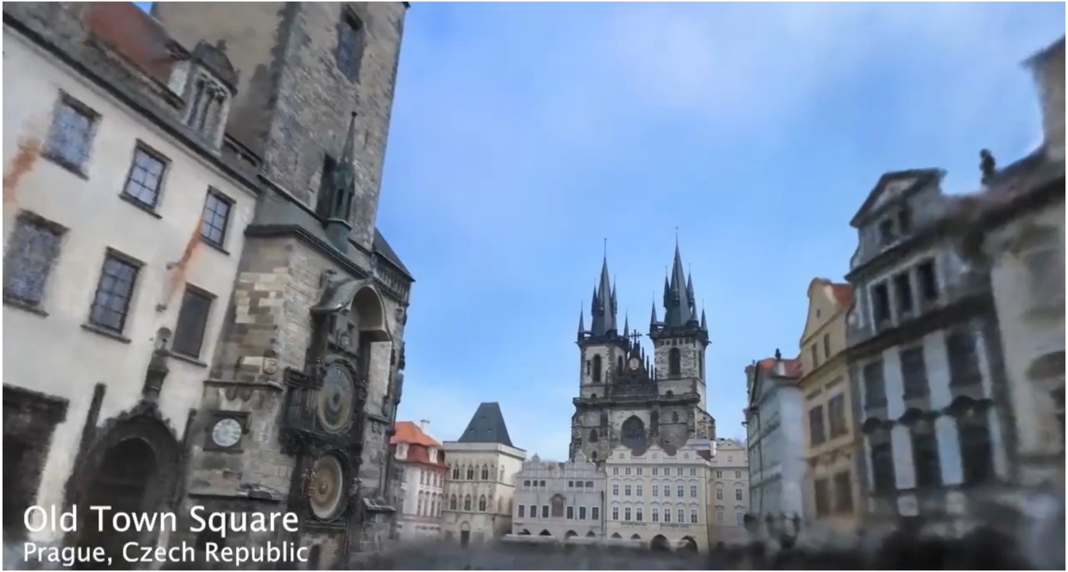
Computer Vision

Lecture 1 – Introduction

Prof. Dr.-Ing. Andreas Geiger

Autonomous Vision Group

University of Tübingen

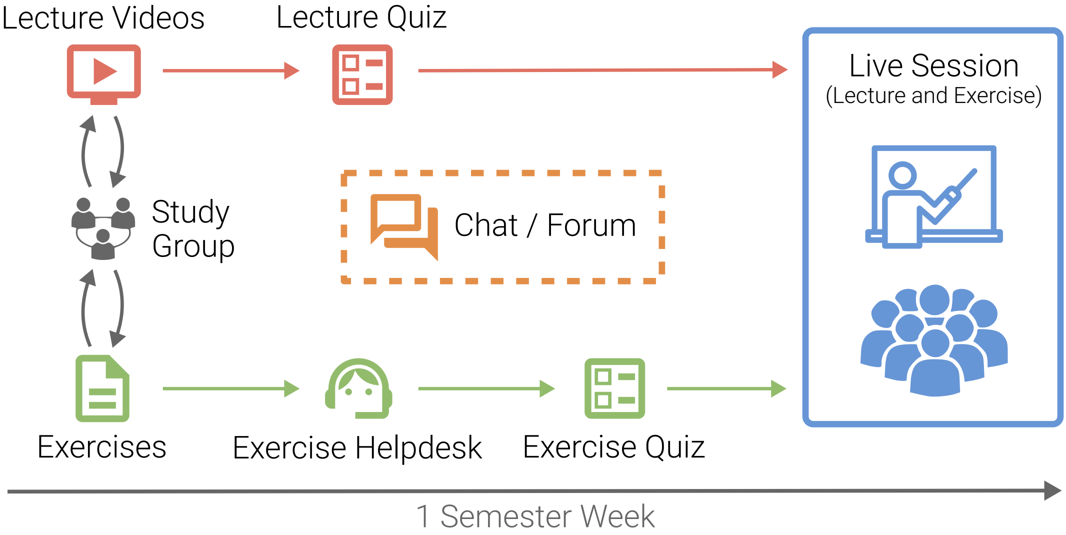


Martin-Brualla, Radwan, Sajjadi, Barron, Dosovitskiy and Duckworth: NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. CVPR, 2021.

Our Model
0.96 hours to train

- `https://cvlibs.net/mipsplatting/0579.html`
- `https://cvlibs.net/mipsplatting/0588.html`

Flipped Classroom



Agenda

1.1 Organization

1.2 Introduction

1.3 History of Computer Vision

1.1
Organization

Team

Prof. Dr.-Ing.
Andreas Geiger

Patricia
Gschoßmann

Christina
Tze

Gege
Gao

Lectures offered by our research group:

- Deep Learning (recommended for 1st semester MSc)
- Computer Vision (recommended for **2nd semester MSc**)
- Self-Driving Cars (recommended for 3rd semester MSc)

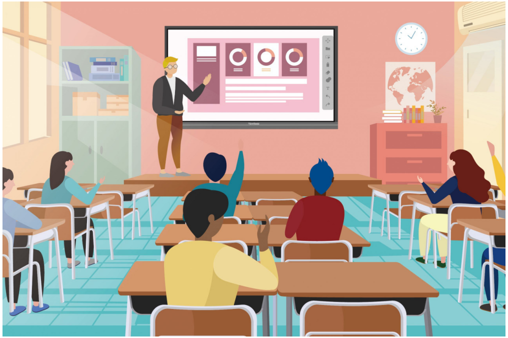
Agenda

Goal: Students gain an understanding of the theoretical and practical concepts of computer vision using deep neural networks and graphical models. A strong focus is put on 3D vision. After this course, students should be able to develop and train computer vision models, reproduce research results and conduct original research.

- History of computer vision
- Image formation
- Structure-from-Motion
- Stereo Reconstruction
- Probabilistic Graphical Models
- Applications of Graphical Models
- Learning in Graphical Models
- Shape-from-X
- Coordinate-based Networks
- Recognition
- Self-Supervised Learning
- Diverse Topics in Computer Vision

Flipped Classroom

Flipped Classroom



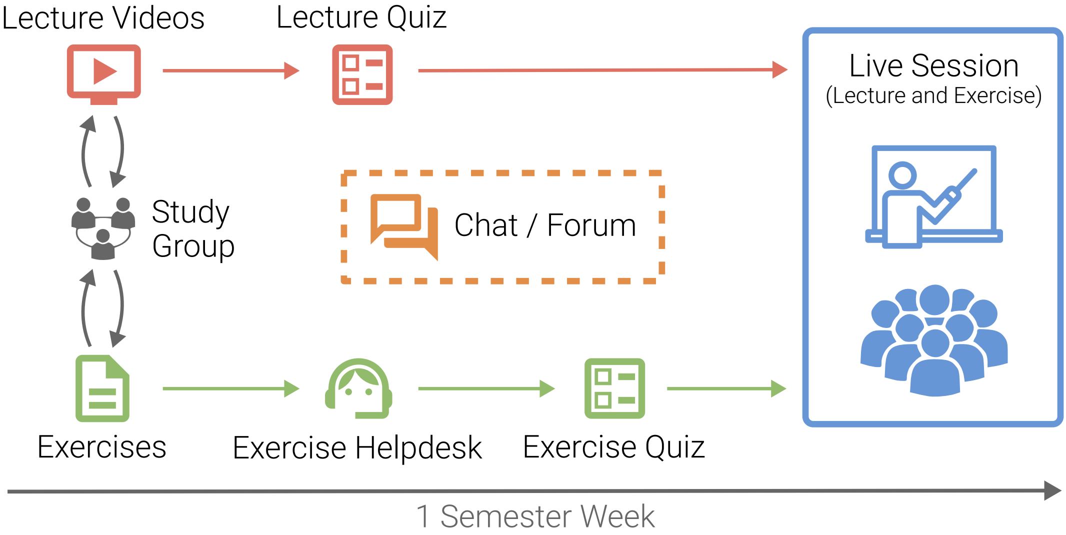
Regular Classroom

- Lecturer “reads” lecture in class
- Students digest content at home
- Little “quality time”, no fun

Flipped Classroom

- Students watch lectures at home
- Content discussed during lecture
- Maximize interaction time

Flipped Classroom



Organization

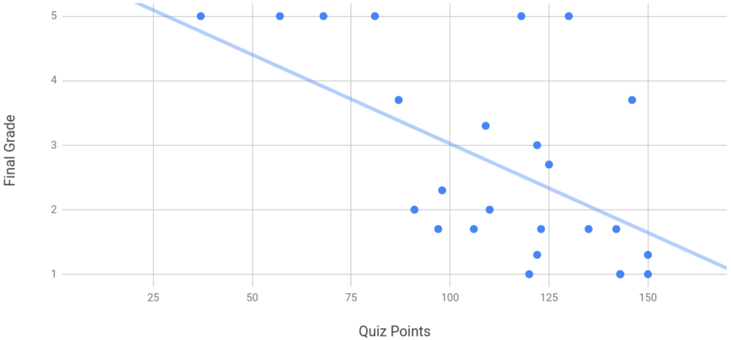
Organization

- **Lecture videos** provided in advance via the course website
  - Watch lecture before live session! (mark your calendar!)
  - Ask (many) questions via the quiz server chat room (dare to ask!)
- **Live sessions** (Wednesdays)
  - Lecture Q&A and examples, exercise Q&A and task solving
  - Will not be hybrid and or recorded
- **No tutorials** in small groups, instead we provide:
  - Practice via **exercises** with quizzes and solutions (solve them!)
  - Plenary feedback via **live sessions** (join us!)
  - Individual help via weekly **helpdesk** (make use of it!)
  - Fast ($<$24h) feedback via **chat** (ask and reply!)
- **Exam:** Written (bonus upon quiz participation)
- **Course Website:** `https://uni-tuebingen.de/de/203146`
  - **Register** on Lecture Quiz Server today! (no ILIAS for this course)

Exercises

- Exercises are offered every 2 weeks, **6 assignments** in total
- Exercises deepen the understanding with **pen & paper and coding tasks**
- Solving the exercises is not a prerequisite to participate in the exam
- Exercise tasks are **highly relevant** to the exam $\Rightarrow$ use this opportunity

Slides / Exercises

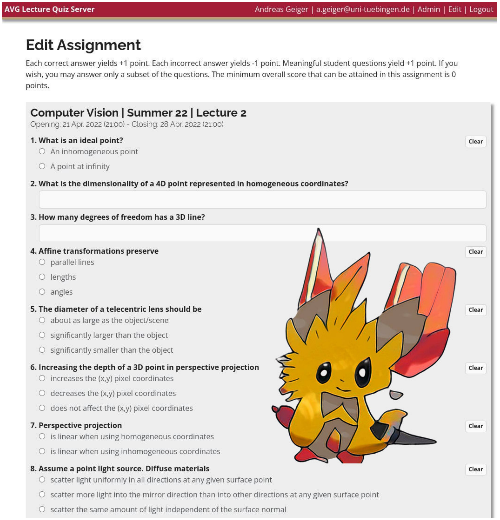


Exercises

- **Problems** handed out, introduced and discussed in the live sessions
- Can be conducted in groups of **up to 4 students**
- We offer a **helpdesk** every week to answer your questions
- At the end of each exercise, we will hand out and discuss the **solutions**
- But how do I stay motivated to watch the lectures and solve the exercises?

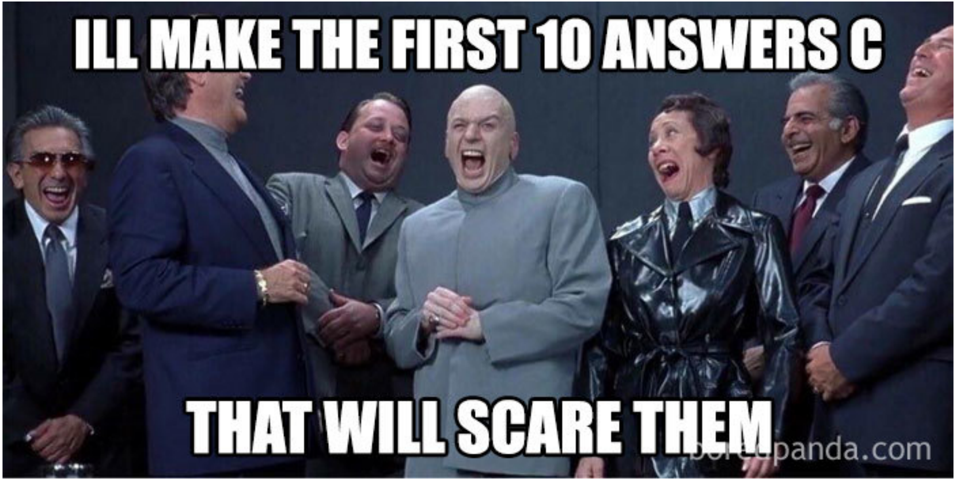
AVG Lecture Quiz Server

- We provide **quiz questions** to students during lectures 2-12 and exercises 1-6
- Collect exam **bonus points**
- Collect AI generated **Pokemon**
- Answers may **not** be **shared**
- Participation is **voluntary**
- Past quizzes only **visible to participants**
- Deadline: **Mondays, 9pm**
- **Register today**



https://uni-tuebingen.de/de/203146

AVG Lecture Quiz Server: Quiz Questions



- Ask (via chat) if a **question is ambiguous** – after the deadline it is too late!

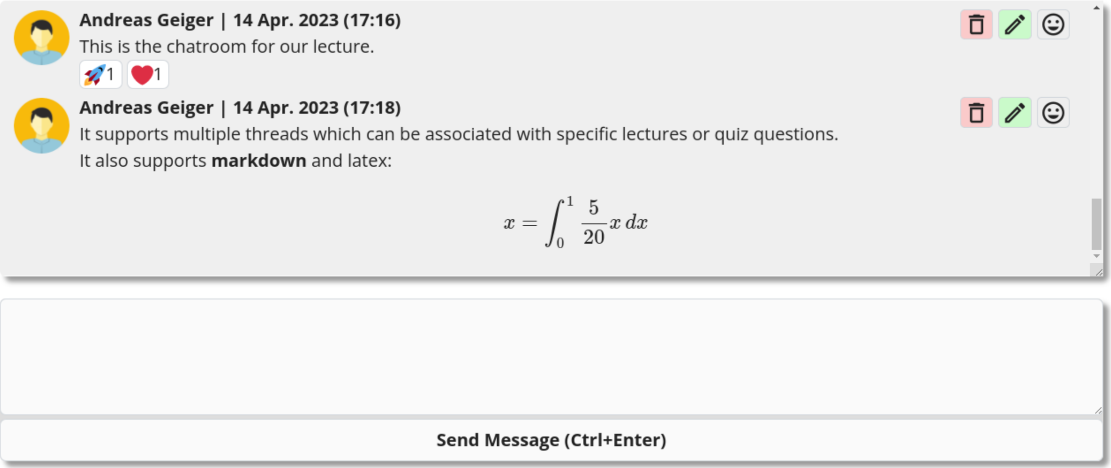
AVG Lecture Quiz Server: ChatGPT

- **Spending time** on the exercises is the **most effective** way to prepare for the exam
- Using **ChatGPT** or solutions from previous years will effectively **harm you**
- ChatGPT will not help you in the exam: you need to deeply understand the topic
- **Simple rule to success:** Don't use ChatGPT, don't copy solution
- The main goal of our quizzes is to **motivate you** to engage with the materials

Exercise Helpdesk

- We offer a **weekly helpdesk** where our TAs provide individual support:
  - Day?
  - Time?
- **Just show up** (no registration)
- Ask **any question** about the exercise
- Share your screen to show a problem
- Make use of this opportunity
- Start working on your exercise **early**

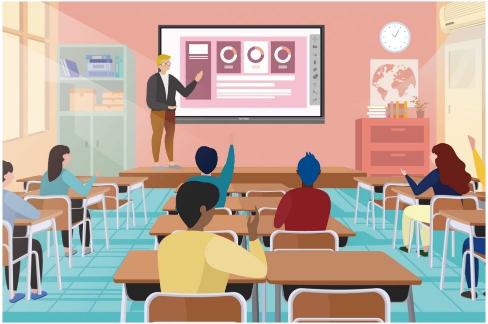
Chat



- **Ask** (many) questions! **Answer** questions if you think you know the answer!

Live Sessions

1. **Lecture Q&A** (70 min)
   - Live lecture and example walk throughs
   - Discussion of quiz questions
   - Discussion of student questions
2. Break (15 min)
3. **Exercise Q&A** (70 min)
   - Exercise introduction
   - Exercise live solving and Q&A
   - Start solving before joining!



What will the exam look like?

- **Written** main and make-up exam
- You may **choose freely** (but no 3rd exam)
- Registration via Quiz Server
- Only **pen and ruler** allowed (no notes)
- Duration: 90 minutes (can be solved in 60)
- 5 Tasks, 10 points each, 25 points will pass
- Bonus added only to passed exams
- Tasks cover both **lectures** and **exercises**
- Mix of knowledge, calculation, multiple choice
- **Old exams** available in shared folder

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Science
Department of Computer Science
Prof. Dr. Andreas Geiger
Maria-von-Linden-Straße 6
72076 Tübingen

Computer Vision

Exam, Summer 2021

Tübingen, August 10, 2021

Student Number:
Seat Number:
First Name:
Last Name:
Date of Birth:
Program of Study (e.g., "Informatik"):
Master ☐ Bachelor ☐

Task	1	2	3	4	5	$\sum$
Maximal Score	10	10	10	10	10	50
Attained Score						
Correction						

Remarks:

- Leave your bag and jacket at the front desk. Only take your **pen, ruler, student ID** with you.
- Notes, books, printouts, smartphones or calculators are not permitted.
- Have your student ID ready for inspection.
- Write your full name at the top of **every** page.
- Use the space provided for your answers. You can use the back side if you need more space.
- Additional paper can be obtained at the front desk.
- Use black or blue pen. Pencil or red pens are not allowed and will lead to a score of 0.
- A total score of **25** is sufficient for passing the exam. The duration of the exam is **90 minutes**.

Good Luck!

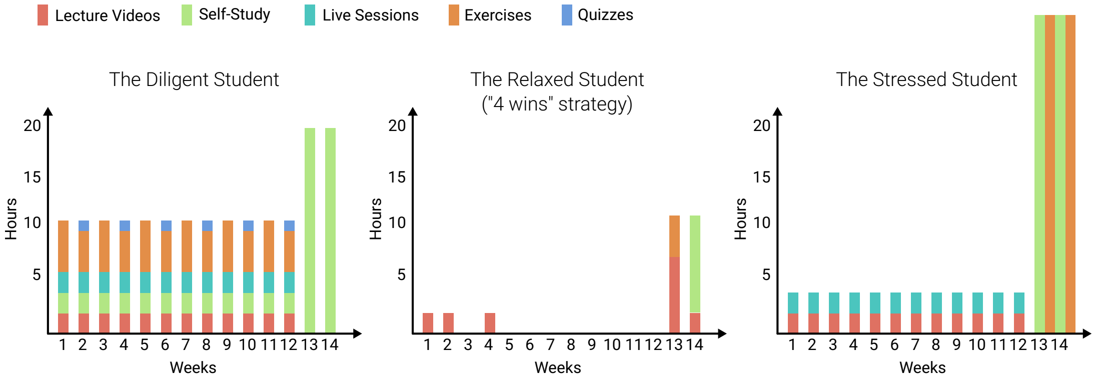
More information: `https://uni-tuebingen.de/de/171060`

Bonus Points

- **11 lecture quizzes** (10 points per quiz)
- **6 exercise quizzes** (10 points per quiz)
- 110 lec. points + 60 ex. points in total
- Bonus Points $= \left\lfloor \frac{\text{lec. points}}{36} \right\rceil + \left\lfloor \frac{\text{ex. points}}{20} \right\rceil$
- Maximum: 3 + 3 = 6 bonus points
- Bonus added only to **passed exams**

Work Ethics

This lecture has **9 ECTS**, corresponding to a total workload of **270 hours** (MHB)



Course Materials and Prerequisites

Course Materials

Books:

- Szeliski: Computer Vision: Algorithms and Applications
  `https://szeliski.org/Book/`
- Hartley and Zisserman: Multiple View Geometry in Computer Vision
  `https://www.robots.ox.ac.uk/~vgg/hzbook/`
- Nowozin and Lampert: Structured Learning and Prediction in Computer Vision
  `https://pub.ist.ac.at/~chl/papers/nowozin-fnt2011.pdf`
- Goodfellow, Bengio, Courville: Deep Learning
  `http://www.deeplearningbook.org`
- Deisenroth, Faisal, Ong: Mathematics for Machine Learning
  `https://mml-book.github.io`
- Inoffical lecture notes written by students in summer 2021
  `https://uni-tuebingen.de/de/203146`

Course Materials

Tutorials:

- The Python Tutorial
  `https://docs.python.org/3/tutorial/`
- NumPy Quickstart
  `https://numpy.org/devdocs/user/quickstart.html`
- PyTorch Tutorial
  `https://pytorch.org/tutorials/`

Frameworks / IDEs:

- Visual Studio Code
  `https://code.visualstudio.com/`
- Google Colab
  `https://colab.research.google.com`

Course Materials

Related Courses:

- Gkioulekas (CMU): Computer Vision
  `http://www.cs.cmu.edu/~16385/`
- Owens (University of Michigan): Foundations of Computer Vision
  `https://web.eecs.umich.edu/~ahowens/eecs504/w20/`
- Lazebnik (UIUC): Computer Vision
  `https://slazebni.cs.illinois.edu/spring19/`
- Freeman and Isola (MIT): Advances in Computer Vision
  `http://6.869.csail.mit.edu/sp21/`
- Seitz (University of Washington): Computer Vision
  `https://courses.cs.washington.edu/courses/cse576/20sp/`
- Serge Belongie (UCSD): Computer Vision II
  `https://cseweb.ucsd.edu/classes/sp04/cse252b/`
- Slide Decks covering Szeliski Book
  `http://szeliski.org/Book/`

Prerequisites

Math:

- Linear algebra, probability and information theory. If unsure, have a look at:
  Goodfellow et al.: Deep Learning (Book), Chapters 1-4
  Luxburg: Mathematics for Machine Learning (Lecture)
  Deisenroth et al.: Mathematics for Machine Learning (Book)

Computer Science:

- Variables, functions, loops, classes, algorithms

Deep Learning:

- Geiger: Deep Learning (Lecture)

Python and PyTorch:

- `https://docs.python.org/3/tutorial/`
- `https://pytorch.org/tutorials/`

Prerequisites

Linear Algebra:

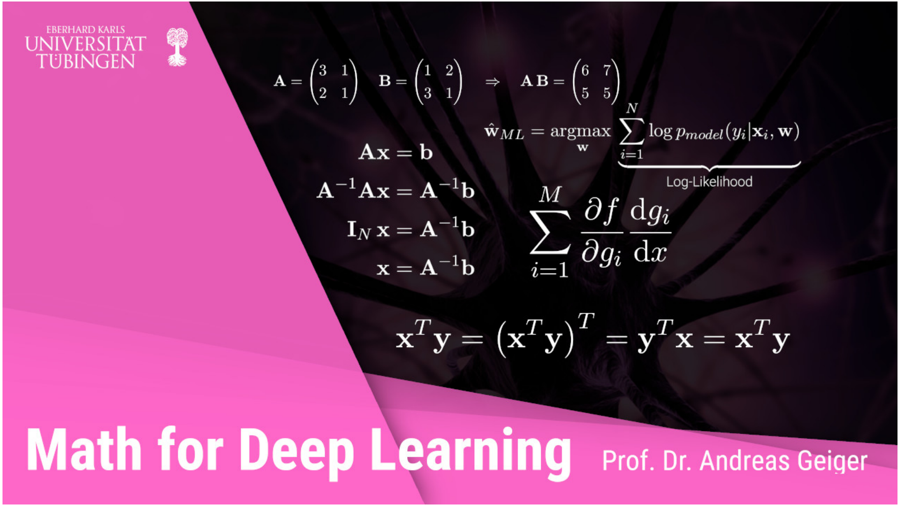
- Vectors: $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- Matrices: $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$
- Operations: $\mathbf{A}^T$, $\mathbf{A}^{-1}$, $\mathrm{Tr}(\mathbf{A})$, $\det(\mathbf{A})$, $\mathbf{A} + \mathbf{B}$, $\mathbf{AB}$, $\mathbf{Ax}$, $\mathbf{x}^\top \mathbf{y}$
- Norms: $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty$, $\|\mathbf{A}\|_F$
- SVD: $\mathbf{A} = \mathbf{UDV}^\top$

Prerequisites

Probability and Information Theory:

- Probability distributions: $P(X = x)$
- Marginal/conditional: $p(x) = \int p(x,y)dy$, $p(x,y) = p(x|y)p(y)$
- Bayes rule: $p(x|y) = p(y|x)p(x)/p(y)$
- Conditional independence: $x \perp\!\!\!\perp y \mid z \Leftrightarrow p(x,y|z) = p(x|z)p(y|z)$
- Expectation: $\mathbb{E}_{x\sim p}\left[f(x)\right] = \int_x p(x)f(x)dx$
- Variance: $\mathrm{Var}(f(x)) = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right]$
- Distributions: Bernoulli, Categorical, Gaussian, Laplace
- Entropy: $H(x)$, KL Divergence: $D_{KL}(p\|q)$

Prerequisites



`https://uni-tuebingen.de/de/241678`

Prerequisites

Deep Learning:

- Machine learning basics, linear and logistic regression
- Computation graphs, backpropagation algorithm
- Activation and loss functions, initialization
- Regularization and optimization of deep neural networks
- Convolutional neural networks
- Recurrent neural networks
- Graph neural networks
- Autoencoders and generative adversarial networks

Thank You!

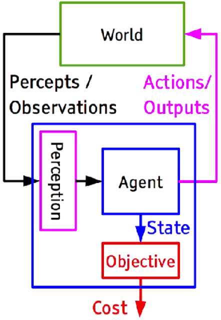
Looking forward to our discussions

1.2
Introduction

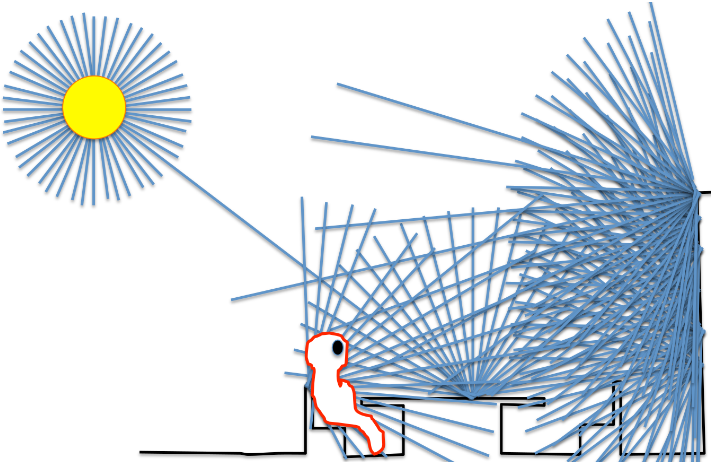
Artificial Intelligence

"An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves." [John McCarthy]

- Machine Learning
- Computer Vision
- Computer Graphics
- Natural Language Processing
- Robotics & Control
- Art, Industry 4.0, Education ...



Computer Vision

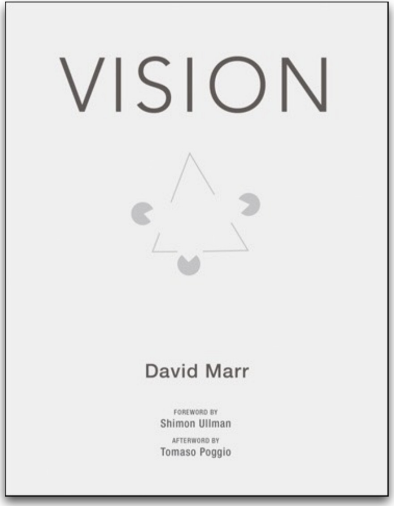


Goal of Computer Vision is to **convert light into meaning** (geometric, semantic, . . .)

Image Credits: Antonio Torralba

Computer Vision

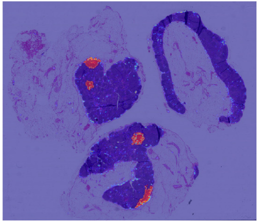
- What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking.
- To discover from images what is present in the world, where things are, what actions are taking place, to predict and anticipate events in the world.



Computer Vision Applications

Robotics

Medical applications



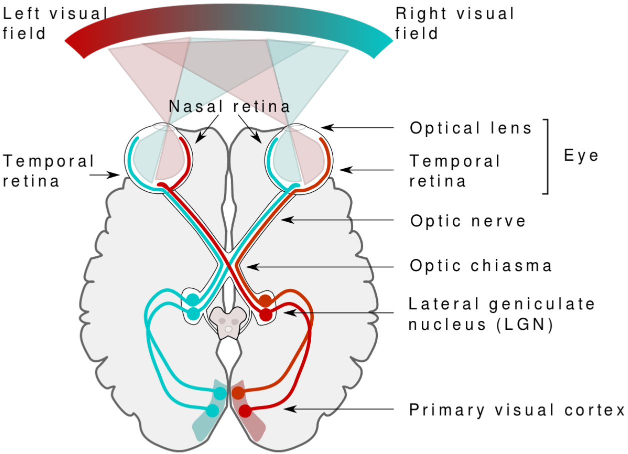
3D modeling

Driving

Mobile devices

Accessibility

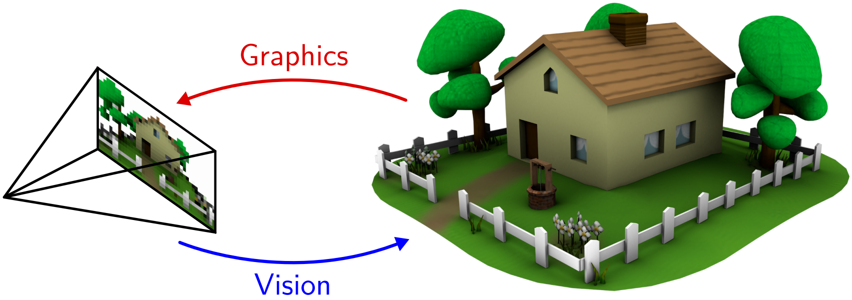
Computer Vision vs. Biological Vision



Over **50%** of the processing in the **human brain** is dedicated to **visual information.**

Computer Vision vs. Computer Graphics

2D Image **3D Scene**



Pixel Matrix

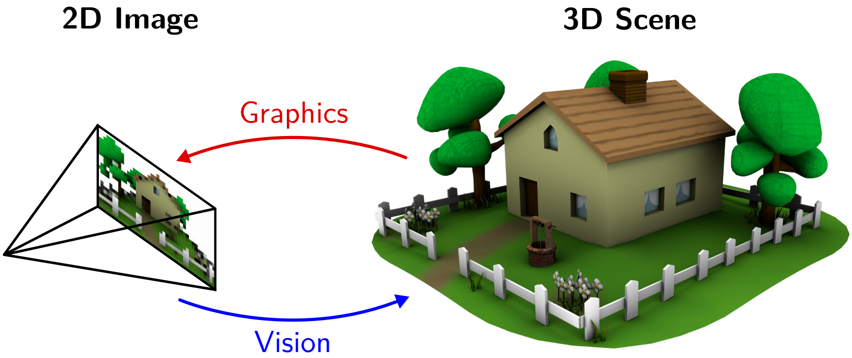
217	191	252	255	239
102	80	200	146	138
159	94	91	121	138
179	106	136	85	41
115	129	83	112	67
94	114	105	111	89

Objects Material

Shape/Geometry Motion

Semantics 3D Pose

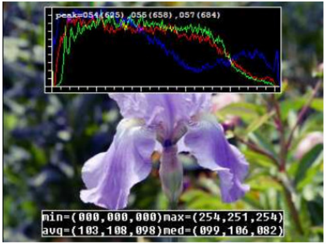
Computer Vision vs. Computer Graphics

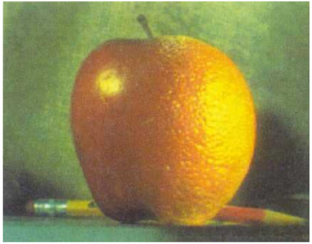


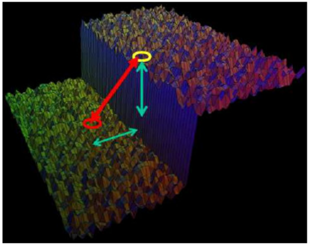
Computer Vision is an ill-posed inverse problem:

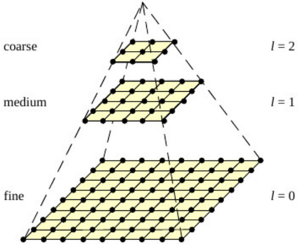
- Many 3D scenes yield the same 2D image
- Additional constraints (knowledge about world) required

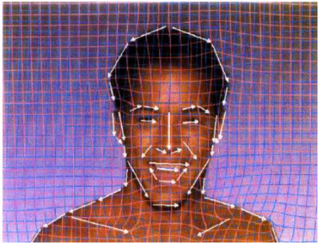
Computer Vision vs. Image Processing







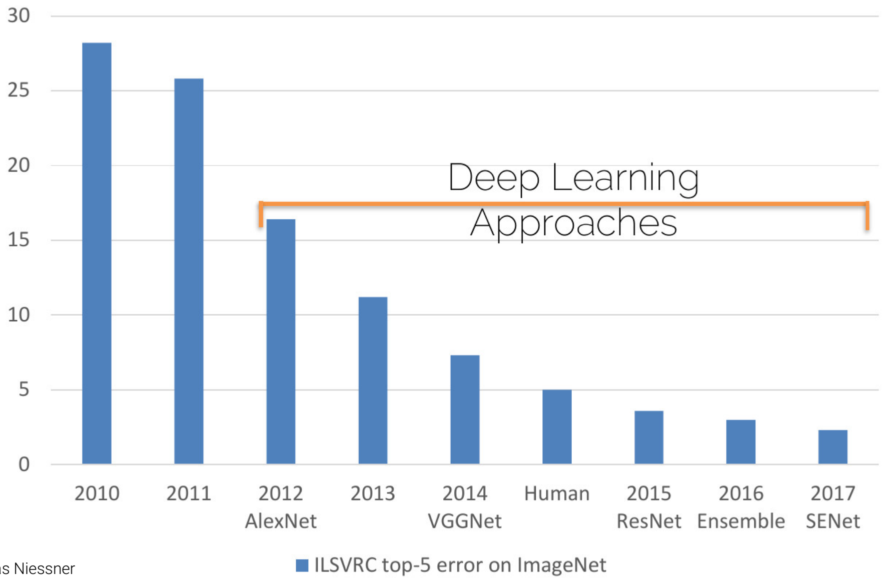




Computer Vision vs. Machine Learning

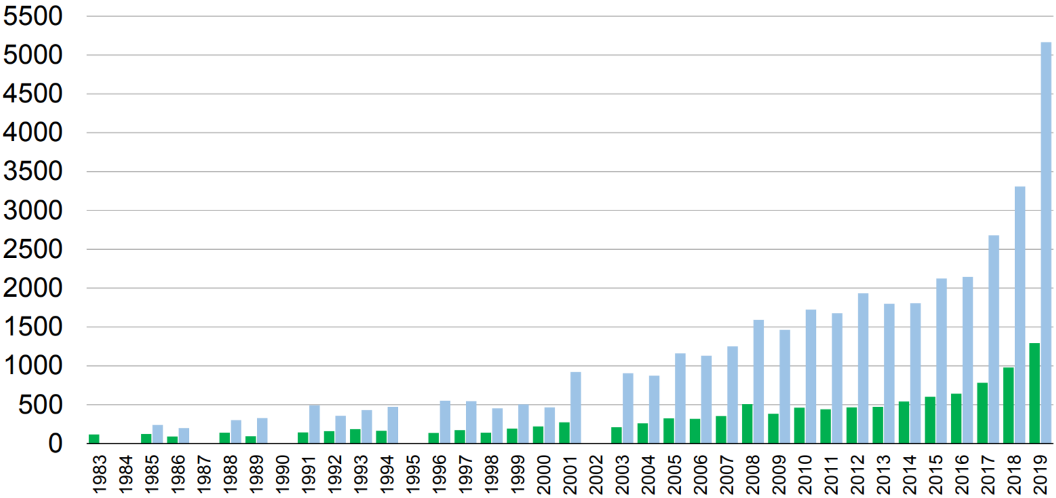
Deng, Dong, Socher, Li, Li and Li: ImageNet: A large-scale hierarchical image database. CVPR, 2009.

The Deep Learning Revolution



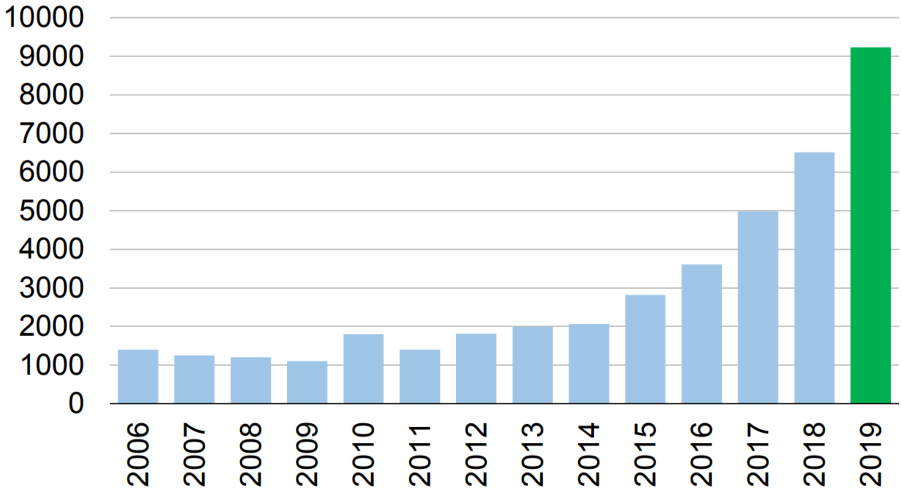
Slide Credit: Matthias Niessner

CVPR Submitted and Accepted Papers



Slide Credit: CVPR 2019 Welcome Slides

CVPR Attendance



Slide Credit: CVPR 2019 Welcome Slides

CVPR 2019 Sponsors

Diamond Sponsors

Microsoft, MEGVII 旷视, 商汤 sensetime

Platinum Sponsors

amazon, ANDURIL, anyvision A BETTER TOMORROW, Apple, ARGO, Baidu 百度, ByteDance, COGNEX, cruise, DeepBlue, facebook, Google, HUAWEI, IBM Research, inspur, intel AI, lyft, NAVER, LINE, NVIDIA, pony.ai, Qualcomm, 快手, SAMSUNG, Tencent YouTu Lab, TOYOTA RESEARCH INSTITUTE, Uber, WAYMO, wayfair, XILINX

Gold Sponsors

Adobe, AIBEE, Alibaba Group 阿里巴巴集团, Horizon Robotics, DiDi, Disney Research, HRI Honda Research Institute US, dji, IIAI Inception Institute of Artificial Intelligence, iRobot, JD.COM, Lambda, MALONG 码隆, Micron, Panasonic, PathAI, scale, Snap, SUPER ANNOTATE, voyage

Silver Sponsors

ABEJA, algolux, AMAX, appen data with a human touch, AQUIFI, Aurora, BAOBAB, BasicFinder, cadence, CEVA, Cirrascale CLOUD SERVICES, cloudfactory, CVTE, Dream-Future 视源股份, deepen, 深睿医疗 deepwise, figure eight, FORESIGHT.AI, gumgum, GENIUS PROS, HITACHI Inspire the Next, HOVER, 虎牙直播 HUYA.com, isee, iMerit, JUMIO, KINATRAX, Kitware, Matroid, Mercedes-Benz, meitu, MTLab, MVTec, nullmax 纽劢科技, ObjectVideo LABS, Orbital Insight, Playment, PROPHESEE, SAP, 眼控科技 EMDATA, SIEMENS Ingenuity for life, SIGOPT, SK telecom, SPORT LOGiQ, Tencent AI Lab, TESLA, Velodyne Lidar, Verisk, VinAI, VOXEL51, Weights & Biases, ZILLOW GROUP, ZOOX

Why is Visual Perception hard?

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

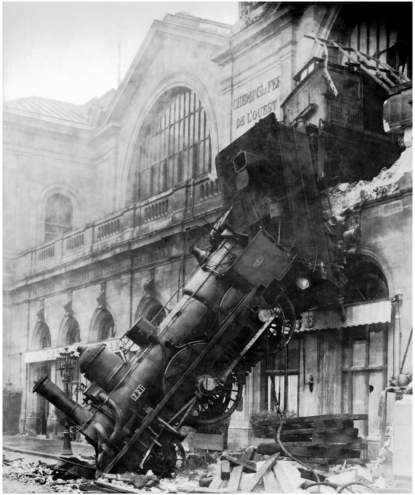
Seymour Papert

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".

Papert: The Summer Vision Project. MIT AI Memos, 1966.

Why is Visual Perception hard?

Why is Visual Perception hard?



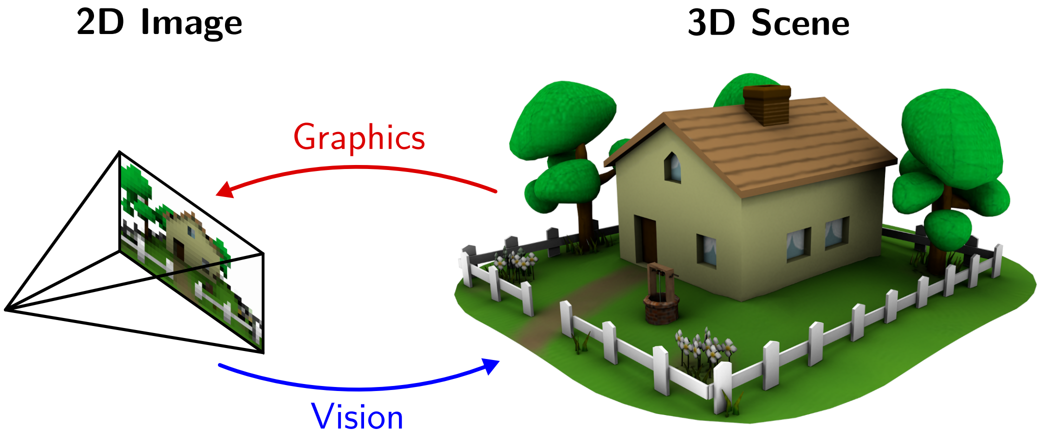
200	133	110	103	117	90	47	30	32	79	66	65
197	122	123	138	98	100	46	45	22	11	43	55
140	116	165	159	90	56	58	47	26	13	54	102
132	148	119	108	123	57	64	46	21	22	79	94
125	121	80	143	101	55	61	38	20	21	81	65
50	71	74	63	52	39	41	39	32	26	97	66
51	59	62	44	40	40	36	28	27	31	29	44
59	62	70	50	48	35	34	35	26	21	24	32
49	59	65	64	58	34	40	28	26	21	23	124
39	45	47	64	54	34	40	24	19	47	133	207
37	42	39	38	39	50	75	74	105	170	197	167
37	47	33	35	50	108	162	184	184	157	125	112
45	48	35	37	75	148	183	156	83	91	91	116
49	48	54	50	75	158	110	66	74	128	155	149
48	51	57	50	65	91	79	92	101	105	132	132
51	58	66	55	58	52	91	91	88	115	158	174
57	60	61	52	56	61	60	55	92	146	188	190
65	50	54	56	57	51	54	56	80	115	177	187
67	40	40	61	65	48	39	30	36	75	151	181
53	32	36	35	61	43	37	26	29	35	126	189
29	42	107	20	28	41	40	26	30	36	113	200
30	21	32	24	34	37	33	23	25	39	105	171
32	28	19	23	29	36	47	89	132	169	183	128
31	25	62	54	47	44	81	190	227	231	206	155
44	66	99	72	67	63	89	128	127	115	109	157
53	47	47	41	29	32	25	20	41	81	89	175
38	44	61	73	54	48	37	87	90	111	126	189
39	41	83	97	86	91	74	134	131	153	143	185
42	56	98	102	112	111	94	137	121	141	146	181
94	114	114	114	122	113	77	117	117	154	149	169
157	176	116	121	130	139	103	161	148	180	145	125
143	178	182	178	139	153	129	168	175	187	170	152
127	183	203	197	153	164	143	180	195	182	165	211
88	107	127	125	101	107	100	123	149	186	167	215

What we see

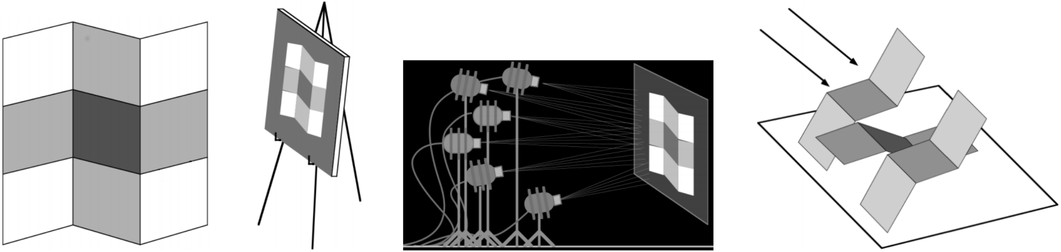
What the computer sees

Slide Credits: Li Fei-Fei

Challenges: Images are 2D Projections of the 3D World



Challenges: Images are 2D Projections of the 3D World


(a) (b) (c) (d)

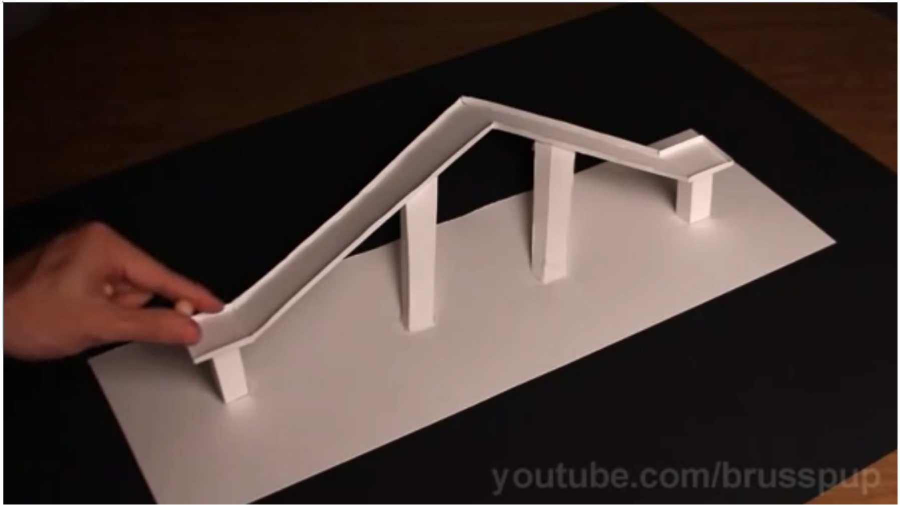
Adelson and Pentland's workshop metaphor:

To explain an image (a) in terms of reflectance, lighting and shape, (b) a painter, (c) a light designer and (d) a sculptor will design three different, but plausible, solutions.

Adelson and Pentland: The perception of shading and reflectance. Perception as Bayesian inference, 1996.

Ames Room Illusion

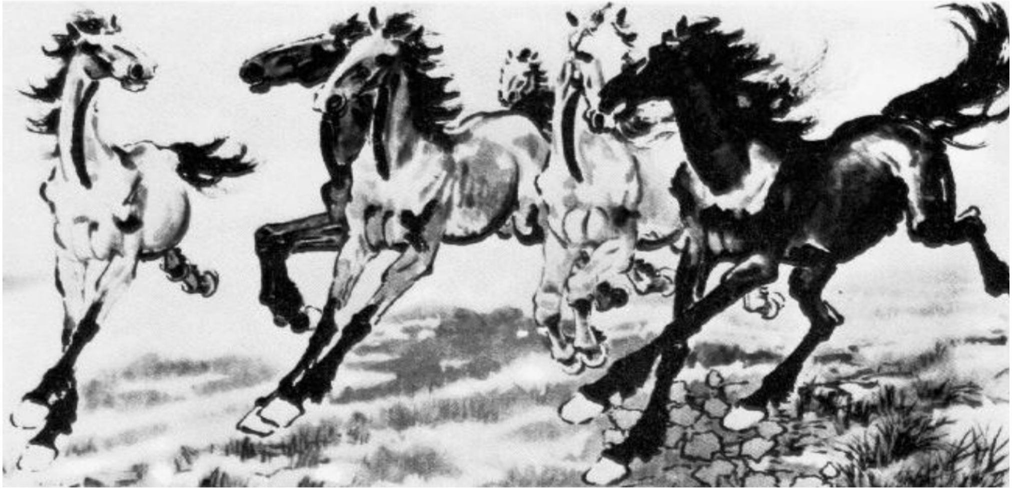
Perspective Illusion



Challenges: Viewpoint Variation

Michelangelo (1475-1564)

Challenges: Deformation



Xu Beihong (1943)

Challenges: Occlusion

René Magritte (1957)

Challenges: Illumination

Slide Credits: Antonio Torralba

Challenges: Illumination

Challenges: Illumination

Slide Credits: Antonio Torralba

Challenges: Illumination

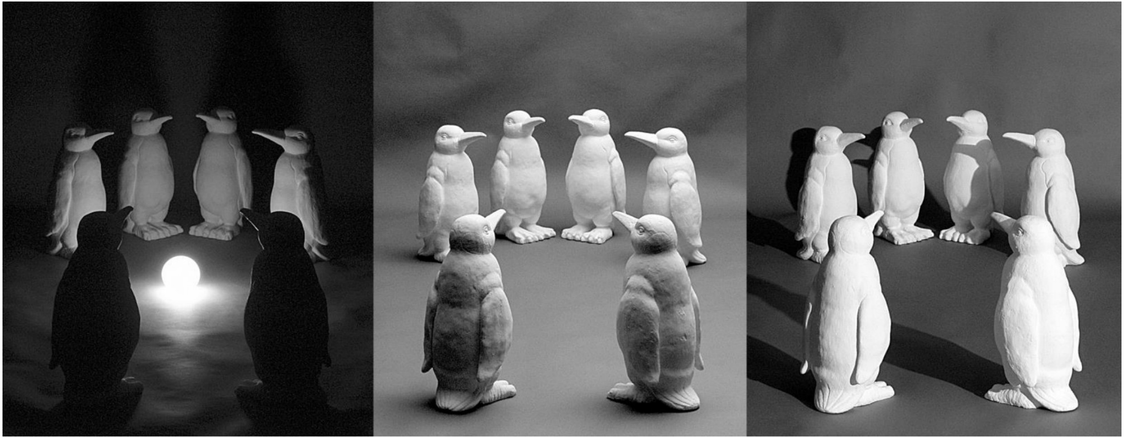
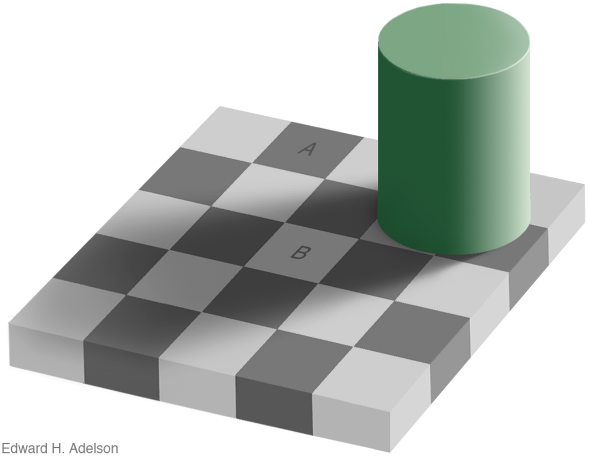


Image Credits: Jan Koenderink

Challenges: Motion

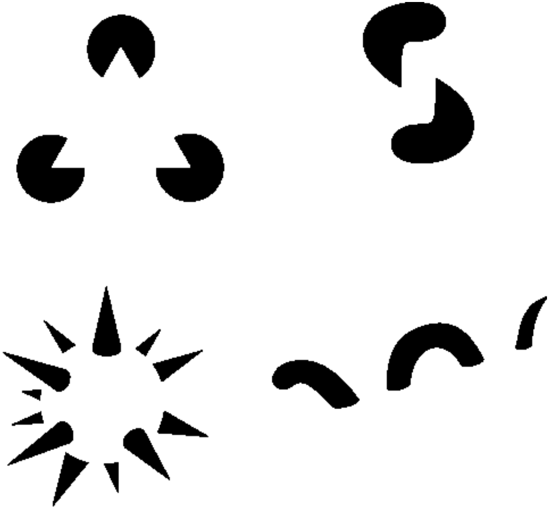
Challenges: Perception vs. Measurement



http://persci.mit.edu/gallery/checkershadow

Image Credits: Edward H. Adelson

Challenges: Perception vs. Measurement



Gaetano Kanizsa (1976)

Challenges: Perception vs. Measurement

Challenges: Local Ambiguities

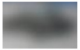


Image Credits: Antonio Torralba

Challenges: Local Ambiguities

Image Credits: Antonio Torralba

Challenges: Local Ambiguities

Image Credits: Antonio Torralba

Challenges: Intra Class Variation

http://www.homeworkshop.com/

Image Credits: Antonio Torralba

Challenges: Number of Object Categories

Image Credits: Antonio Torralba

1.3
History of Computer Vision

Credits

Svetlana Lazebnik (UIUC): Computer Vision: Looking Back to Look Forward

- `https://slazebni.cs.illinois.edu/spring20/`

Steven Seitz (Univ. of Washington): 3D Computer Vision: Past, Present, and Future

- `http://www.youtube.com/watch?v=kyIzMr917Rc`
- `http://www.cs.washington.edu/homes/seitz/talks/3Dhistory.pdf`

Pre-History

Perspective
Leonardo da Vinci
(1452–1519)

Photometry
Johann Heinrich Lambert
(1728–1777)

Least Squares
Carl Friedrich Gauss
(1777–1855)

Stereopsis
Charles Wheatstone
(1802–1875)

1510: Perspectograph

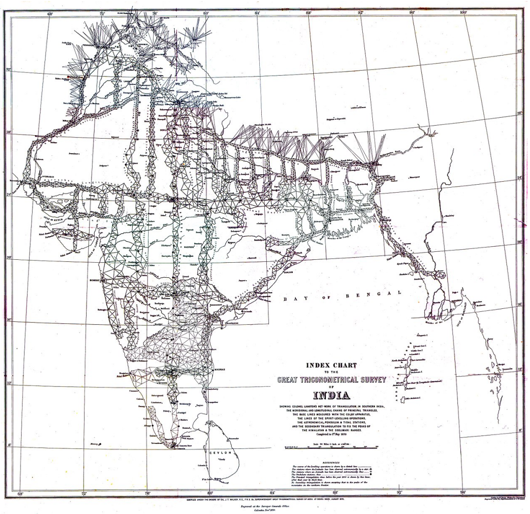
"Perspective is nothing else than the seeing of an object behind a sheet of glass, smooth and quite transparent, on the surface of which all the things may be marked that are behind this glass. All things transmit their images to the eye by pyramidal lines, and these pyramids are cut by the said glass. The nearer to the eye these are intersected, the smaller the image of their cause will appear."
– Leonardo da Vinci

1839: Daguerreotype

- First publicly available photographic process invented by Louis Daguerre
- Widely used during the 1840s and 1850s
- Polish a sheet of silver-plated copper and treat with fumes to make light sensitive
- Make resulting latent image visible by fuming it with mercury vapor and remove its sensitivity to light by chemical treatment
- Rinse, dry and seal behind glass

1802-1871: Great Trigonometrical Survey

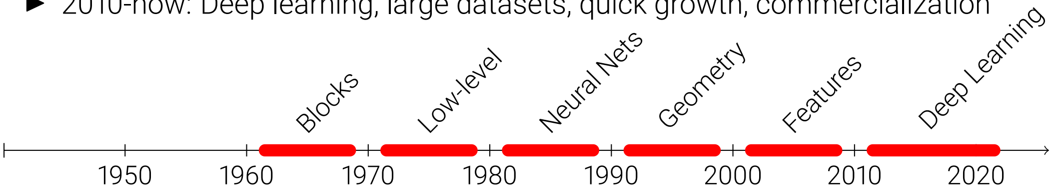
- Multi-decade project to measure the entire Indian subcontinent with scientific precision
- Under the leadership of George Everest, the project was made responsible of the Survey of India
- Manual bundle adjustment proves Mt. Everest highest mountain on earth mountain on earth



Overview

Waves of development:

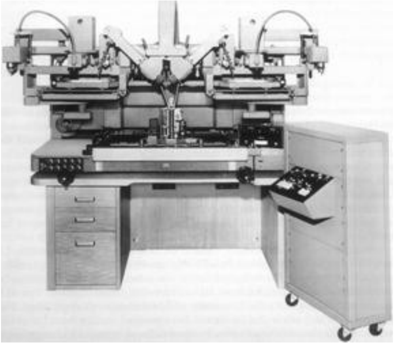
- 1960-1970: Blocks Worlds, Edges and Model Fitting
- 1970-1981: Low-level vision: stereo, flow, shape-from-shading
- 1985-1988: Neural networks, backpropagation, self-driving
- 1990-2000: Dense stereo and multi-view stereo, MRFs
- 2000-2010: Features, descriptors, large-scale structure-from-motion
- 2010-now: Deep learning, large datasets, quick growth, commercialization

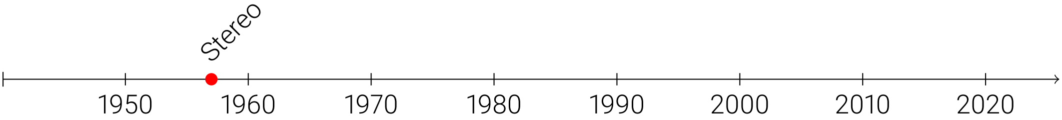


A Brief History of Computer Vision

1957: Stereo

- Gilbert Hobrough demonstrated an analog implementation of stereo image correlation
- This led to the creation of the Raytheon-Wild B8 Stereomat
- Used to create Elevation Maps (Photogrammetry, since 1840)





Roberts: Machine perception of 3-d solids. PhD Thesis, 1965.

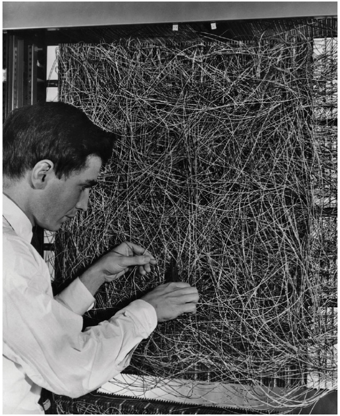
A Brief History of Computer Vision

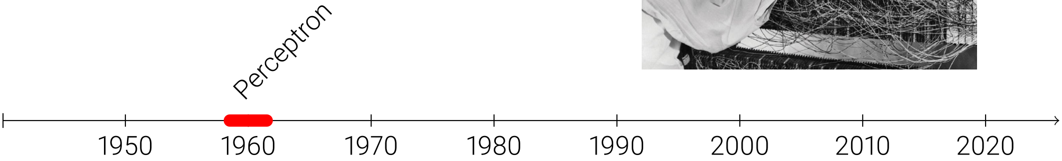
1958-1962: Rosenblatt's Perceptron

- First algorithm and implementation to train single linear threshold neuron
- Optimization of perceptron criterion:

$$\mathcal{L}(\mathbf{w}) = -\sum_{n \in \mathcal{M}} \mathbf{w}^T \mathbf{x}_n \, y_n$$

- Novikoff proved convergence



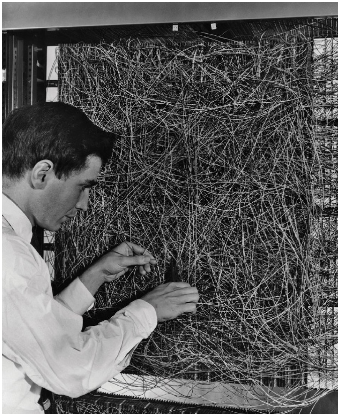


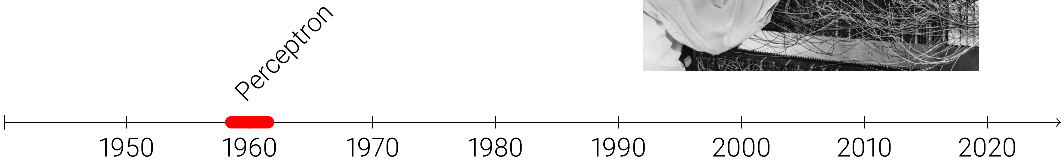
Rosenblatt: The perceptron - a probabilistic model for information storage and organization in the brain. Psychological Review, 1958.

A Brief History of Computer Vision

1958-1962: Rosenblatt's Perceptron

- First algorithm and implementation to train single linear threshold neuron
- Overhyped: Rosenblatt claimed that the perceptron will lead to computers that walk, talk, see, write, reproduce and are conscious of their existence



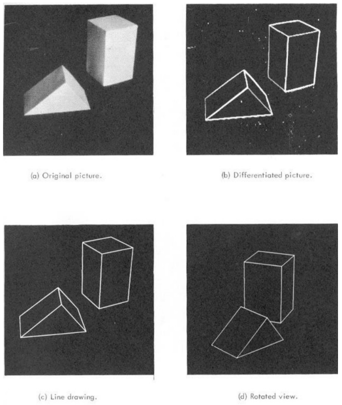


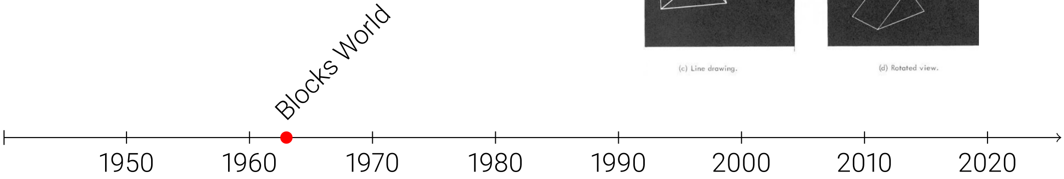
Rosenblatt: The perceptron - a probabilistic model for information storage and organization in the brain. Psychological Review, 1958.

A Brief History of Computer Vision

1963: Larry Robert's Blocks World

- Scene understanding for robotics
- Extracts edges as primitives
- Infers 3D structure of an object from topological structure of the 2D lines
- Interpret images as projections of 3D scenes, not 2D pattern recognition


(a) Original picture.
(b) Differentiated picture.
(c) Line drawing.
(d) Rotated view.

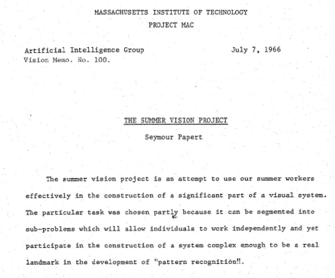


Roberts: Machine Perception of Three-Dimensional Solids. PhD Thesis, 1965.

A Brief History of Computer Vision

1966: MIT Summer Vision Project

- Underestimated the challenge of computer vision, committed to "blocks world"



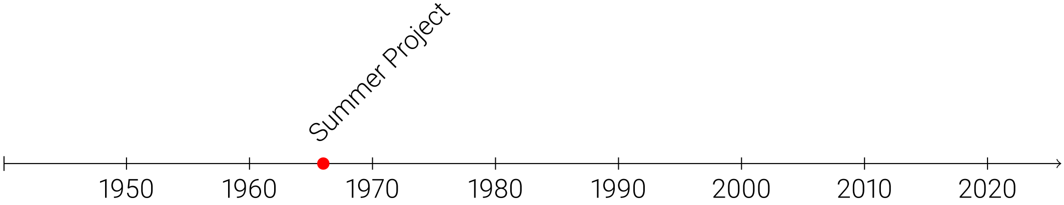
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group   July 7, 1966
Vision Memo. No. 100.

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".

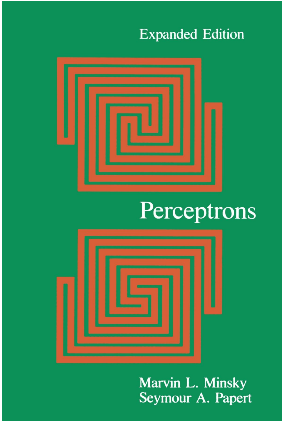


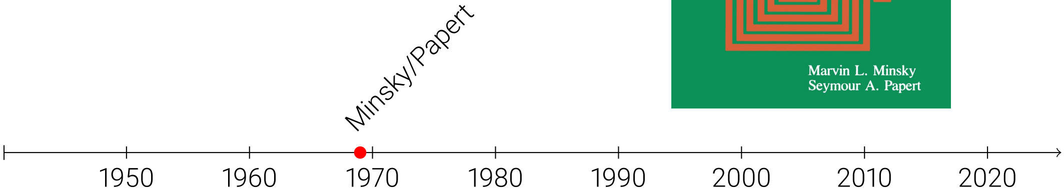
Papert: The Summer Vision Project. MIT AI Memos, 1966.

A Brief History of Computer Vision

1969: Minsky and Papert publish book

- Several discouraging results
- Showed that single-layer perceptrons cannot solve some very simple problems (XOR problem, counting)
- Symbolic AI research dominates 70s



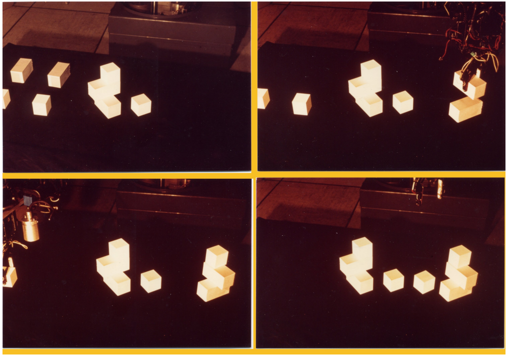


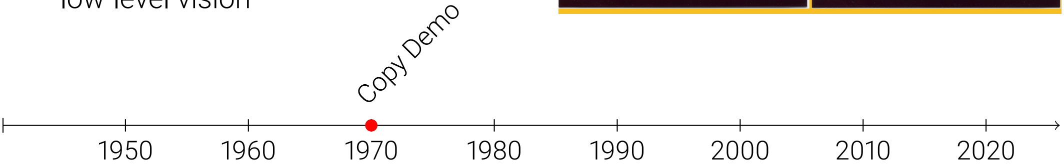
Minsky and Papert: Perceptrons: An introduction to computational geometry. MIT Press, 1969.

A Brief History of Computer Vision

1970: MIT Copy Demo

- Vision system recovers structure of a blocks scene, robot plans and builds copy from another set of blocks
- Vision, planning and manipulation
- But low-level edge finding not robust enough for task, led to attention on low-level vision





Papert: The Summer Vision Project. MIT AI Memos, 1966.

A Brief History of Computer Vision

1970: Shape from Shading

- Recover 3D from single 2D image
- Assumes Lambertian surface and constant albedo
- Applies smoothness reglarization to constrain the ill-posed problem

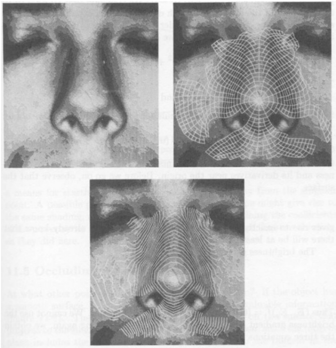
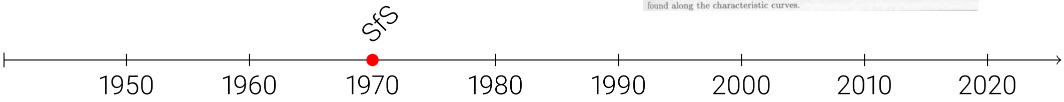


Figure 11-7. The shape-from-shading method is applied here to the recovery of the shape of a nose. The first picture shows the (crudely quantized) gray-level image available to the program. The second picture shows the base characteristics superimposed, while the third shows a contour map computed from the elevations found along the characteristic curves.

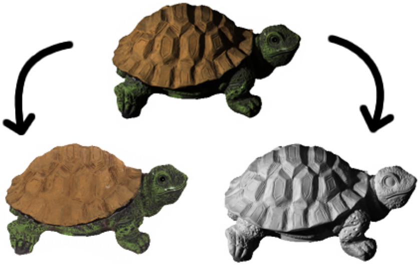


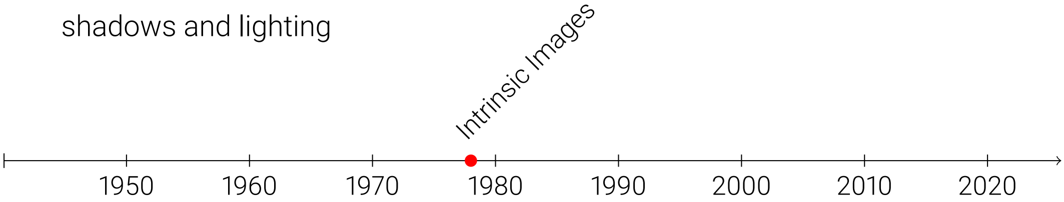
Horn: Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View. MIT TR, 1970.

A Brief History of Computer Vision

1978: Intrinsic Images

- Decomposing an image into its different intrinsic 2D layers, such as reflectance, shading, shape and motion components
- Useful for downstream tasks, e.g., object detection independent from shadows and lighting



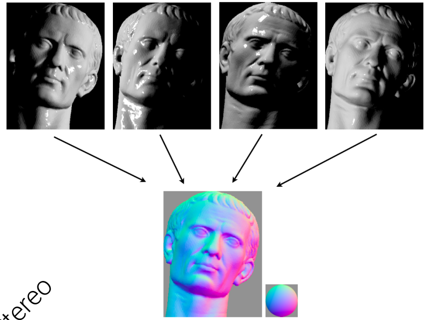


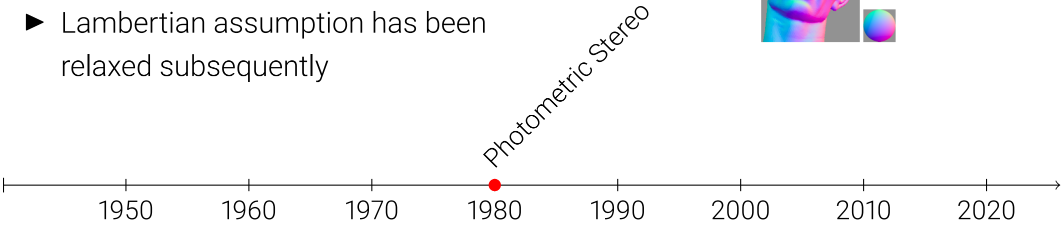
Barrow and Tenenbaum: Recovering intrinsic scene characteristics from images. Computer Vision Systems, 1978.

A Brief History of Computer Vision

1980: Photometric Stereo

- Recover 3D from multiple 2D images, taken from the same viewpoint with different lighting conditions
- Requires at least 3 images
- Unprecedented detail and accuracy
- Lambertian assumption has been relaxed subsequently



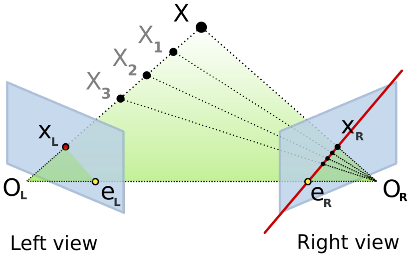


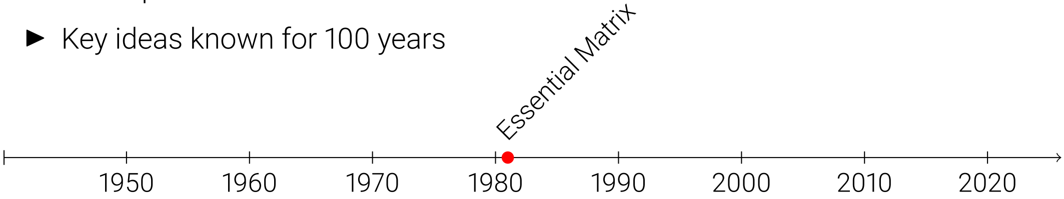
Woodham. Photometric method for determining surface orientation from multiple images. Optical Engineerings, 1980.

A Brief History of Computer Vision

1981: Essential Matrix

- Defines two-view geometry as matrix mapping points to epipolar lines
- Reduces correspondence search to a 1D problem
- Can be estimated from a set of 2D correspondences
- Key ideas known for 100 years



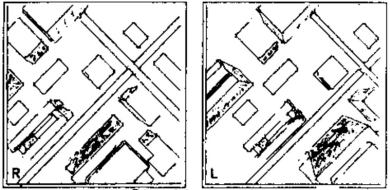


Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. Nature, 1981.

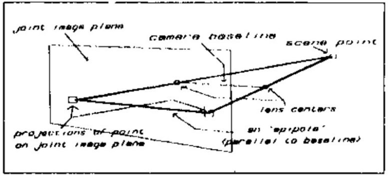
A Brief History of Computer Vision

1981: Binocular Scanline Stereo

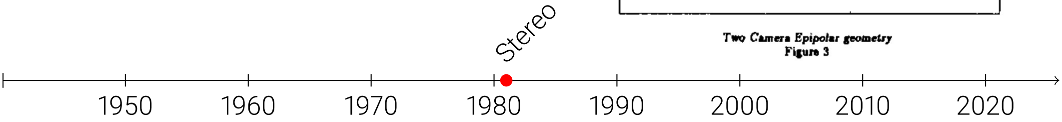
- Correlate points along epipolar lines
- Use dynamic programming to introduce constraints along scanlines (image rows)
- Allows for overcoming abiguities, but streaking artifacts between rows



Edges of the stereo pair
Figure 2



Two Camera Epipolar geometry
Figure 3

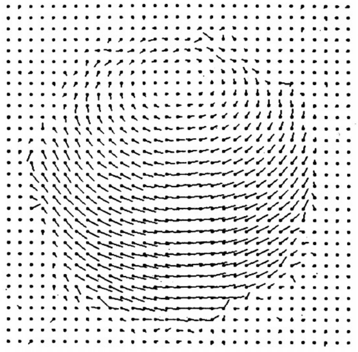


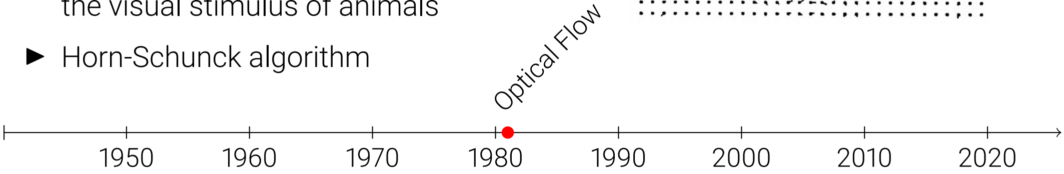
Baker and Binford: Depth from Edge and Intensity Based Stereo. IJCAI, 1981.

A Brief History of Computer Vision

1981: Dense Optical Flow

- Pattern of apparent motion of objects, surfaces, and edges in a visual scene
- Measured by (densely) tracking pixels between two frames
- Investigated by Gibson to describe the visual stimulus of animals
- Horn-Schunck algorithm



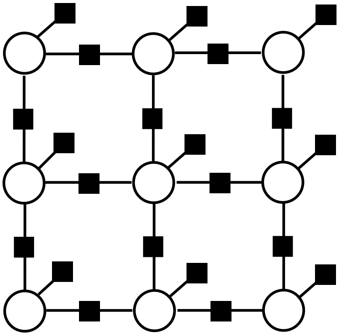


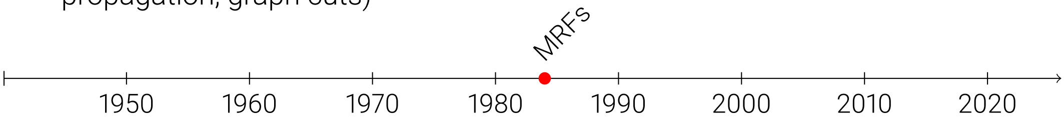
Horn and Schunck: Determining Optical Flow. Artificial Intelligence, 1981.

A Brief History of Computer Vision

1984: Markov Random Fields

- MRFs for encoding prior knowledge (e.g., about smoothness)
- Resolves ambiguities in many ill-posed vision problems (e.g., stereo, flow, denoising)
- Global optimization (e.g., variational inference, sampling, belief propagation, graph cuts)



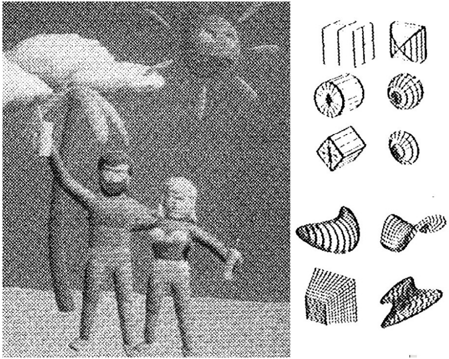


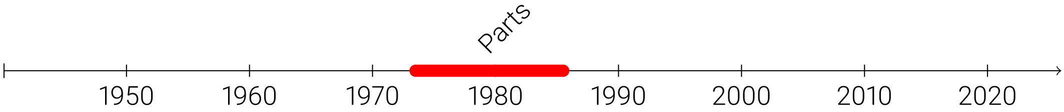
Geman and Geman: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. TPAMI, 1984.

A Brief History of Computer Vision

1980s: Part-based Models

- 1973: Pictorial Structures
- 1976: Generalized Cylinders (solids of revolution, swept curves)
- 1986: Superquadrics (generalization of quadric surfaces)
- Express complex relationships
- Compact representation



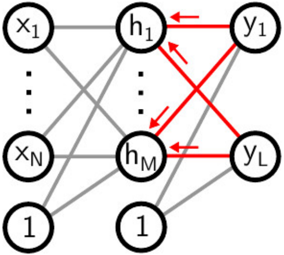


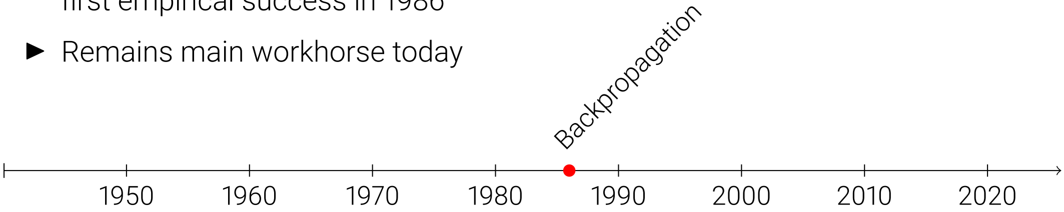
Pentland: Parts: Structured descriptions of shape. AAAI, 1986.

A Brief History of Computer Vision

1986: Backpropagation Algorithm

- Efficient calculation of gradients in a deep network wrt. network weights
- Enables application of gradient based learning to deep networks
- Known since 1961, but first empirical success in 1986
- Remains main workhorse today



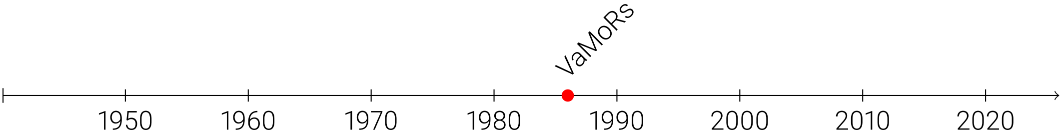


Rumelhart, Hinton and Williams: Learning representations by back-propagating errors. Nature, 1986.

A Brief History of Computer Vision

1986: Self-Driving Car VaMoRs

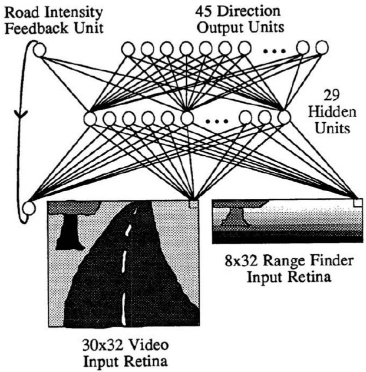
- Developed by Ernst Dickmanns in context of EUREKA-Prometheus
- Demonstration to Daimler-Benz Research 1986 in Stuttgart
- Longitudinal & lateral guidance with lateral acceleration feedback
- Speed: 0 to 36 km/h

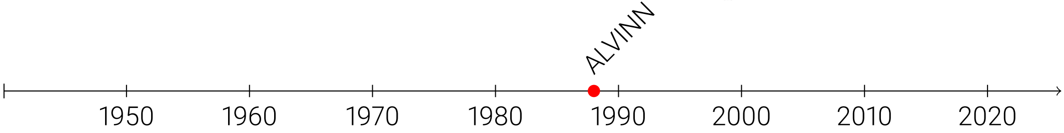


A Brief History of Computer Vision

1988: Self-Driving Car ALVINN

- Forward-looking, vision based driving
- Fully connected neural network maps road images to vehicle turn radius
- Trained on simulated road images
- Tested on unlined paths, lined city streets and interstate highways
- 90 consecutive miles at up to 70 mph



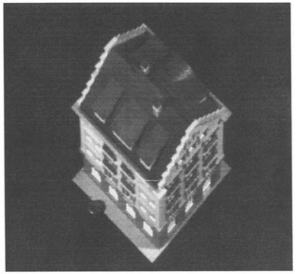


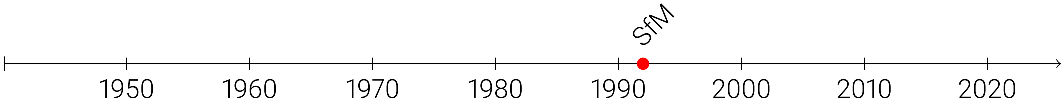
Pomerleau: ALVINN: An Autonomous Land Vehicle in a Neural Network. NIPS, 1988.

A Brief History of Computer Vision

1992: Structure-from-Motion

- Estimating 3D structures from 2D image sequences of static scenes
- Requires only a single camera
- Tomasi-Kanade factorization provides closed-form (SVD-based) solution for orthographic case
- Today: non-linear least squares



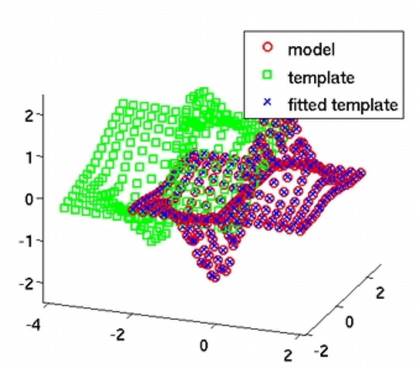


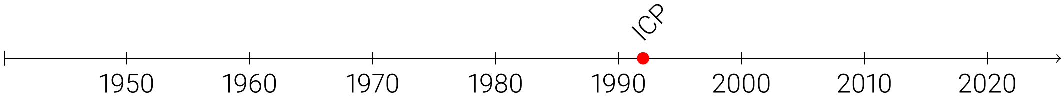
Tomasi and Kanade: Shape and motion from image streams under orthography: a factorization method. IJCV, 1992.

A Brief History of Computer Vision

1992: Iterative Closest Points

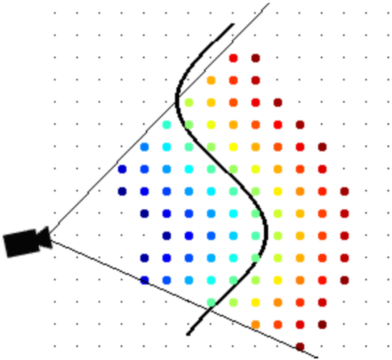
- Registering two point clouds by iteratively optimizing a (rigid or non-rigid) transformation
- Used to aggregate partial 2D or 3D surfaces from different scans, to estimate relative camera poses from point clouds or to localize wrt. a map





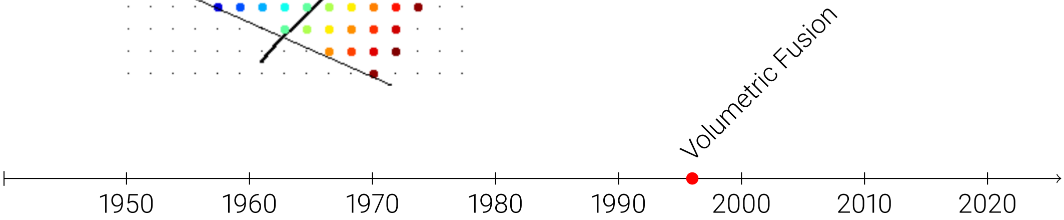
Besl and McKay: A Method for Registration of 3-D Shapes. PAMI, 1992.

A Brief History of Computer Vision



1996: Volumetric Fusion

- Aggregation of multiple implicitly represented surfaces by averaging signed distance values
- Mesh-extraction as post-processing



Curless and Levoy: A Volumetric Method for Building Complex Models from Range Images. SIGGRAPH, 1996.

A Brief History of Computer Vision

1998: Multi-View Stereo

- 3D reconstruction from multiple input images using level-set methods
- Reconstruction vs. image matching
- Proper model of visibility
- Flexible topology
- Provable convergence
- Other approaches (dead-ends): Voxel-coloring, space carving

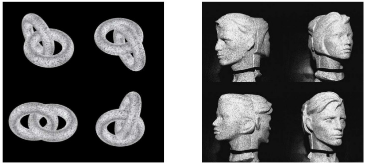


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

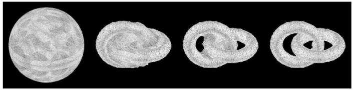
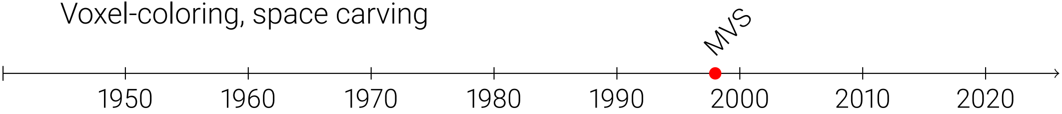


Fig. 4. Evolution of the surface for the two tori.

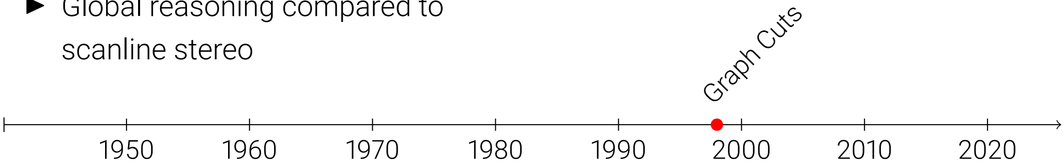


Faugeras and Keriven: Complete Dense Stereovision Using Level Set Methods. ECCV, 1998.

A Brief History of Computer Vision

1998: Stereo with Graph Cuts

- Popular discrete MAP inference algorithm for Markov Random Fields
- First versions included unary and pairwise terms
- Later versions also included specific forms of higher-order potentials
- Global reasoning compared to scanline stereo

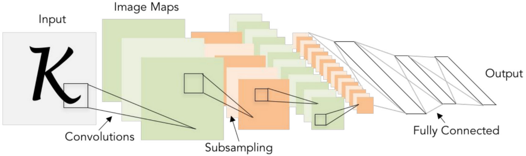


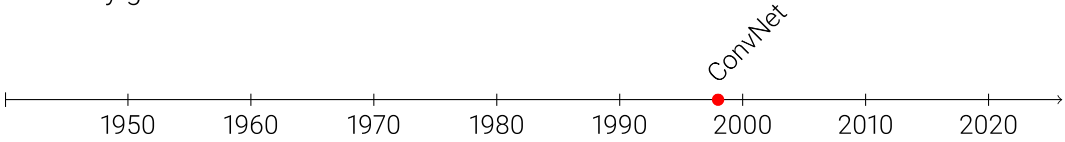
Boykov, Veksler and Zabih: Markov Random Fields with Efficient Approximations. CVPR, 1998.

A Brief History of Computer Vision

1998: Convolutional Neural Networks

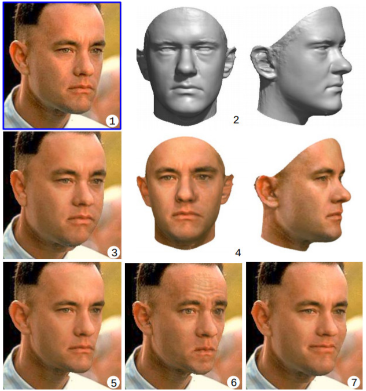
- Similar to Neocognitron, but trained end-to-end using backpropagation
- Implements spatial invariance via convolutions and max-pooling
- Weight sharing reduces parameters
- Tanh/Softmax activations
- Only good results on MNIST so far





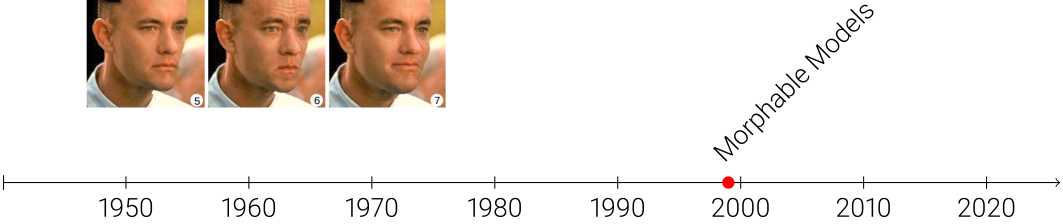
LeCun, Bottou, Bengio and Haffner: Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.

A Brief History of Computer Vision



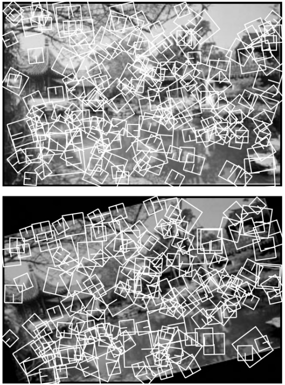
1999: Morphable Models

- Single-view 3D face reconstruction
- Linear combination of
  200 laser scans of faces
- Stunning results at the time



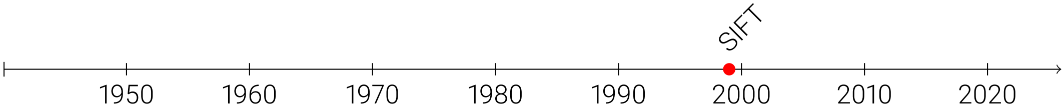
Blanz and Vetter: A Morphable Model for the Synthesis of 3D Faces. SIGGRAPH, 1999.

A Brief History of Computer Vision



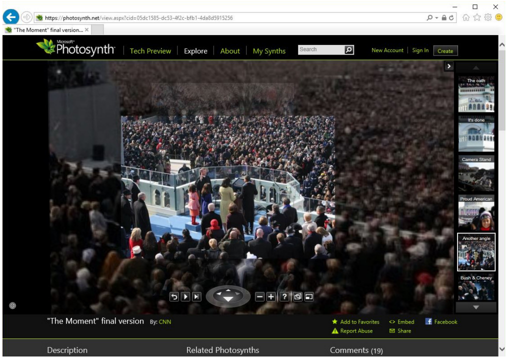
1999: SIFT

- Scale Invariant Feature Transform
- Detection and description of salient local features in an image
- Enabled many applications (e.g., image stitching, reconstruction, motion estimation, . . .)



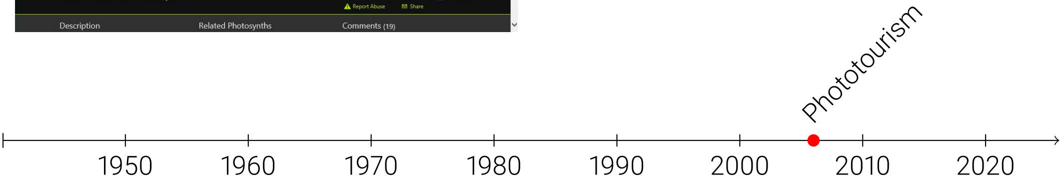
Lowe: Object Recognition from Local Scale-Invariant Features. ICCV, 1999.

A Brief History of Computer Vision



2006: Photo Tourism

- Large-scale 3D reconstruction from internet photos
- Key ingredients: SIFT feature matching, bundle adjustment
- Microsoft Photosynth (discont.)

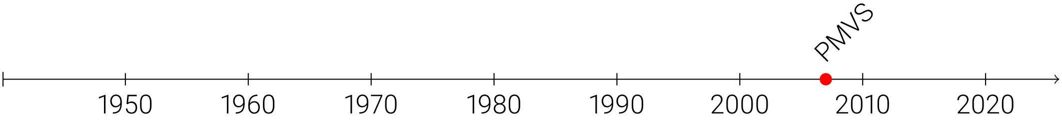


Snavely, Seitz and Szeliski: Photo tourism: exploring photo collections in 3D. SIGGRAPH, 2006.

A Brief History of Computer Vision

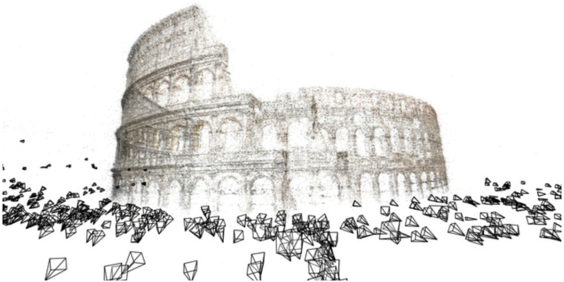
2007: PMVS

- Patch-based Multi View Stereo
- Robust reconstruction of various small and large objects
- Performance of 3D reconstruction techniques continues to increase



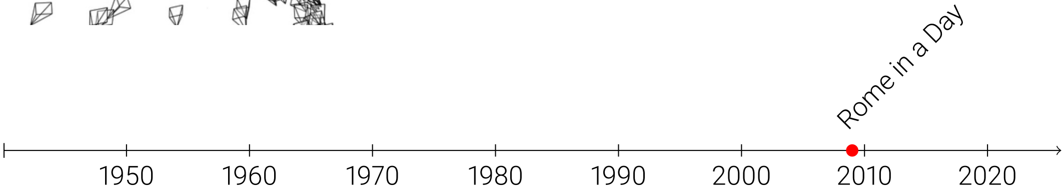
Furukawa and Ponce: Accurate, Dense, and Robust Multi-View Stereopsis. CVPR 2007.

A Brief History of Computer Vision



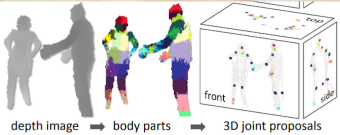
2009: Building Rome in a Day

- 3D reconstruction of landmarks and cities from unstructured Internet photo-collections
- Follow-up: Rome on a Cloudless Day



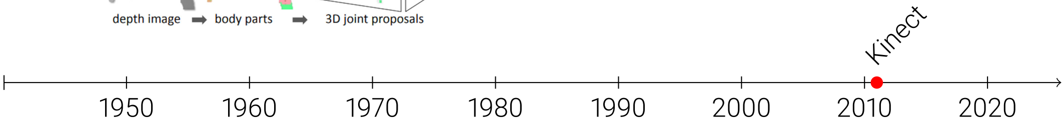
Agarwal, Snavely, Simon, Seitz and Szeliski: Building Rome in a day. ICCV, 2009.

A Brief History of Computer Vision



2011: Kinect

- Active light 3D sensing
- ML for 3D pose estimation
- Multiple hardware generations
- Early versions failed to commercialize but heavily used for robotics and vision research



Shotton et al.: Real-time human pose recognition in parts from single depth images. CVPR, 2011.

A Brief History of Computer Vision

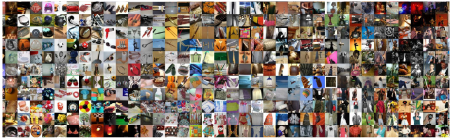
2009-2012: ImageNet and AlexNet

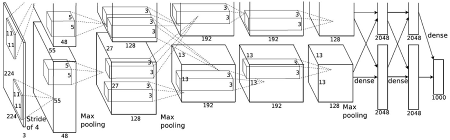
ImageNet

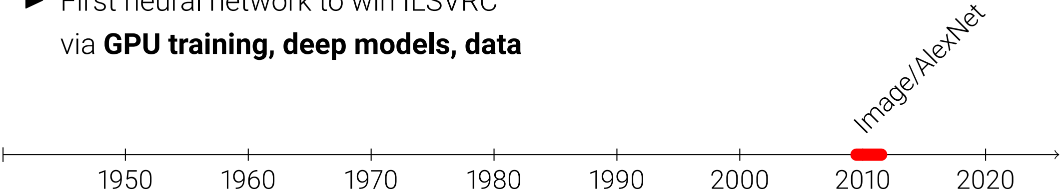
- ► Recognition benchmark (ILSVRC)
- ► 10 million annotated images
- ► 1000 categories

AlexNet

- ► First neural network to win ILSVRC via **GPU training, deep models, data**







Krizhevsky, Sutskever and Hinton: ImageNet classification with deep convolutional neural networks. NIPS, 2012.

A Brief History of Computer Vision

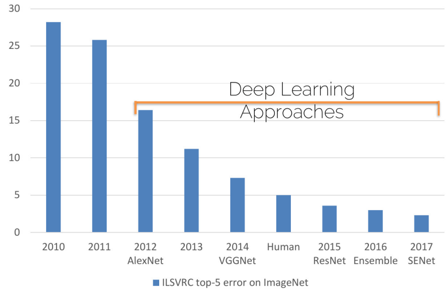
2009-2012: ImageNet and AlexNet

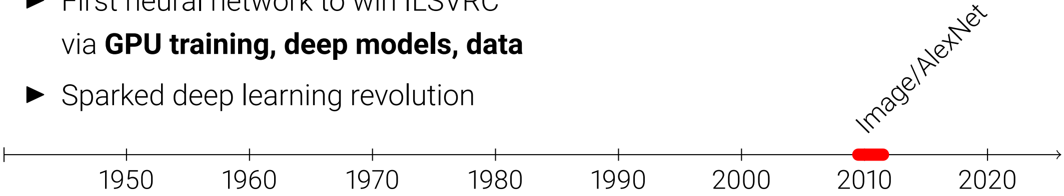
ImageNet

- Recognition benchmark (ILSVRC)
- 10 million annotated images
- 1000 categories

AlexNet

- First neural network to win ILSVRC via **GPU training, deep models, data**
- Sparked deep learning revolution



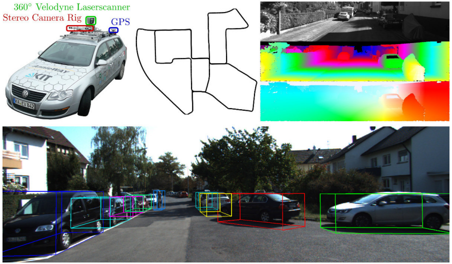


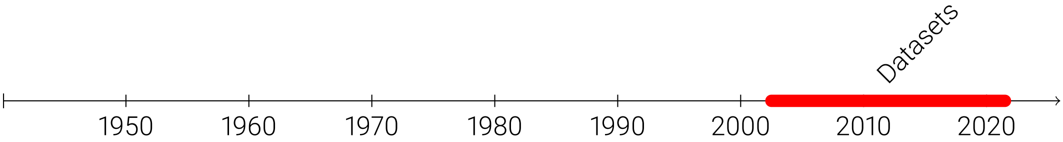
Krizhevsky, Sutskever and Hinton: ImageNet classification with deep convolutional neural networks. NIPS, 2012.

A Brief History of Computer Vision

2002-now: Golden Age of Datasets

- Middlebury Stereo and Flow
- KITTI, Cityscapes: Self-driving
- PASCAL, MS COCO: Recognition
- ShapeNet, ScanNet: 3D DL
- Visual Genome: Vision/Language
- MITOS: Breast cancer



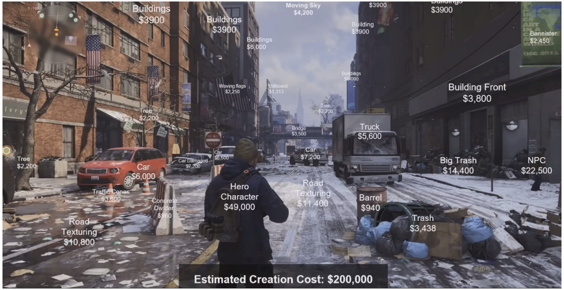


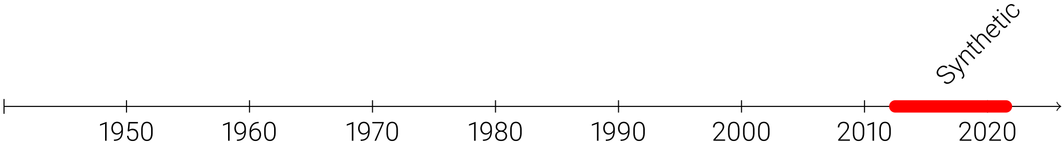
Geiger, Lenz and Urtasun: Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. CVPR, 2012.

A Brief History of Computer Vision

2012-now: Synthetic Data

- Annotating real data is expensive
- Led to surge of synthetic datasets
- Creating 3D assets is also costly



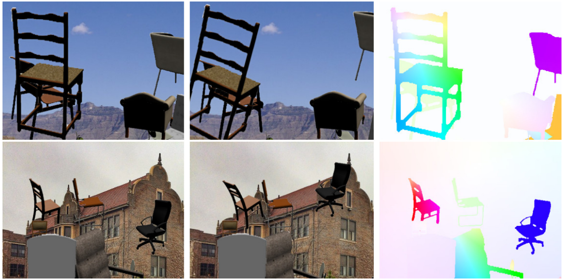


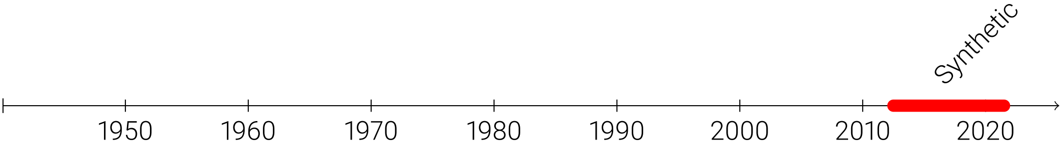
Dosovitskiy et al.: FlowNet: Learning Optical Flow with Convolutional Networks. ICCV, 2015.

A Brief History of Computer Vision

2012-now: Synthetic Data

- Annotating real data is expensive
- Led to surge of synthetic datasets
- Creating 3D assets is also costly
- But even very simple 3D datasets proved tremendously useful for pre-training (e.g., in optical flow)



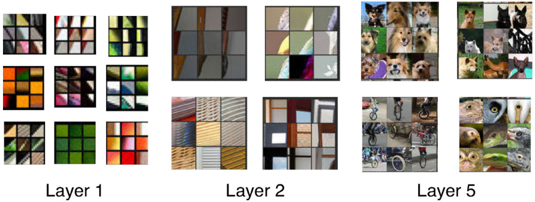


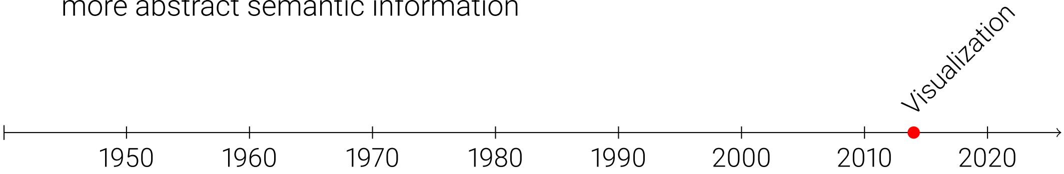
Dosovitskiy et al.: FlowNet: Learning Optical Flow with Convolutional Networks. ICCV, 2015.

A Brief History of Computer Vision

2014: Visualization

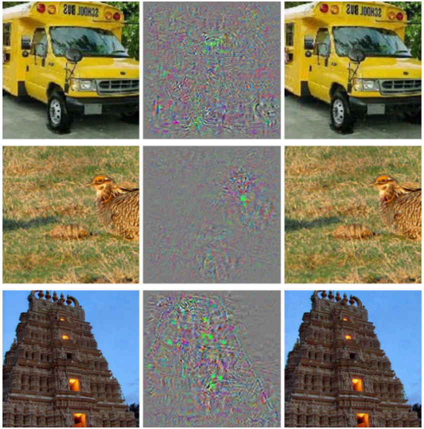
- Goal: provide insights into what the network (black box) has learned
- Visualized image regions that most strongly activate various neurons at different layers of the network
- Found that higher levels capture more abstract semantic information





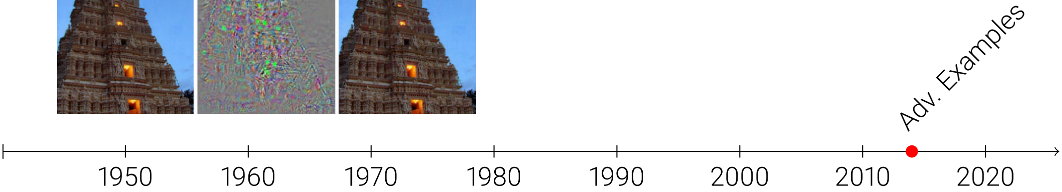
Zeiler and Fergus: CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. CVPR Workshops, 2014.

A Brief History of Computer Vision



2014: Adversarial Examples

- Accurate image classifiers can be fooled by imperceptible changes (here magnified for visibility)
- All images in the right column are classified as "ostrich"

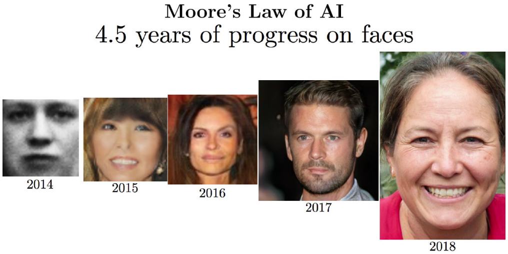


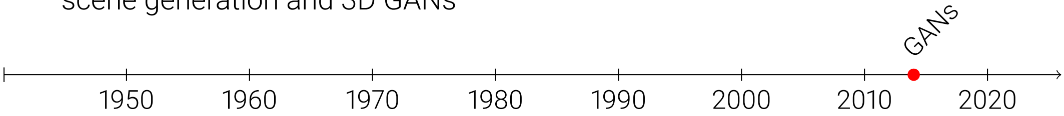
Szegedy et al.: Intriguing properties of neural networks. ICLR, 2014.

A Brief History of Computer Vision

2014: Generative Adversarial Networks

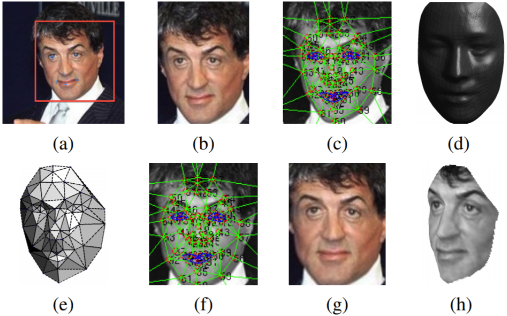
- Deep generative models (VAEs, GANs) produce compelling images
- StyleGAN2 is state-of-the-art
- Results on faces hard to distinguish from real images
- Active research on image translation, domain adaptation, content and scene generation and 3D GANs

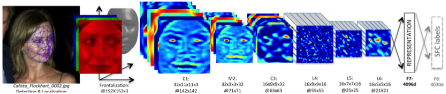




Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, Bengio: Generative Adversarial Networks. NIPS, 2014.

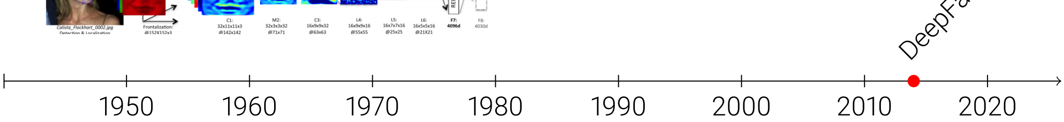
A Brief History of Computer Vision





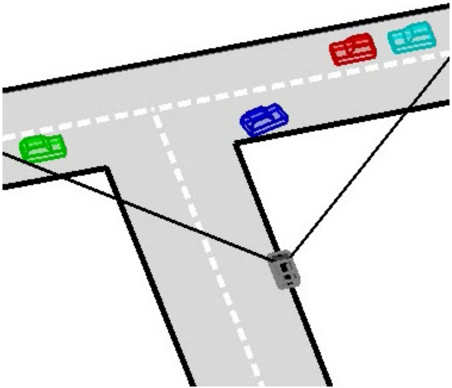
2014: DeepFace

- Combination of model-based alignment with deep learning for face recognition
- First model to reach human-level face recognition performance



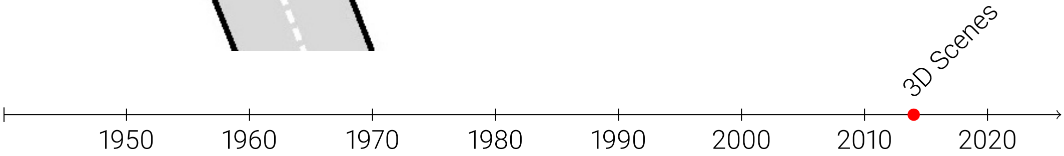
Taigman, Yang, Ranzato and Wolf: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. CVPR, 2014.

A Brief History of Computer Vision



2014: 3D Scene Understanding

- Parsing RGB and RGB-D images into holistic 3D scene representations
- Methods for indoors and outdoors

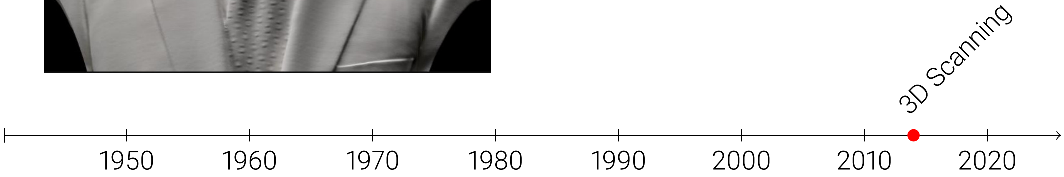


Geiger, Lauer, Wojek, Stiller and Urtasun: 3D Traffic Scene Understanding From Movable Platforms. PAMI, 2014.

A Brief History of Computer Vision

2014: 3D Scanning

- 3D scanning techniques allow for creating accurate replicas
- Debevec's team scans Obama
- Exhibition in Smithsonian
- `https://dpo.si.edu/blog/`

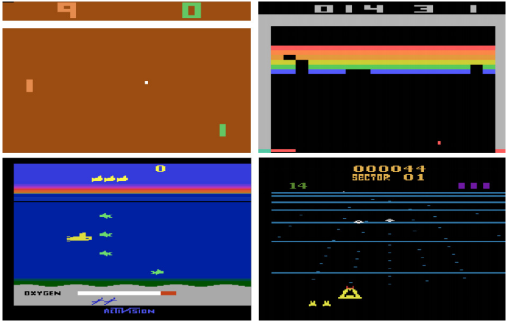


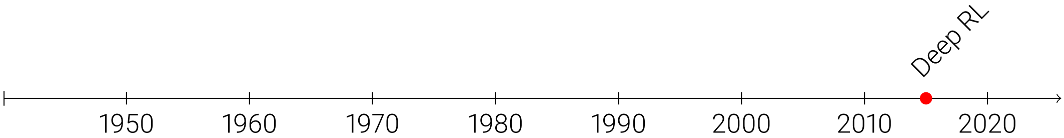
Metallo et al.: Scanning and printing a 3D portrait of president Barack Obama. SIGGRAPH Studio, 2015.

A Brief History of Computer Vision

2015: Deep Reinforcement Learning

- Learning a policy (state→action) through random exploration and reward signals (e.g., game score)
- No other supervision
- Success on many Atari games
- But some games remain hard



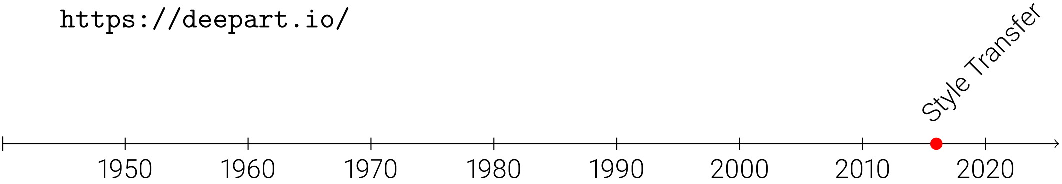


Mnih et al.: Human-level control through deep reinforcement learning. Nature, 2015.

A Brief History of Computer Vision

2016: Style Transfer

- Manipulate photograph to adopt style of a another image (painting)
- Uses deep network pre-trained on ImageNet for disentangling content from style
- It is fun! Try yourself: `https://deepart.io/`

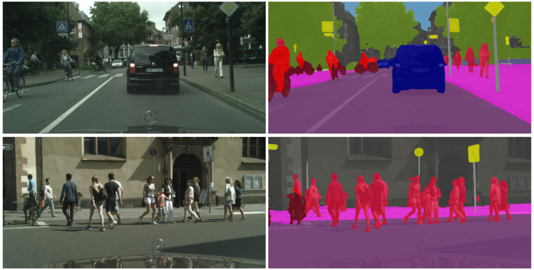


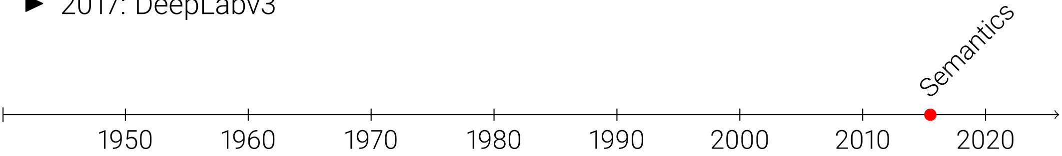
Gatys, Ecker and Bethge: Image Style Transfer Using Convolutional Neural Networks. CVPR, 2016.

A Brief History of Computer Vision

2015-2017: Semantic Segmentation

- Assign semantic class to every pixel
- Semantic segmentation starts to work on challenging real-world datasets (e.g., CityScapes)
- 2015: FCN, SegNet
- 2016: DeepLab, FSO
- 2017: DeepLabv3



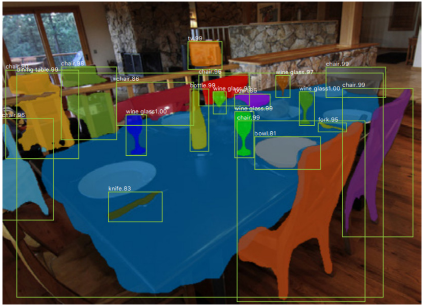


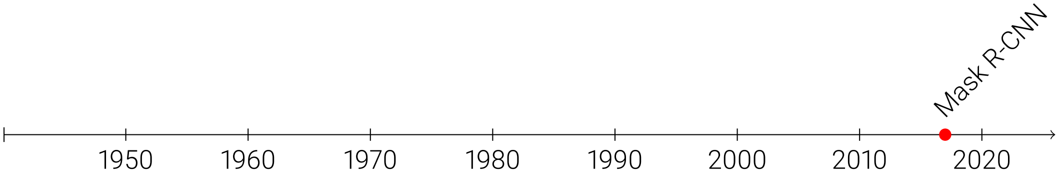
Kundu, Vineet and Koltun: Feature Space Optimization for Semantic Video Segmentation. CVPR, 2016.

A Brief History of Computer Vision

2017: Mask R-CNN

- Deep neural network for joint object detection and instance segmentation
- Outputs “structured object”, not only a single number (class label)
- State-of-the-art on MS-COCO



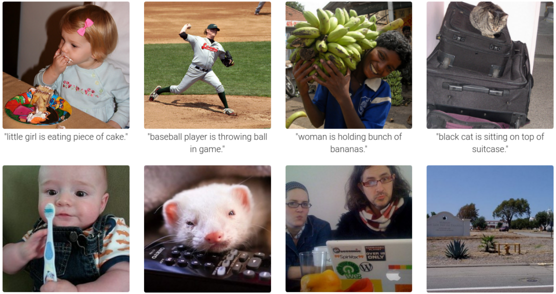


He, Gkioxari, Dollár and Ross Girshick: Mask R-CNN. ICCV, 2017.

A Brief History of Computer Vision

2017: Image Captioning

- Growing interest in combining vision with language
- Several new tasks emerged including image captioning and visual question answering
- However, models still lack understanding / commonsense



"little girl is eating piece of cake."

"baseball player is throwing ball in game."

"woman is holding bunch of bananas."

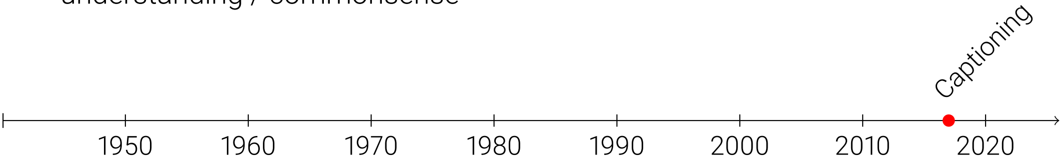
"black cat is sitting on top of suitcase."

"a young boy is holding a baseball bat."

"a cat is sitting on a couch with a remote control."

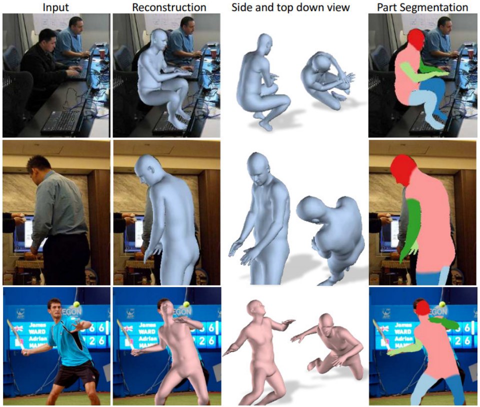
"a woman holding a teddy bear in front of a mirror."

"a horse is standing in the middle of a road."



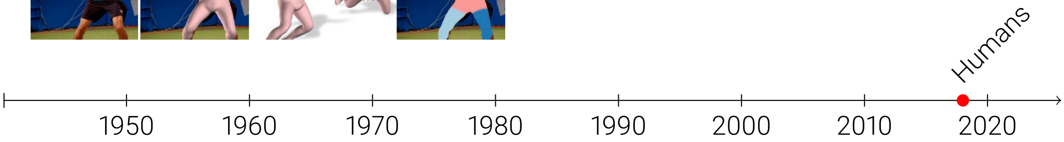
Karpathy and Fei-Fei: Deep Visual-Semantic Alignments for Generating Image Descriptions. PAMI, 2017.

A Brief History of Computer Vision



2018: Human Shape and Pose

- Human pose/shape models mature
- Rich parametric models (e.g., SMPL, STAR)
- Regression from RGB images only
- Models of pose-dependent deformation and clothing

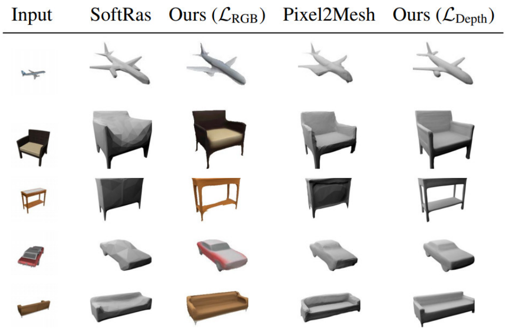


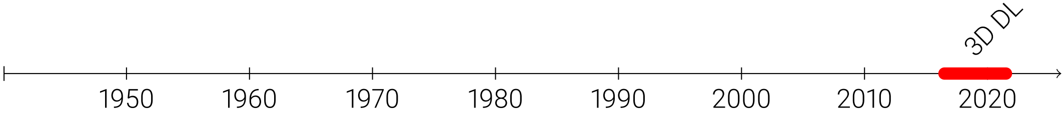
Kanazawa, Black, Jacobs and Malik: End-to-End Recovery of Human Shape and Pose. CVPR, 2018.

A Brief History of Computer Vision

2016-2020: 3D Deep Learning

- First deep models to output 3D representations
- Voxels, point clouds, meshes, implicit representations
- Prediction of 3D models even from a single image
- Geometry, materials, light, motion





Niemeyer, Mescheder, Oechsle, Geiger: Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision. CVPR, 2020.

A Brief History of Computer Vision

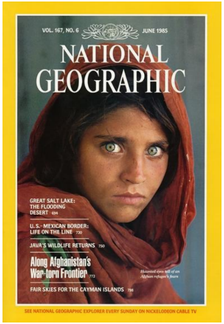
Applications and Commercial Products

Google Portrait Mode

Skydio 2 Drone

Self-Driving Cars

Microsoft Hololens



Iris Recognition

A Brief History of Computer Vision

Current Challenges

- Un-/Self-Supervised Learning
- Interactive learning
- Accuracy (e.g., self-driving)
- Robustness and generalization
- Inductive biases
- Understanding and mathematics
- Memory and compute
- Ethics and legal questions