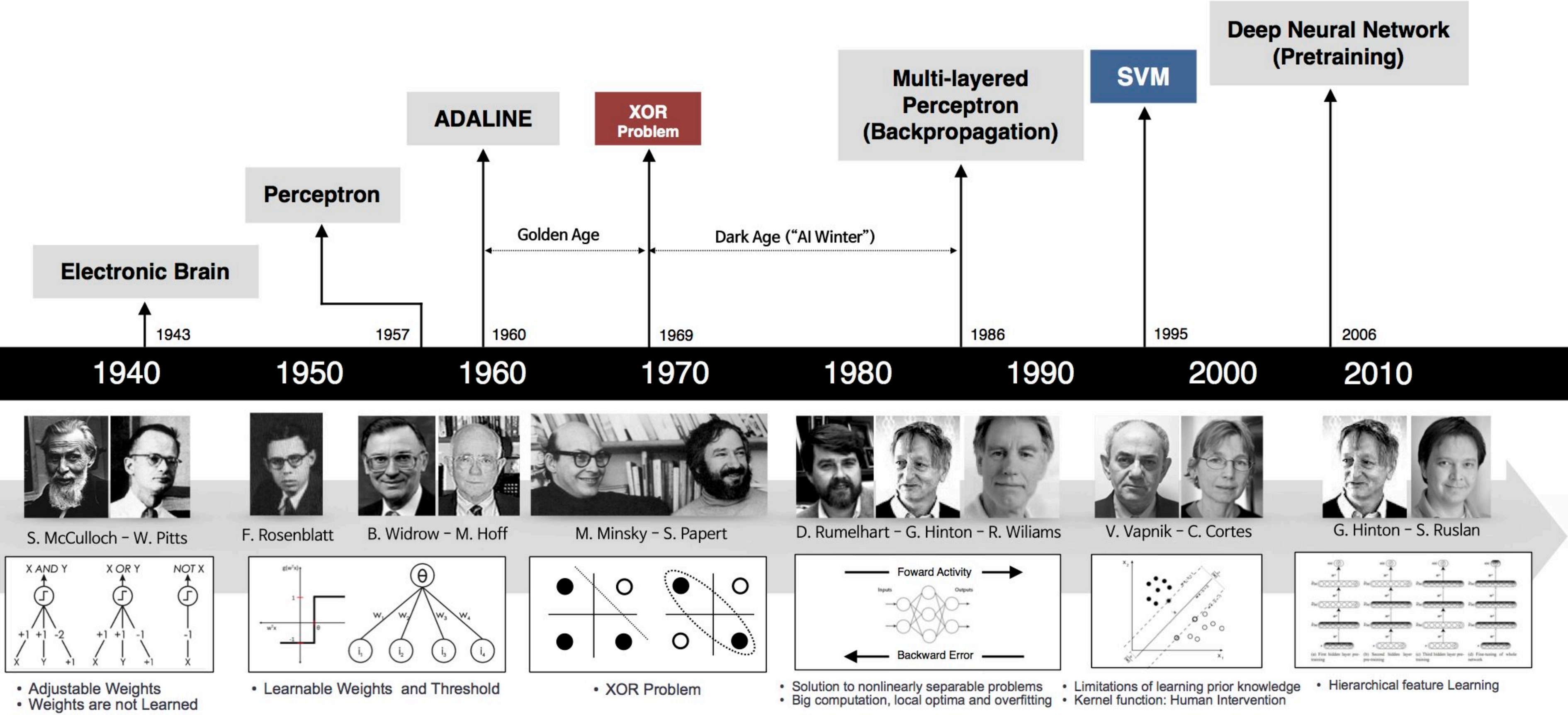
Become magicians

Post 2012

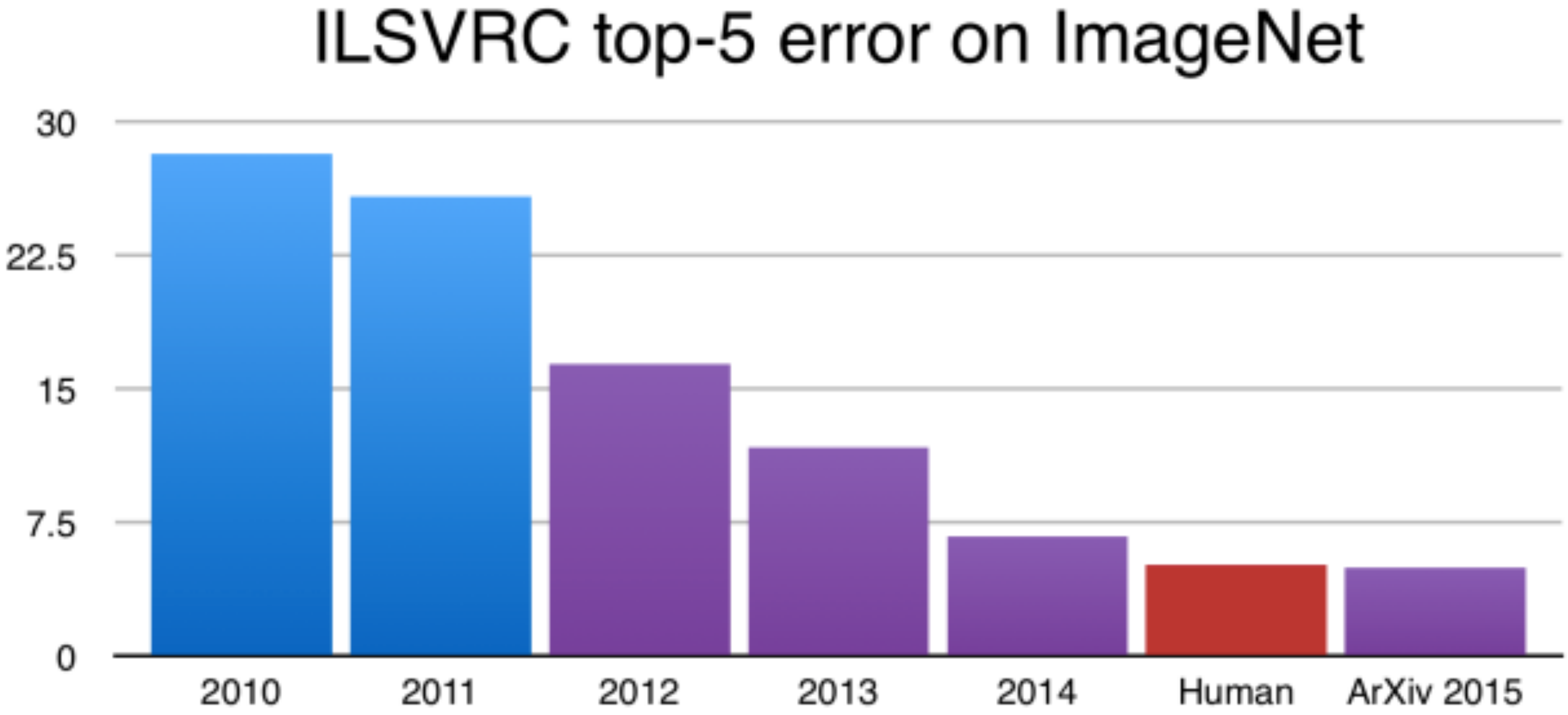
Deep Learning for Computer Vision

… magic?

Deep Learning History

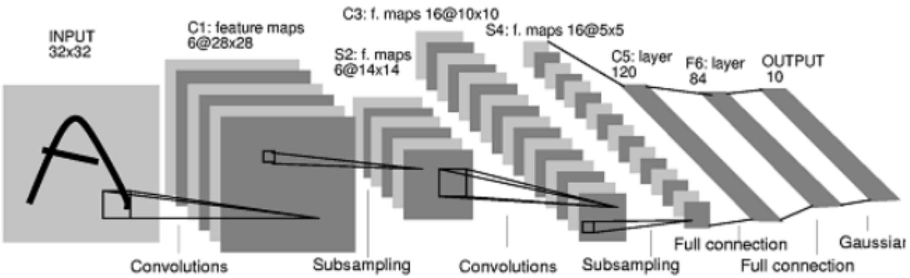


The empire strikes back



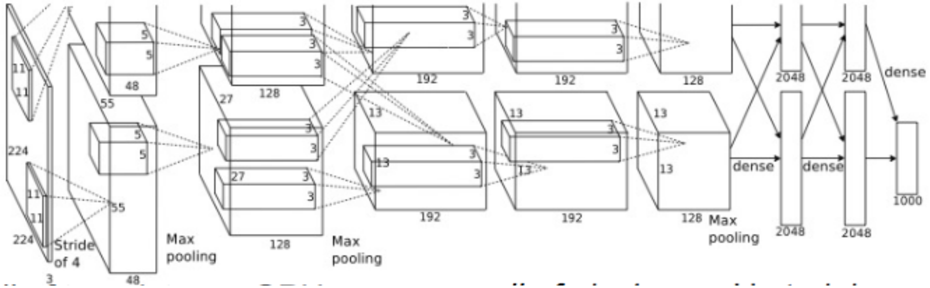
What has changed?

1988
LeCun
et al.



- MNIST digit recognition dataset
- 10^7 pixels used in training

2012
Krizhevsky
et al.



- ImageNet image recognition dataset
- 10^{14} pixels used in training

What made this possible?

Big Data

Models know where to learn from

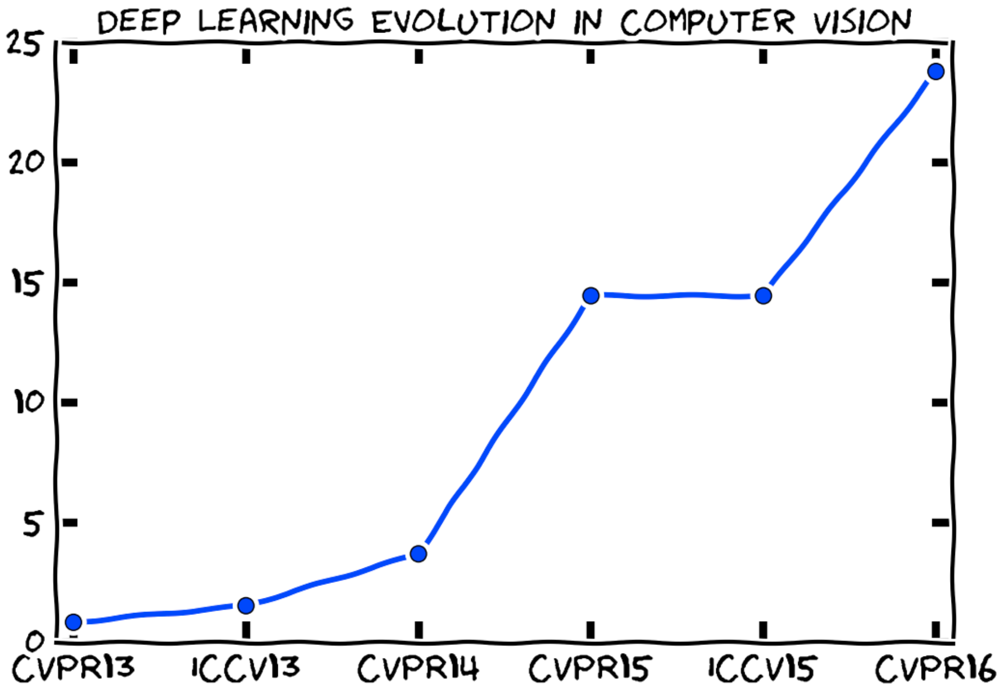
Hardware

Models are trainable

Deep

Models are complex

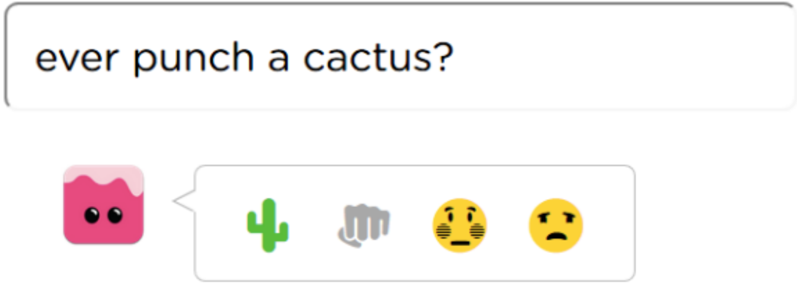
Deep Learning and Computer Vision



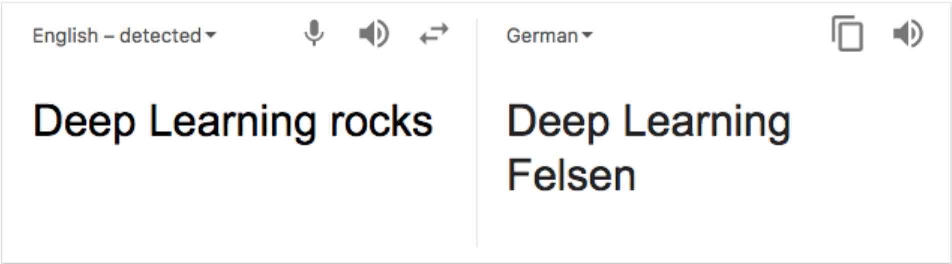
Credits: Dr. Pont-Tuset, ETH Zurich

Deep Learning nowadays

AlphaGo

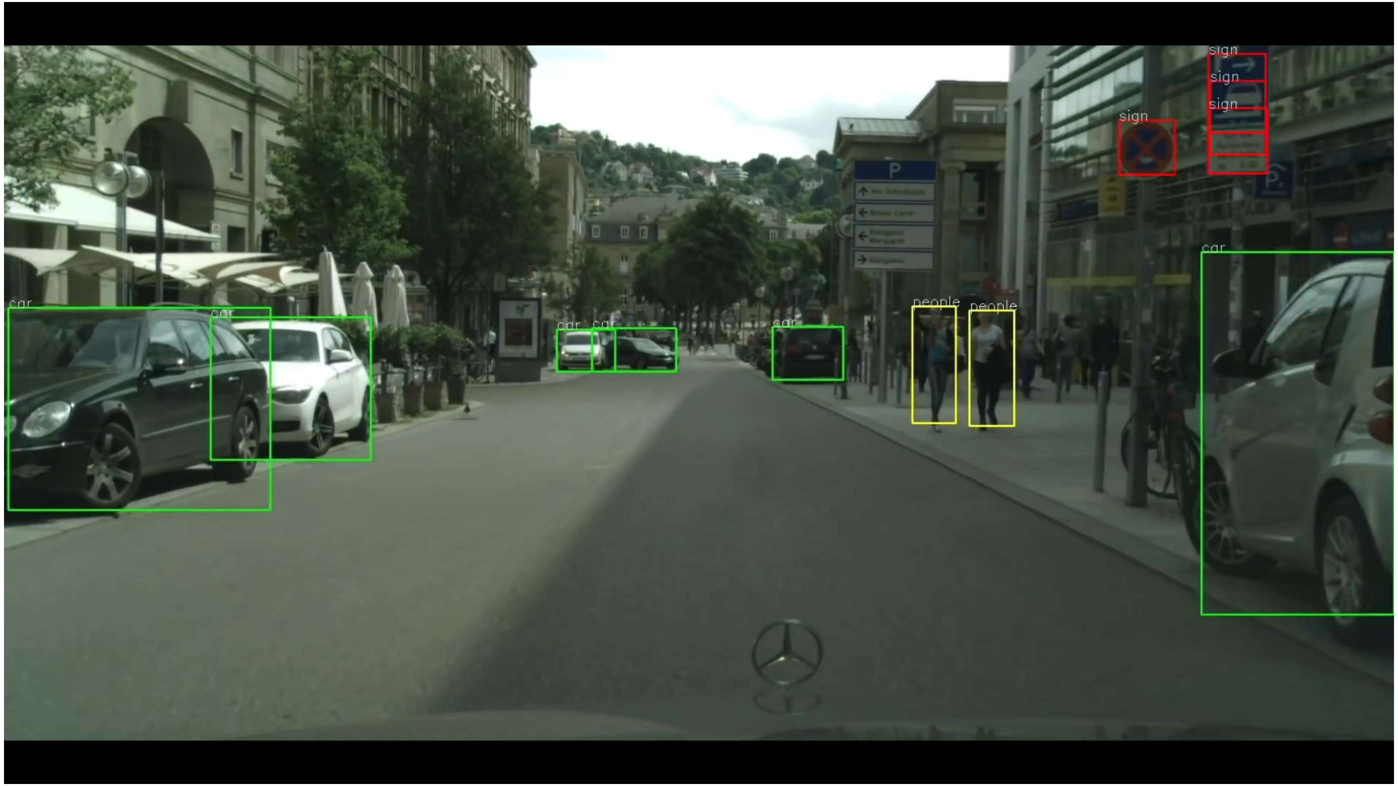


Emoticon suggestion



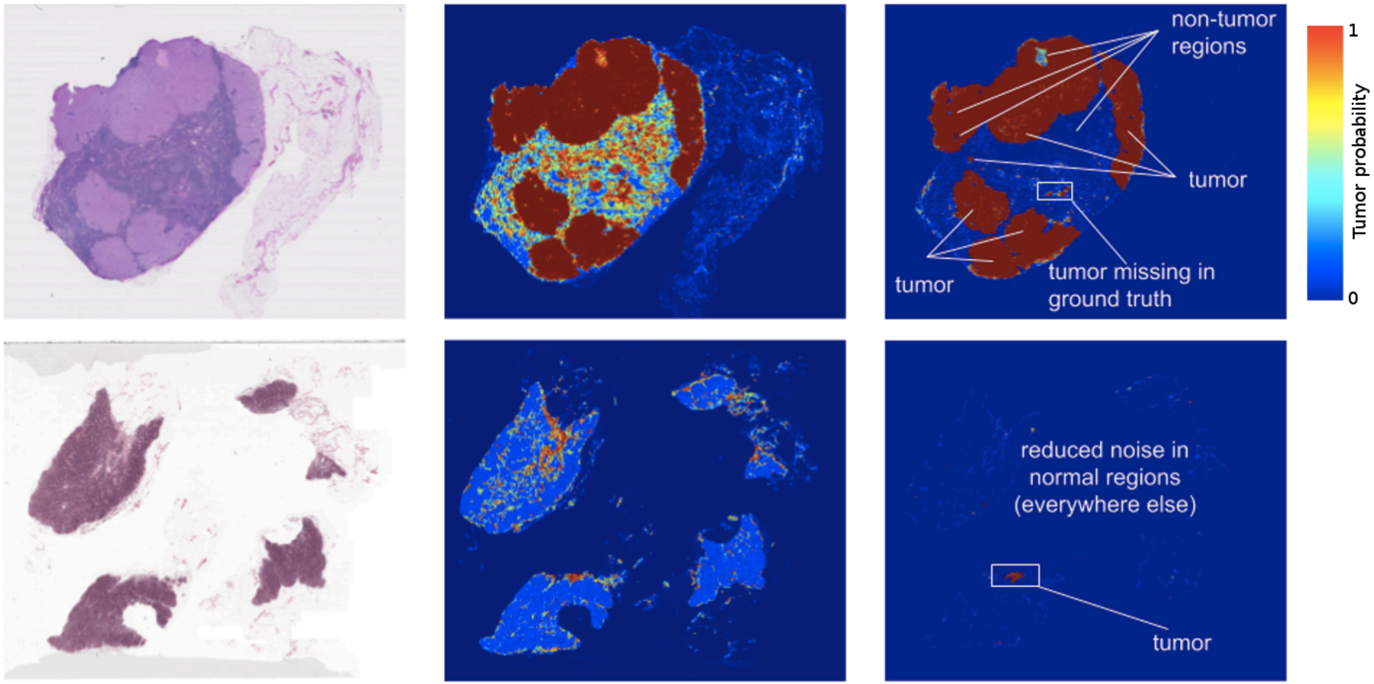
Machine translation

Deep Learning nowadays



Self-driving cars

Deep Learning nowadays



Healthcare, cancer detection

Deep Learning is everywhere

nervana FiveFocal DEXTRO

MARKABLE SHOP ANY PHOTO™ umbo cv spotscale Buildings Decoded SPORTLOGiQ

wrnch MORPX ALMOND RESEARCH Neon BODY x LABS

CVPR 2016 sponsors

Deep Learning is everywhere

CVPR 2016 sponsors

Deep Learning is everywhere

CVPR 2016 sponsors

Deep Learning is everywhere

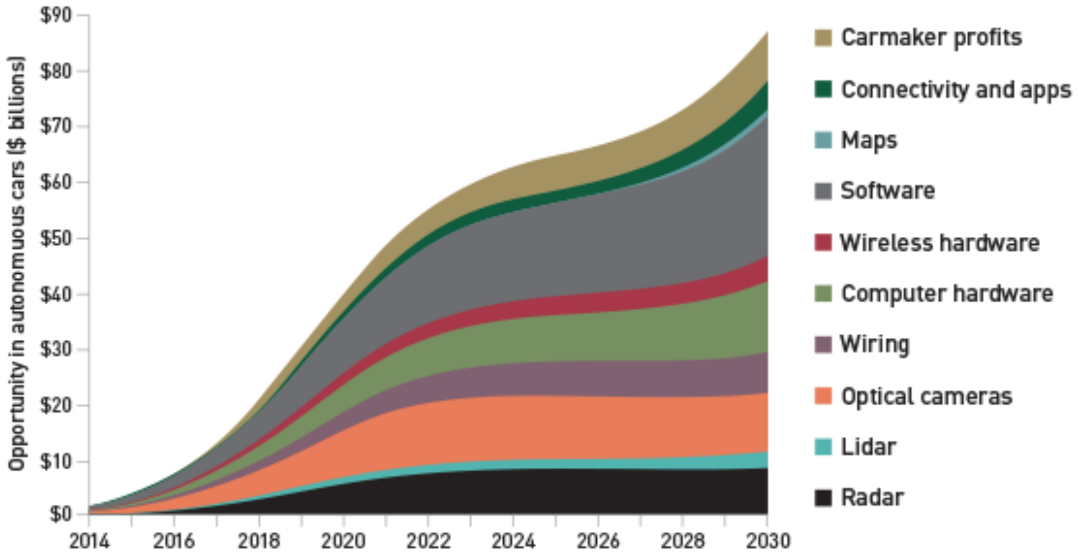
iRobot
SKYDIO
UBER ADVANCED TECHNOLOGIES CENTER
Baidu 百度
DAQRI
SENSETIME 商湯集團
Sighthound
Linkface
COGNEX
IBM Research
ebay
here
spare 5
360 www.360.cn

CVPR 2016 sponsors

Deep Learning is everywhere

CVPR 2016 sponsors

Deep Learning market

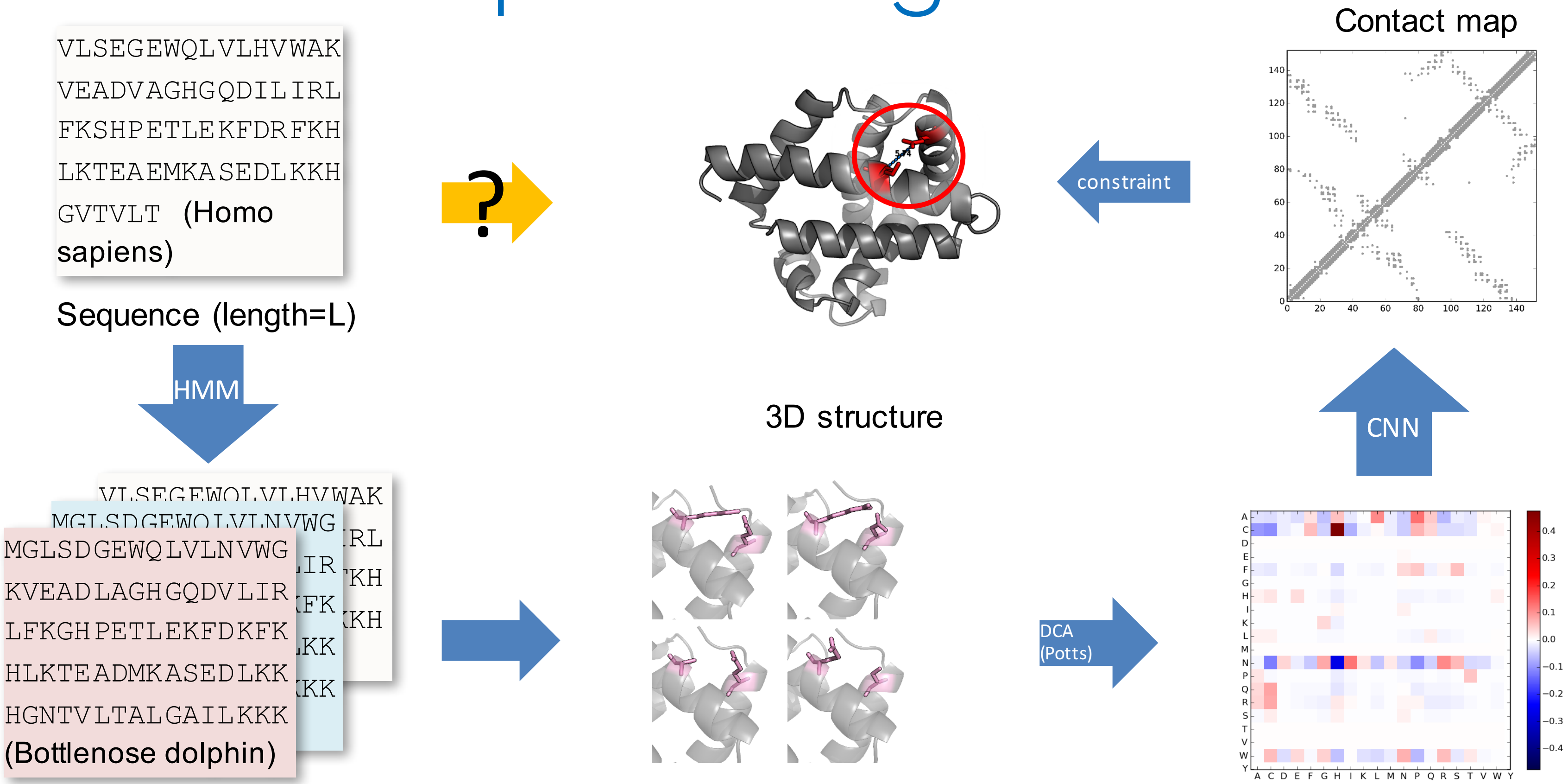


- [...]market research report Deep Learning Market [...] Global Forecasts to 2022", the deep learning market is expected to be worth USD 1,722.9 Million by 2022.

Deep Learning at TUM

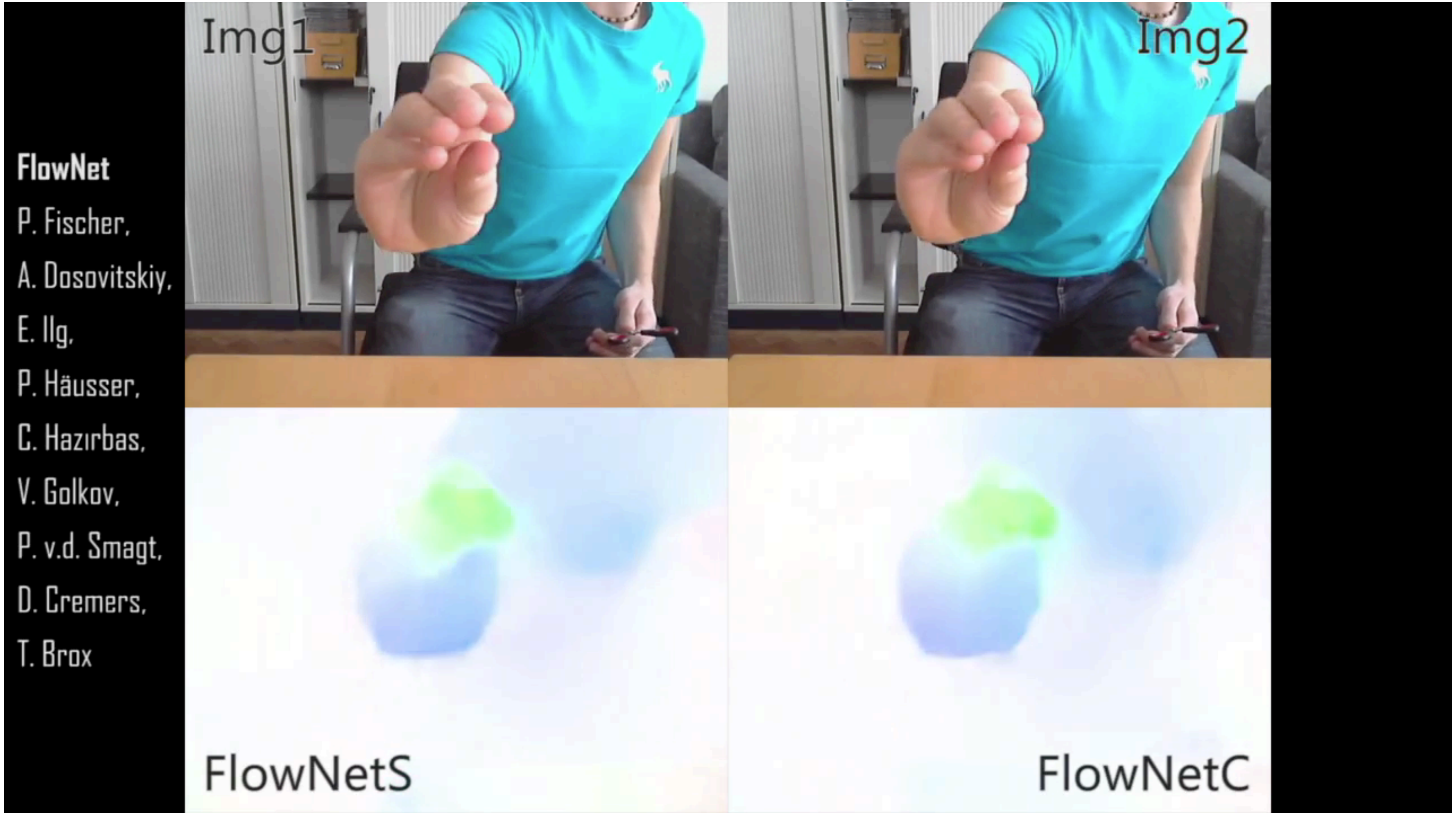
S. Caelles, K.K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool.
One-Shot Video Object Segmentation, *CVPR 2017.*

Deep Learning at TUM



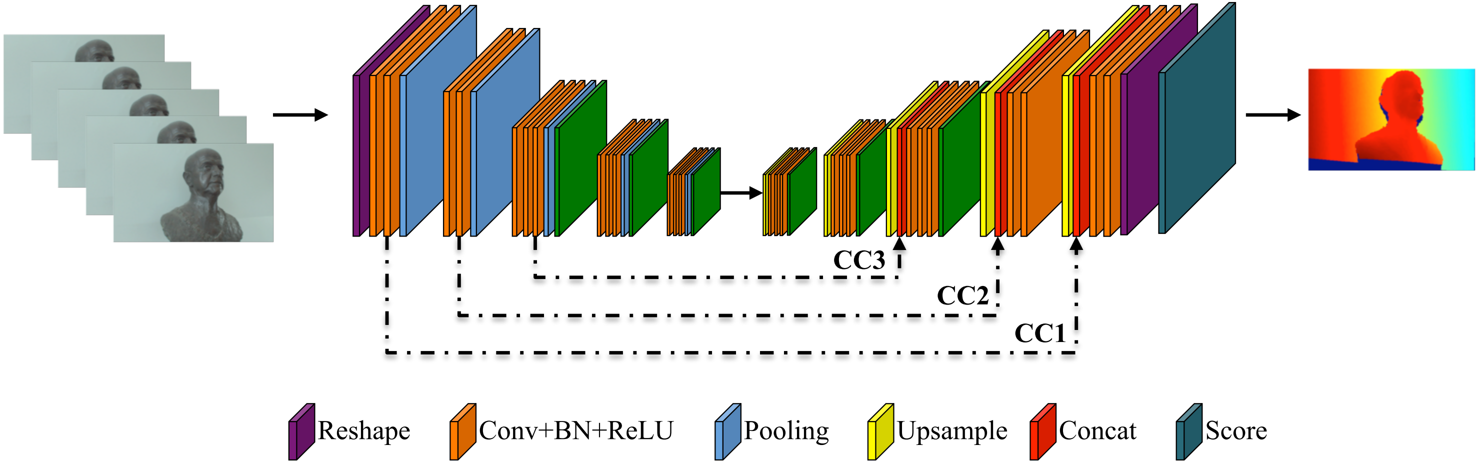
Golkov, Skwark, Golkov, Dosovitskiy, Brox, Meiler, Cremers., *NIPS 2016*

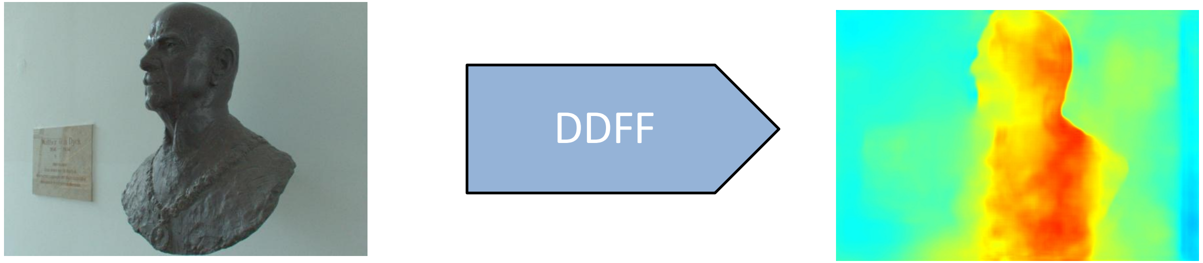
Deep Learning at TUM



Dosovitskiy, Fischer, Ilg, Haeusser, Hazirbas, Golkov, van der Smagt, Cremers, Brox. *ICCV 15*

Deep Learning at TUM





Hazirbas, Leal-Taixé, Cremers. *ICCV 2017 submission*

Computer Vision at TUM

ScanNet Stats:

-Kinect-style RGB-D sensors

-1513 scans of 3D environments

-2.5 Mio RGB-D frames

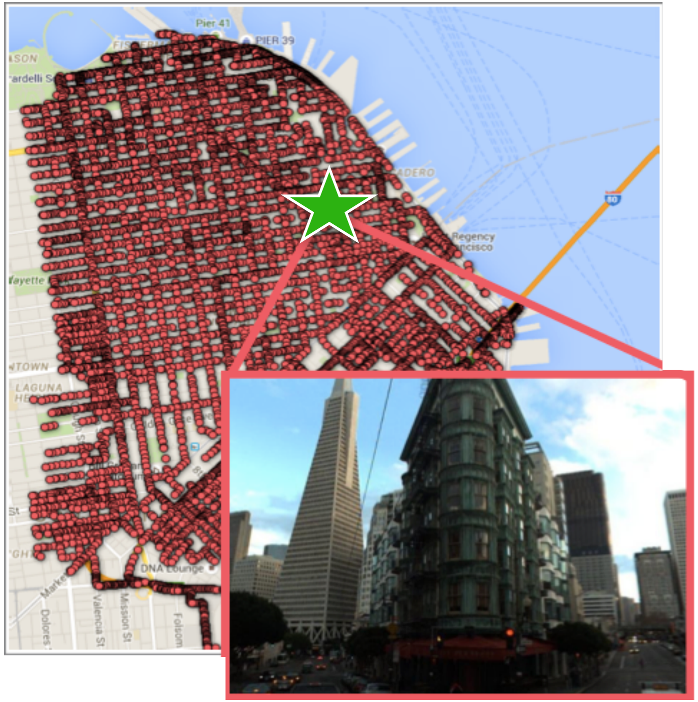
-Dense 3D, crowd-source MTurk labels

-Annotations projected to 2D frames

ScanNet: Dai, Chang, Savva, Halber, Funkhouser, Niessner., CVPR *2017.*

Deep Learning at TUM

Map



Photo

Walch, Hazirbas, Leal-Taixé, Sattler, Hilsenbeck, Cremers. *ICCV 2017 submission*

Deep Learning for Computer Vision

The Team

Lecturers

Dr. Laura Leal-Taixé

Prof. Dr. Matthias Niessner

Tutors

Thomas Frerix

Tim Meinhardt

About the lecture

- Theory: 11 lectures + 2 special lectures
  - Every Thursday (except tomorrow!!)
- Practice: 3 exercises, practical sessions
  - Every Friday
- Project: 1 final project

https://vision.in.tum.de/teaching/ss2017/dl4cv

Grading system

- Theory: 11 lectures + 2 special lectures
  - 60% of the final written exam
- Practice: 3 exercises, practical sessions
  - Bonus 0.3
- Project: 1 final project
  - 40% of the final written exam

https://vision.in.tum.de/teaching/ss2017/dl4cv

Theory lecture

- 3 lectures on Machine Learning Basics
- 3 lectures on Neural Networks
- 2 lectures on Convolutional Neural Networks
- 3 lectures on advanced topics (LSTM, GANs, RL)
- 2 special lectures: research, industry

https://vision.in.tum.de/teaching/ss2017/dl4cv

Practical exercises

- Topics: Linear classifiers, multinomial regression, two-layer neural net.
- Begin: 05.05
- End: 17.05

- Topics: Fully connected nets, dropout, batch normalization.
- Begin: 19.05
- End: 31.05

- Topics: Convolutional neural networks, large-scale project with PyTorch.
- Begin: 02.06
- End: 21.06

Practical exercises

- Friday (05.05): introduction of the exercise
- Next Friday (12.05): Q&A session
- Wednesday (17.05): Delivery deadline at midnight
- Friday (19.05): solution is discussed and new exercise is presented

FIXED DEADLINES!

Final project

- Introduction: 23.06
- Project proposal due date: 28.06
- Starting date: 30.06
- Midterm handout is due: 19.07
- Due date: 10.08
- Poster presentation: 17.08
- Groups of 4

Final information

- Questions regarding the syllabus, exercises or contents of the lecture, use Moodle!
- Slides and exercises will be posted on Moodle
- No recordings will be made!

- Questions regarding organization of the course:

dl4cv@vision.in.tum.de

See you tomorrow!

- Lecture 2 will be held tomorrow!
- When: Friday 27th of April
- Where: MI Hörsaal 2 – Math/Informatics building

The Team