Deep Learning for Computer Vision

The Team

Lecturers

Dr. Laura Leal-Taixé

Prof. Dr. Matthias Niessner

Tutors

Thomas Frerix

Tim Meinhardt

Ji Hou

Andreas Rössler

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.

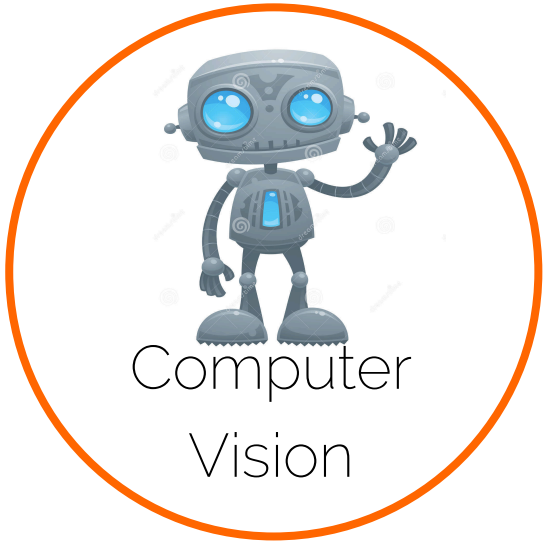
July 7, 1966

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…



Engineering

Mathematics

Computer science

Robotics

Artificial Intelligence ML

NLP Speech

Algorithms Optimization

Computer Vision

Optics Image processing

Neuroscience

Physics

Biology

Psychology

Engineering

Mathematics

Computer science

Robotics

Artificial Intelligence ML

NLP Speech

Algorithms Optimization

Computer Vision

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

Algorithms Optimization

Computer Vision

Optics Image processing

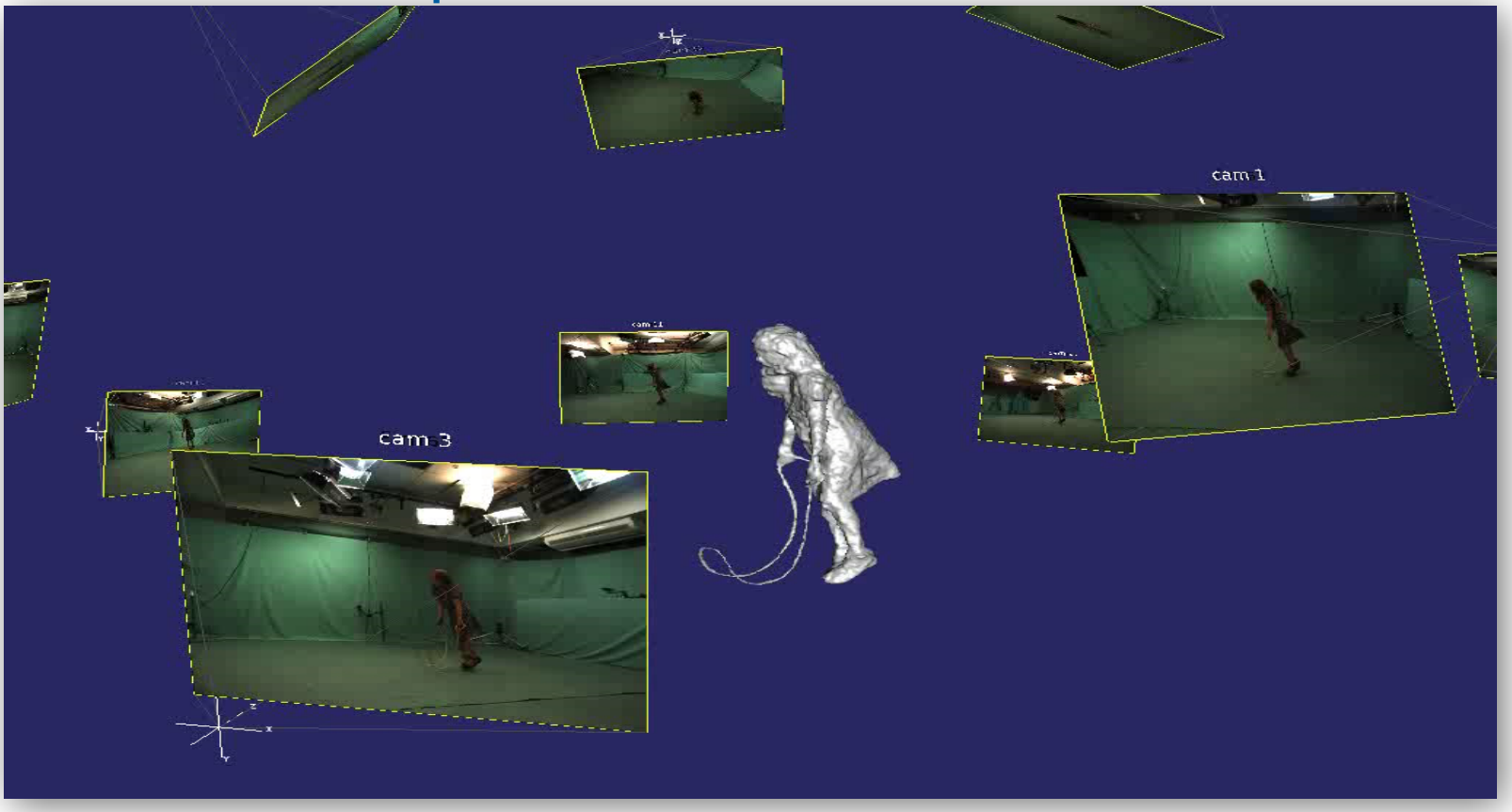
Neuroscience

Physics

Biology

Psychology

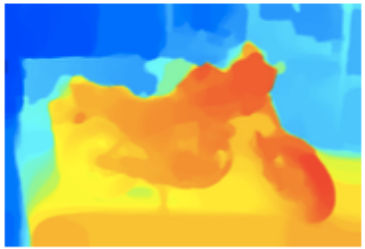
Computer Vision at TUM



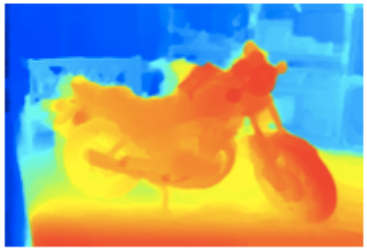
Oswald, Stühmer, Cremers., *ECCV 2014.*

Computer Vision at TUM

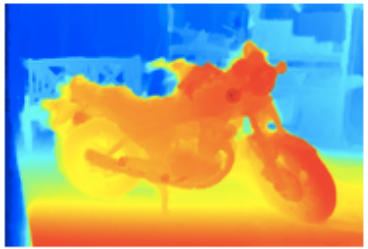
One of the input images



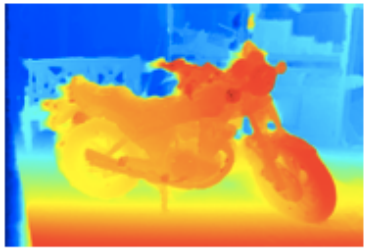
Proposed ($L = 2$)



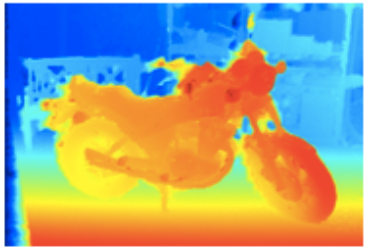
Proposed ($L = 4$)



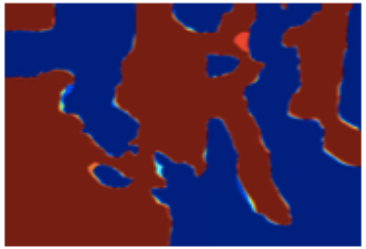
Proposed ($L = 8$)



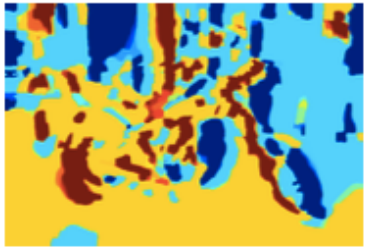
Proposed ($L = 16$)



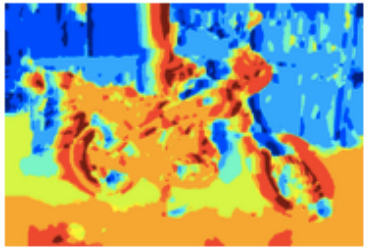
Baseline ($L = 270$)



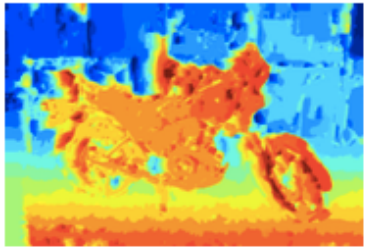
Baseline ($L = 2$)



Baseline ($L = 4$)



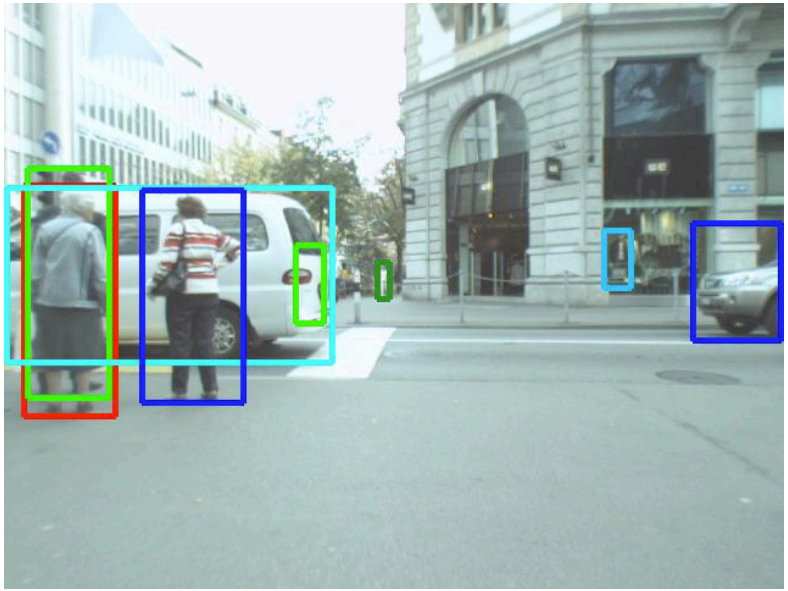
Baseline ($L = 8$)



Baseline ($L = 16$)

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

Computer Vision at TUM

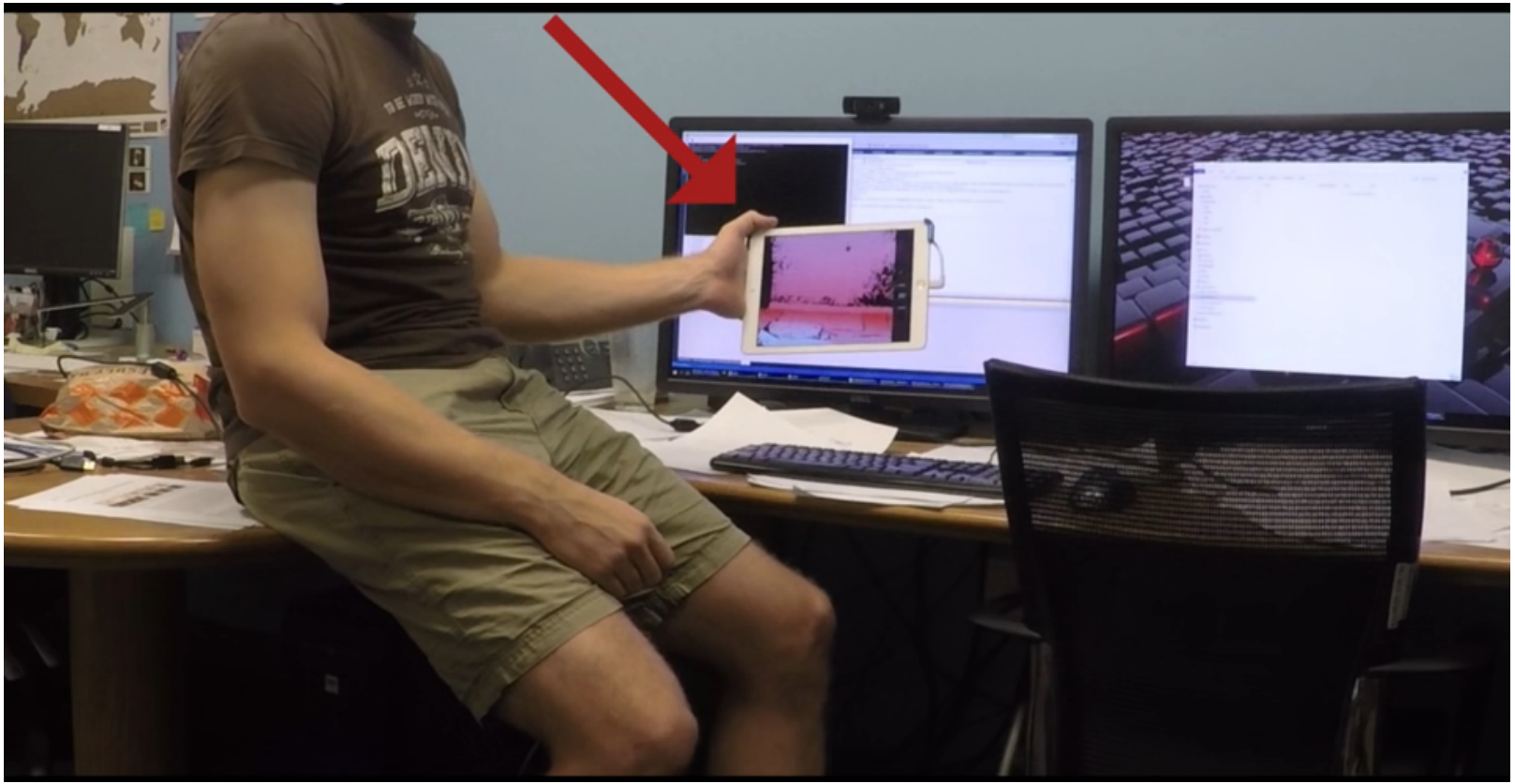


L. Leal-Taixé, A. Milan, I. Reid, S. Roth and K. Schindler. arXiv:1504.01942, 2015.

Computer Vision at TUM

Engel, Schöps, Cremers., *ECCV 2014.*

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

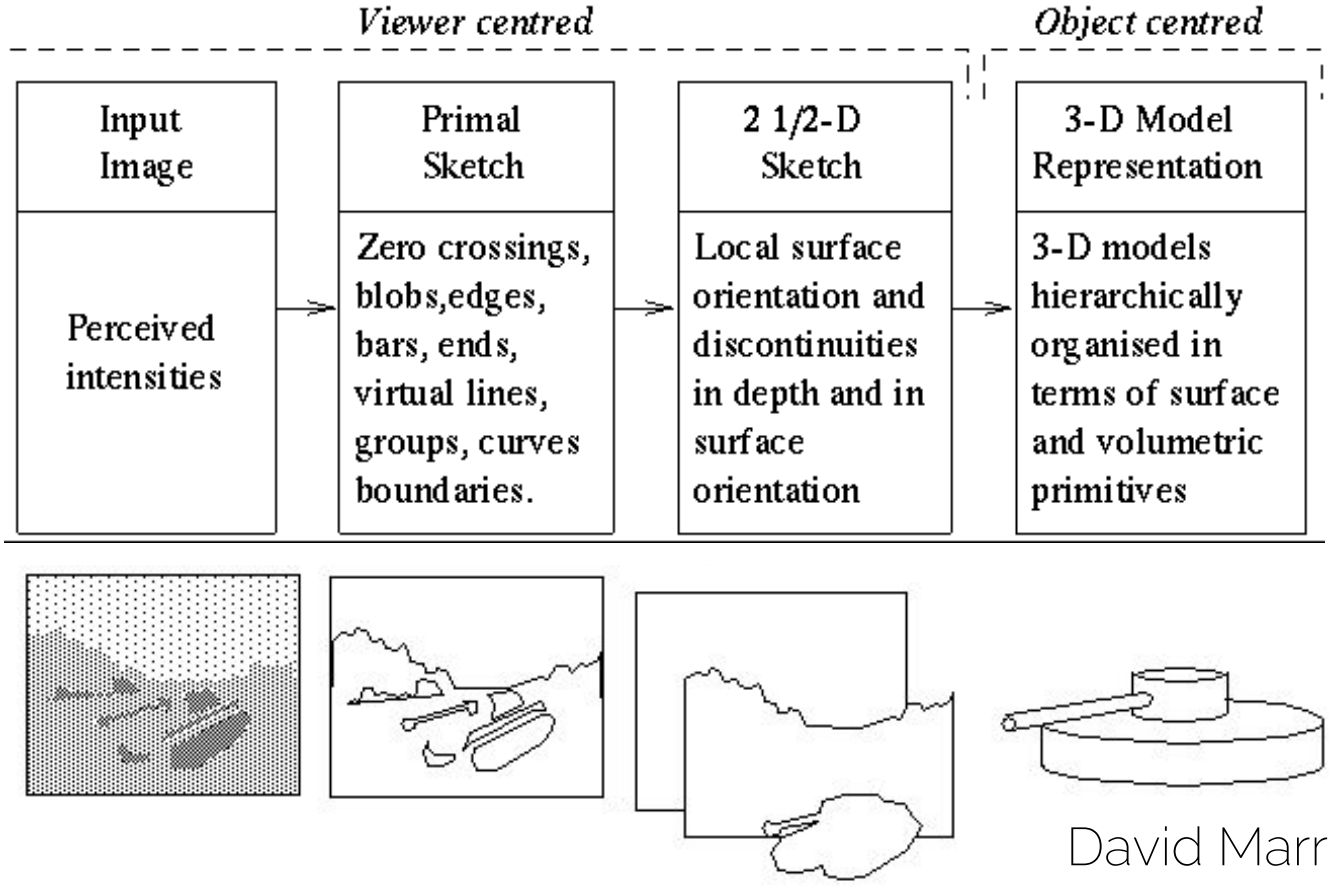
Problem solving

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

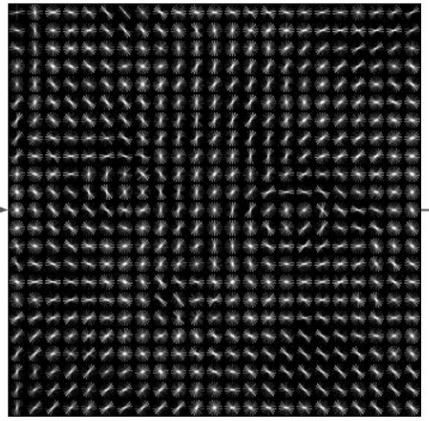


David Marr. Vision. 1982

Image classification

Input image

Preprocessing



Features : HAAR, HOG, SIFT, SURF

Learning Algorithm

SVM, Random Forests, ANN

Label Assignment

Cat or Background

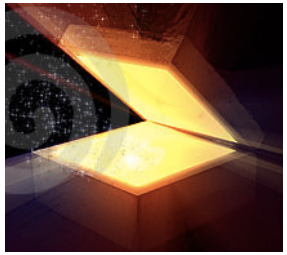
Image classification

Input image

Awesome magic box

Label Assignment

Cat or Background



Open the box

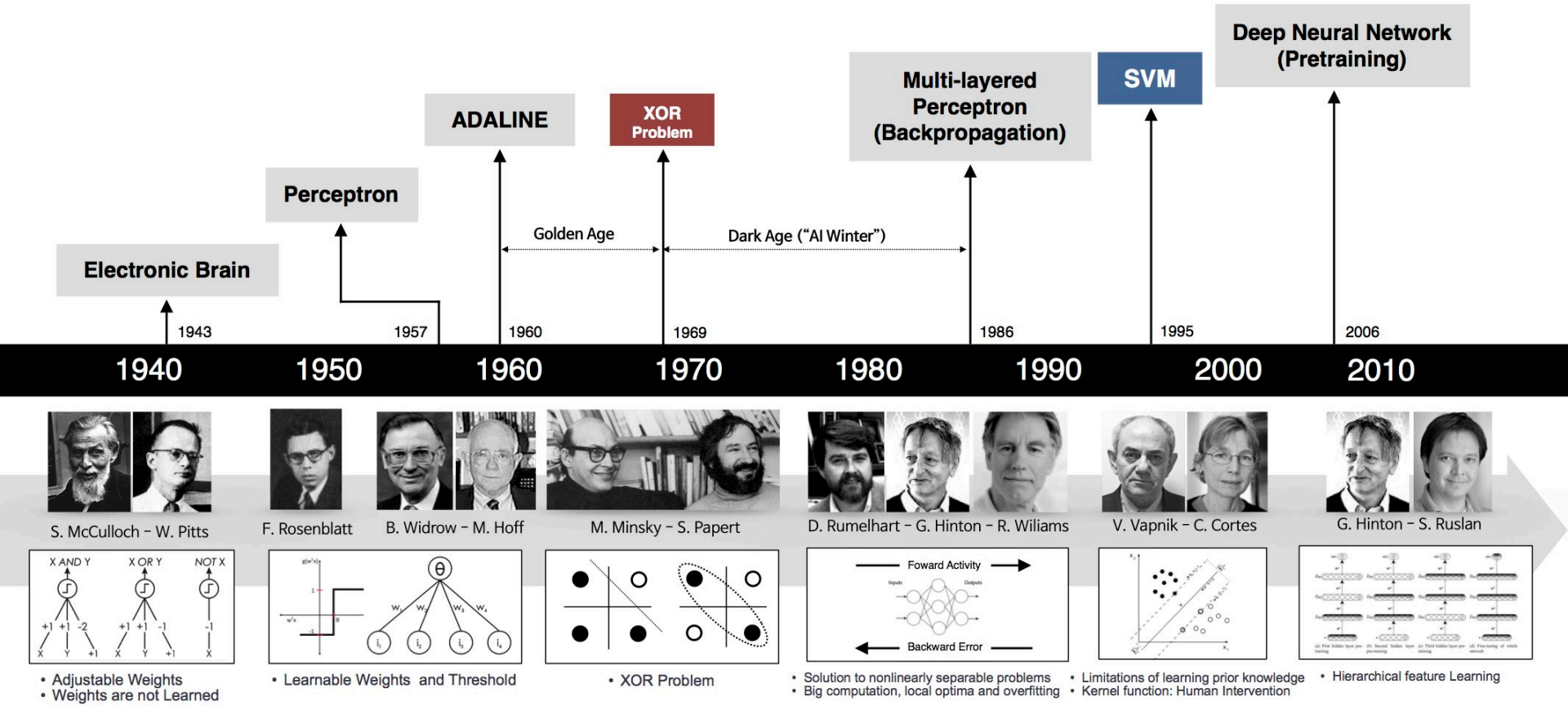
Become magicians

Post 2012

Deep Learning for Computer Vision?

... Vision?

Deep Learning History



- Adjustable Weights
- Weights are not Learned

- Learnable Weights and Threshold

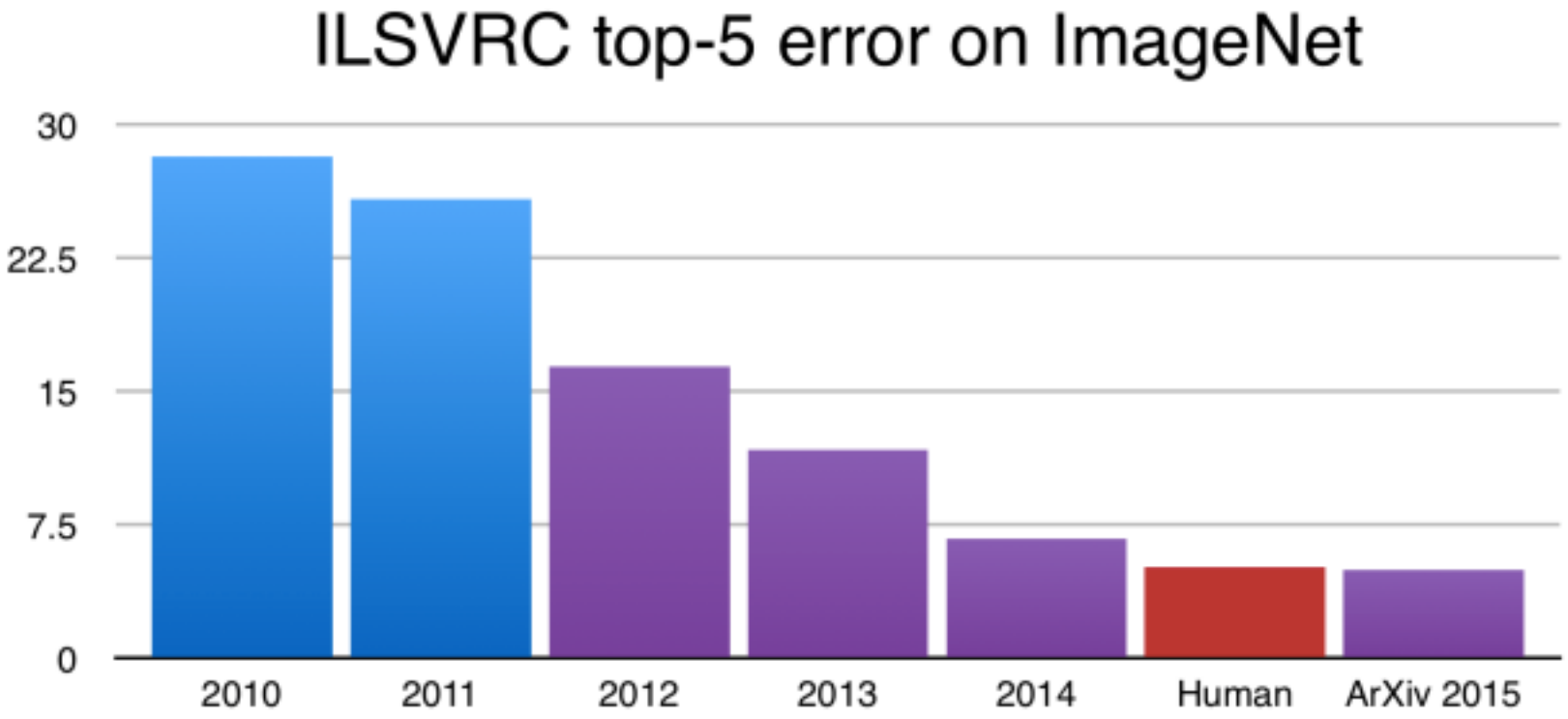
- XOR Problem

- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting

- Limitations of learning prior knowledge
- Kernel function: Human Intervention

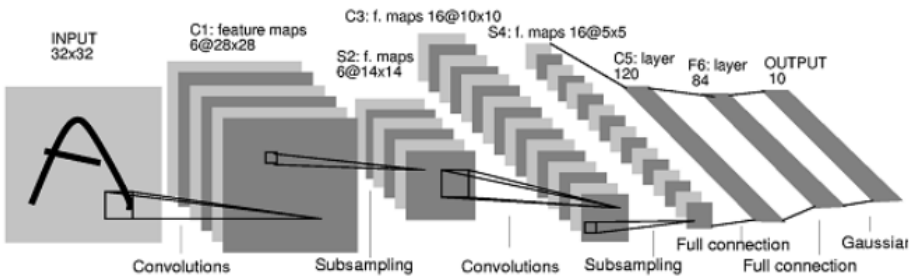
- Hierarchical feature Learning

The empire strikes back



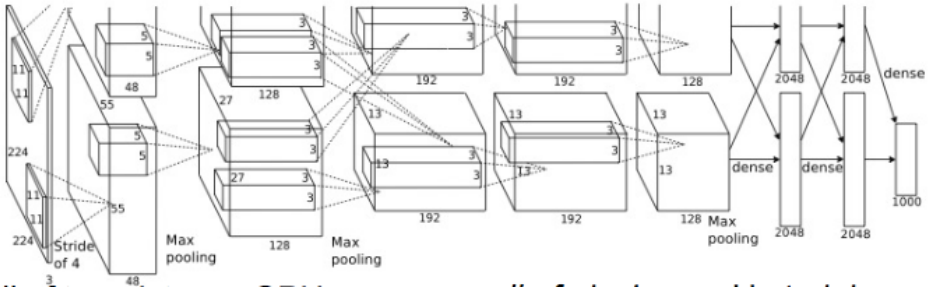
What has changed?

1988
LeCun
et al.



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

2012
Krizhevsky
et al.

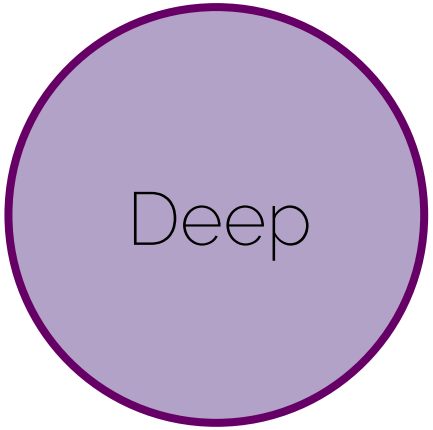


- ImageNet image recognition dataset
- 10^{14} pixels used in training

What made this possible?

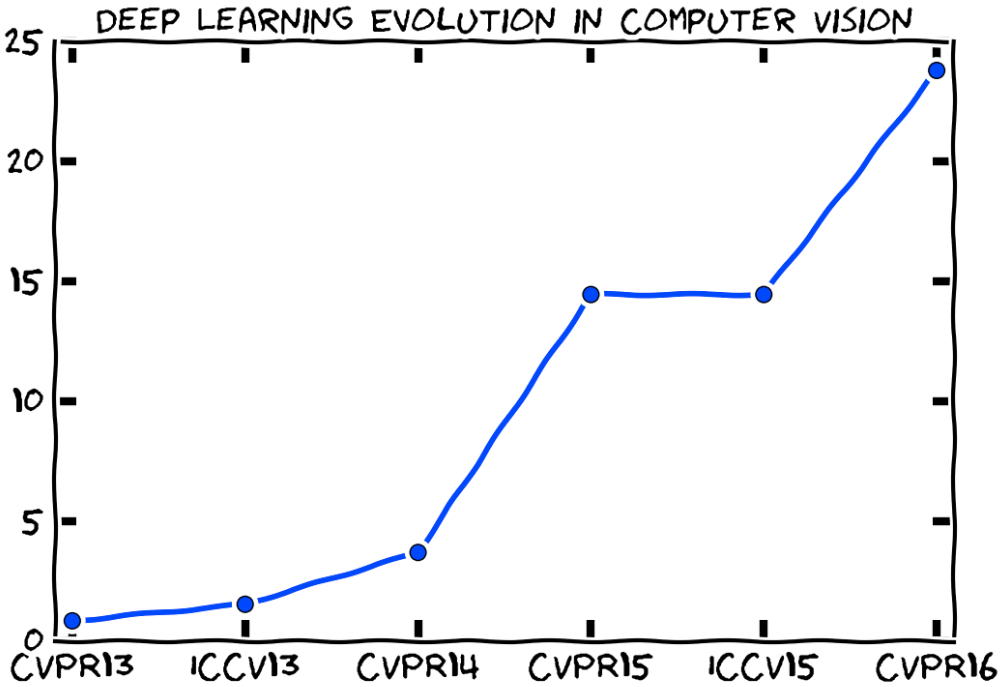
Models know where to learn from

Models are trainable



Models are complex

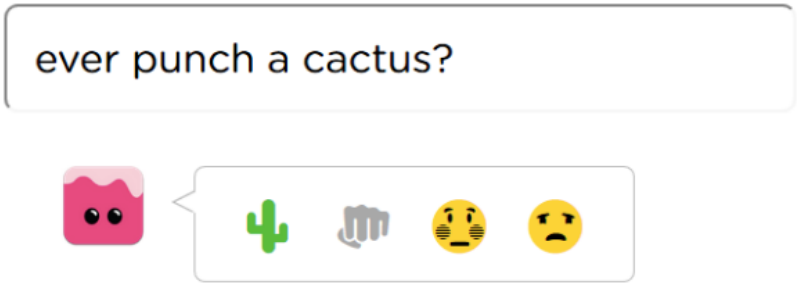
Deep Learning and Computer Vision



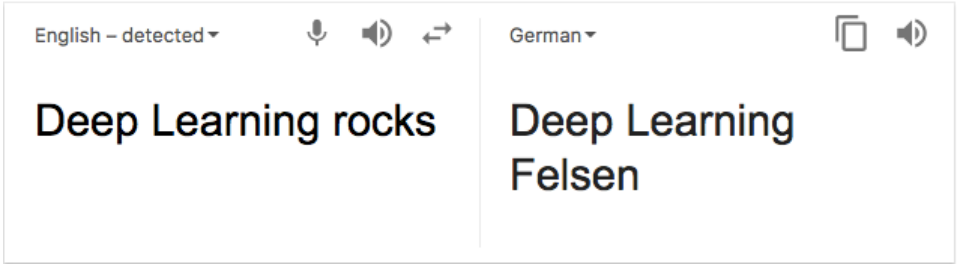
Credits: Dr. Pont-Tuset, ETH Zurich

Deep Learning nowadays

AlphaGo

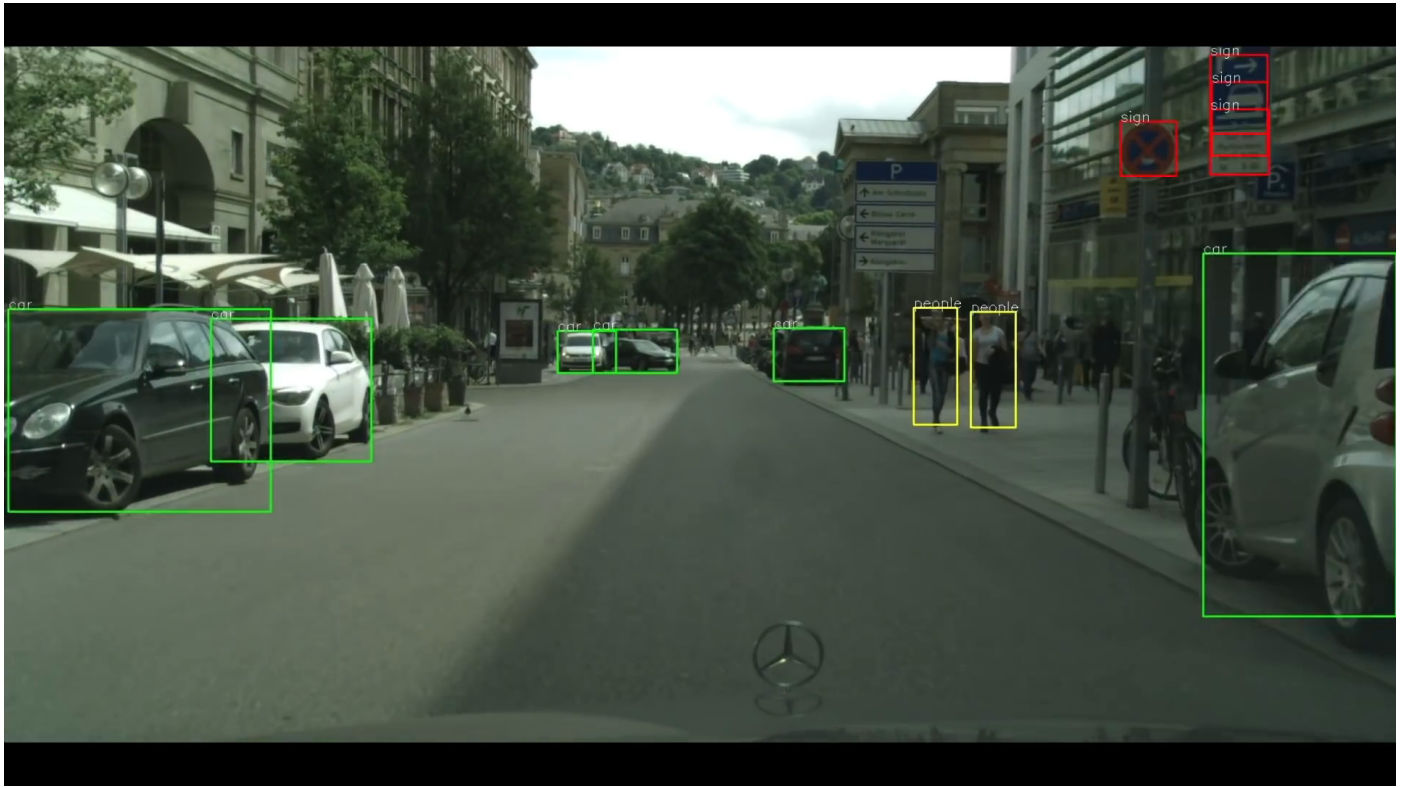


Emoticon suggestion



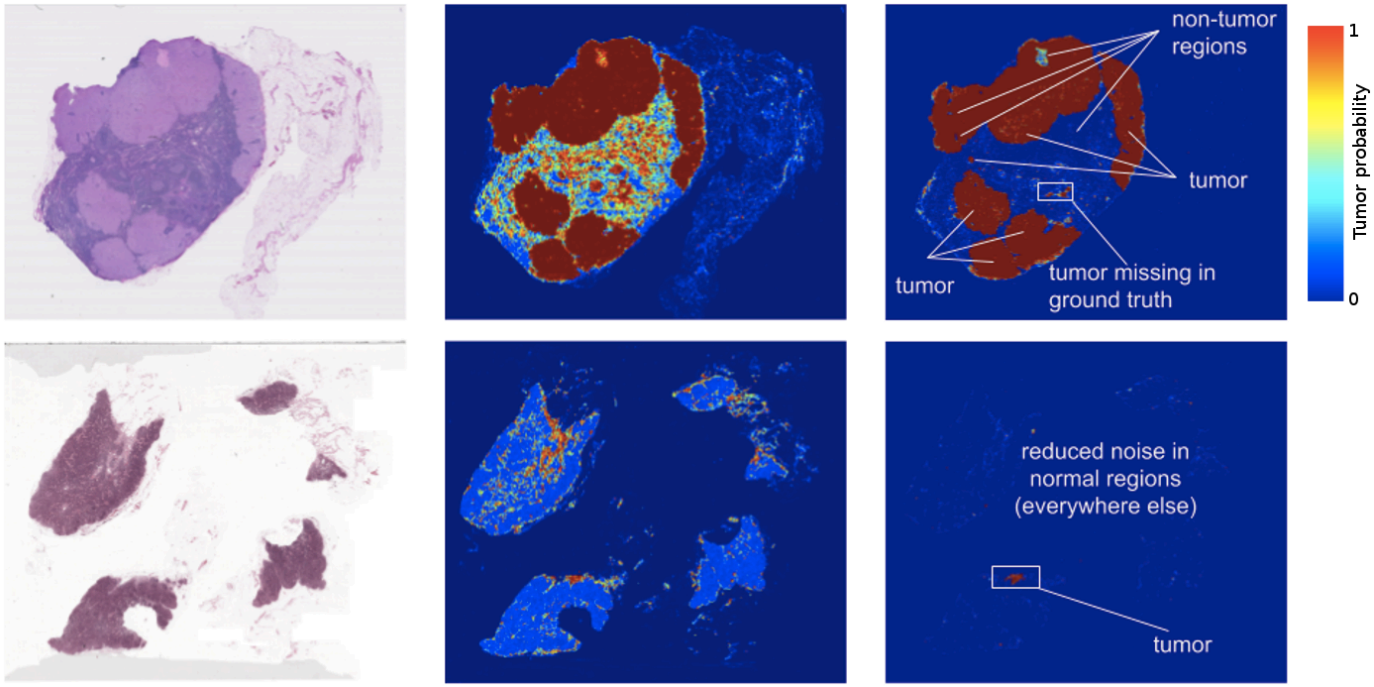
Machine translation

Deep Learning nowadays



Self-driving cars

Deep Learning nowadays



Healthcare, cancer detection

Deep Learning is everywhere

CVPR 2016 sponsors

Deep Learning is everywhere

CVPR 2016 sponsors

Deep Learning is everywhere

CVPR 2016 sponsors

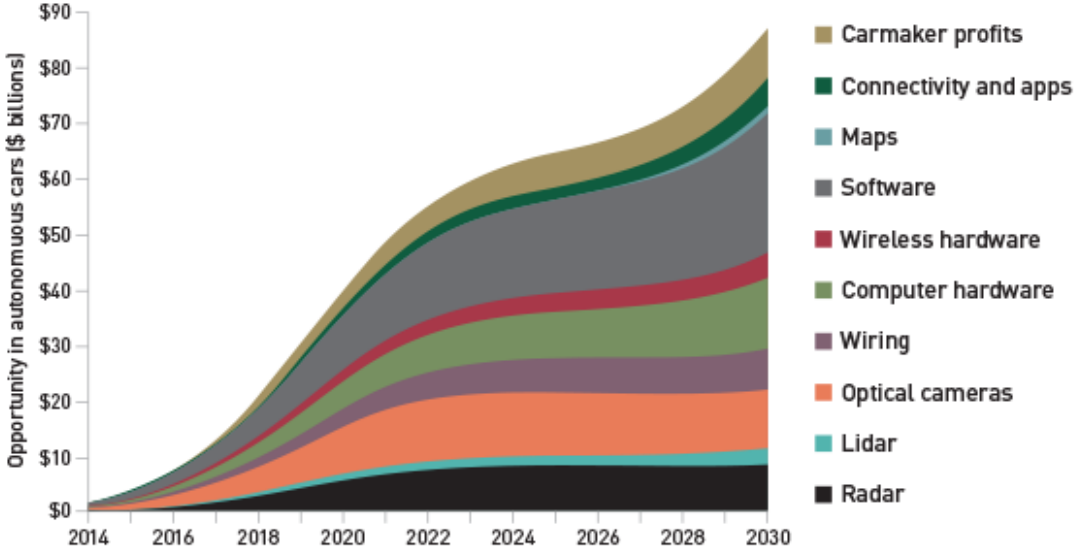
Deep Learning is everywhere

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

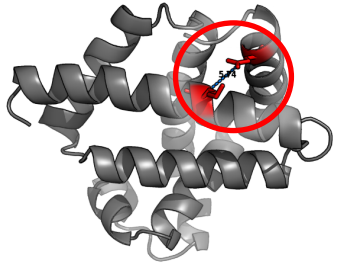
VLSEGEWQLVLHVWAK
VEADVAGHGQDILIRL
FKSHPETLEKFDRFKH
LKTEAEMKASEDLKKH
GVTVLT (Homo sapiens)

Sequence (length=L)

HMM

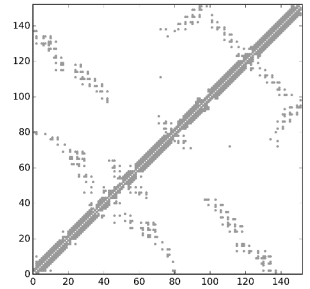
MGLSDGEWQLVLNVWG
KVEADLAGHGQDVLIR
LFKGHPETLEKFDKFK
HLKTEADMKASEDLKK
HGNTVLTALGAILKKK
(Bottlenose dolphin)

?



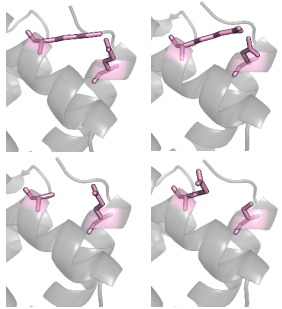
constraint

Contact map

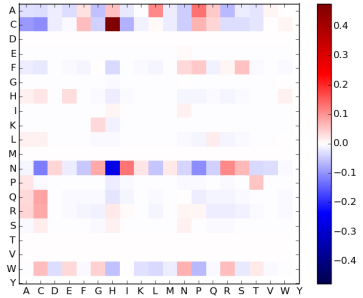


CNN

3D structure

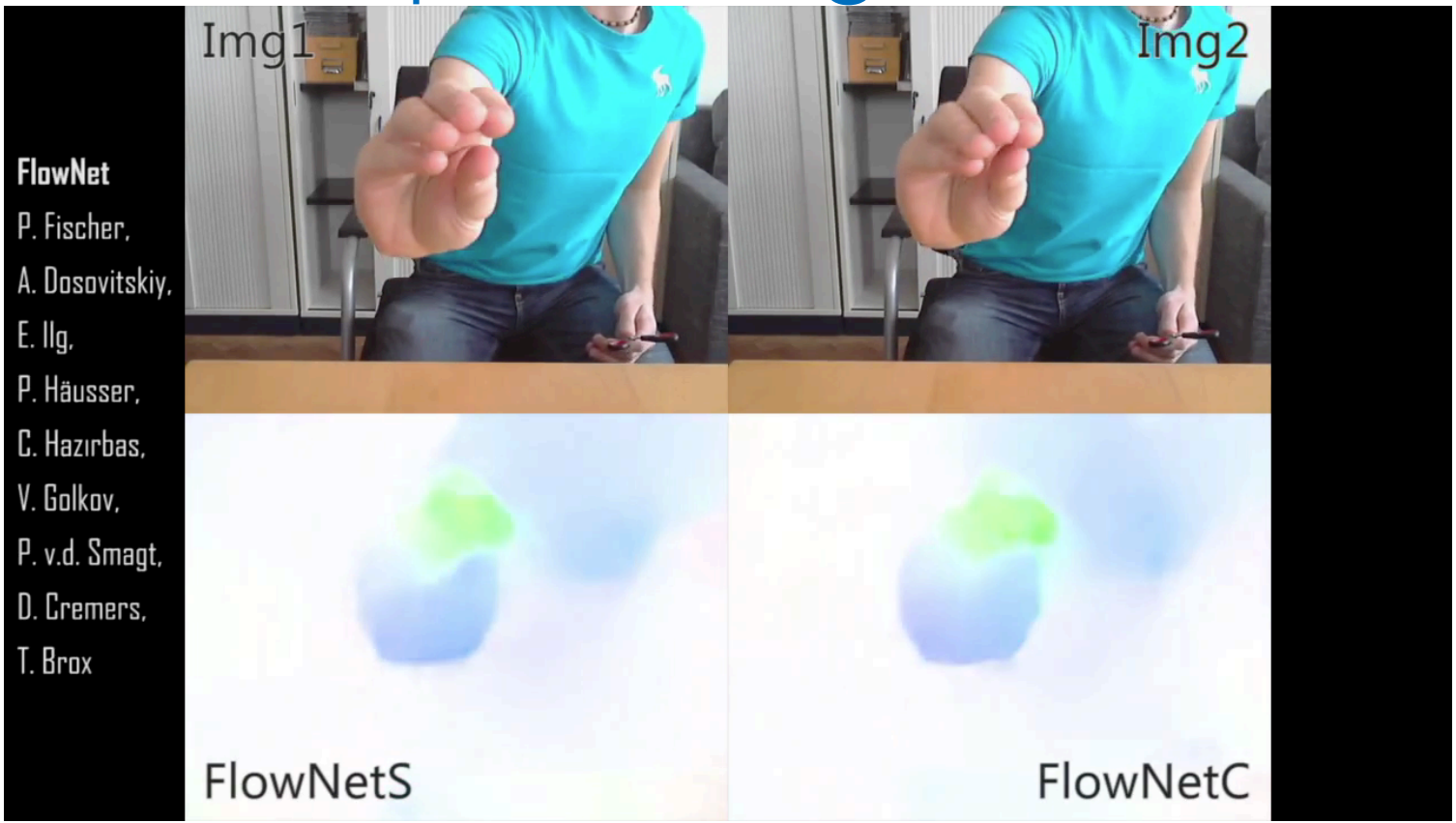


DCA (Potts)



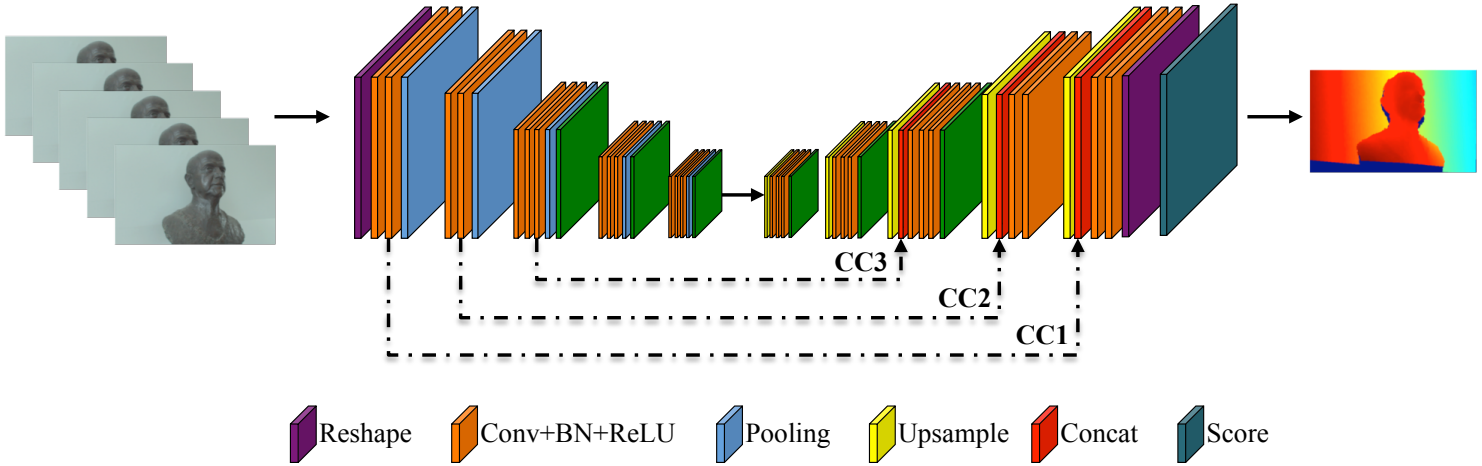
Golkov, Skwark, Golkov, Dosovitskiy, Brox, Meiler, Cremers., *NIPS 2016*

Deep Learning at TUM

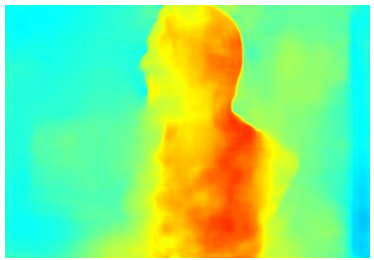


Dosovitskiy, Fischer, Ilg, Haeusser, Hazirbas, Golkov, van der Smagt, Cremers, Brox. *ICCV 15*

Deep Learning at TUM



DDFF



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

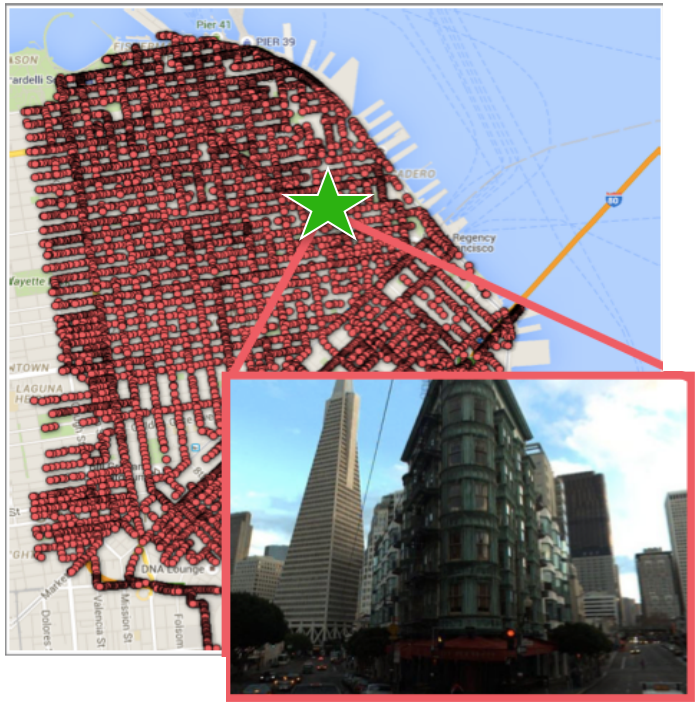
Computer Vision at TUM

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*

Deep Learning for Computer Vision

About the lecture

- Theory: 13 lectures + special lectures
  - Every Tuesday (except in two days!!)
- Practice: 3 exercises, practical sessions
  - Every Thursday
- Project: 1 final project

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

Grading system

- Theory: 13 lectures + 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/WS2017/dl4cv

Theory lecture

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

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

Slides

- Slides will be uploaded here:

https://vision.in.tum.de/teaching/ws2017/dl4cv/coursematerial

- Password: DL4CVws17

- Please DO NOT distribute!

Practical exercises

- Topics: Linear classifiers, multinomial regression, two-layer neural net.
- Begin: 02.11
- End: 15.11

- Topics: Fully connected nets, dropout, batch normalization.
- Begin: 16.11
- End: 29.11

- Topics: Convolutional neural networks, large-scale project with PyTorch.
- Begin: 30.11
- End: 13.12

Practical exercises

- Thursday (02.11): introduction of the exercise
- Next Thursday (09.11): Q&A session
- Wednesday (15.11): Delivery deadline at midnight
- Thursday (16.11): solution is discussed and new exercise is presented

FIXED DEADLINES!

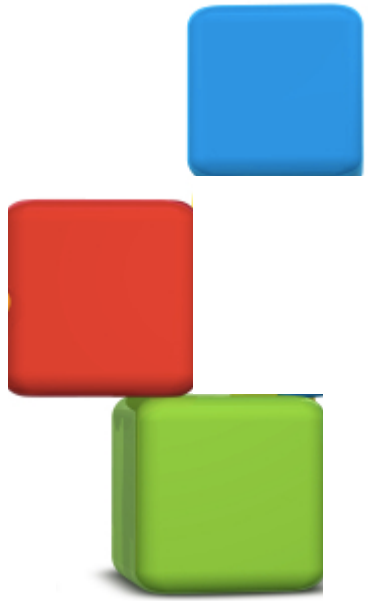
Python

- IMPORTANT: get familiar with Python!
- Exercise 0: starting 17.10 until 02.11

https://github.com/jrjohansson/scientific-python-lectures

Final project

- Introduction: 14.12
- Project proposal due date: 19.12
- Starting date: 21.12
- Midterm handout is due: tba
- Due date: tba
- Poster presentation: tba
- 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 next Thursday!

- Lecture 2 will be held on Thursday!
- When: Friday 19th of October
- Where: MI Hörsaal 1 – Math/Informatics building

The Team