

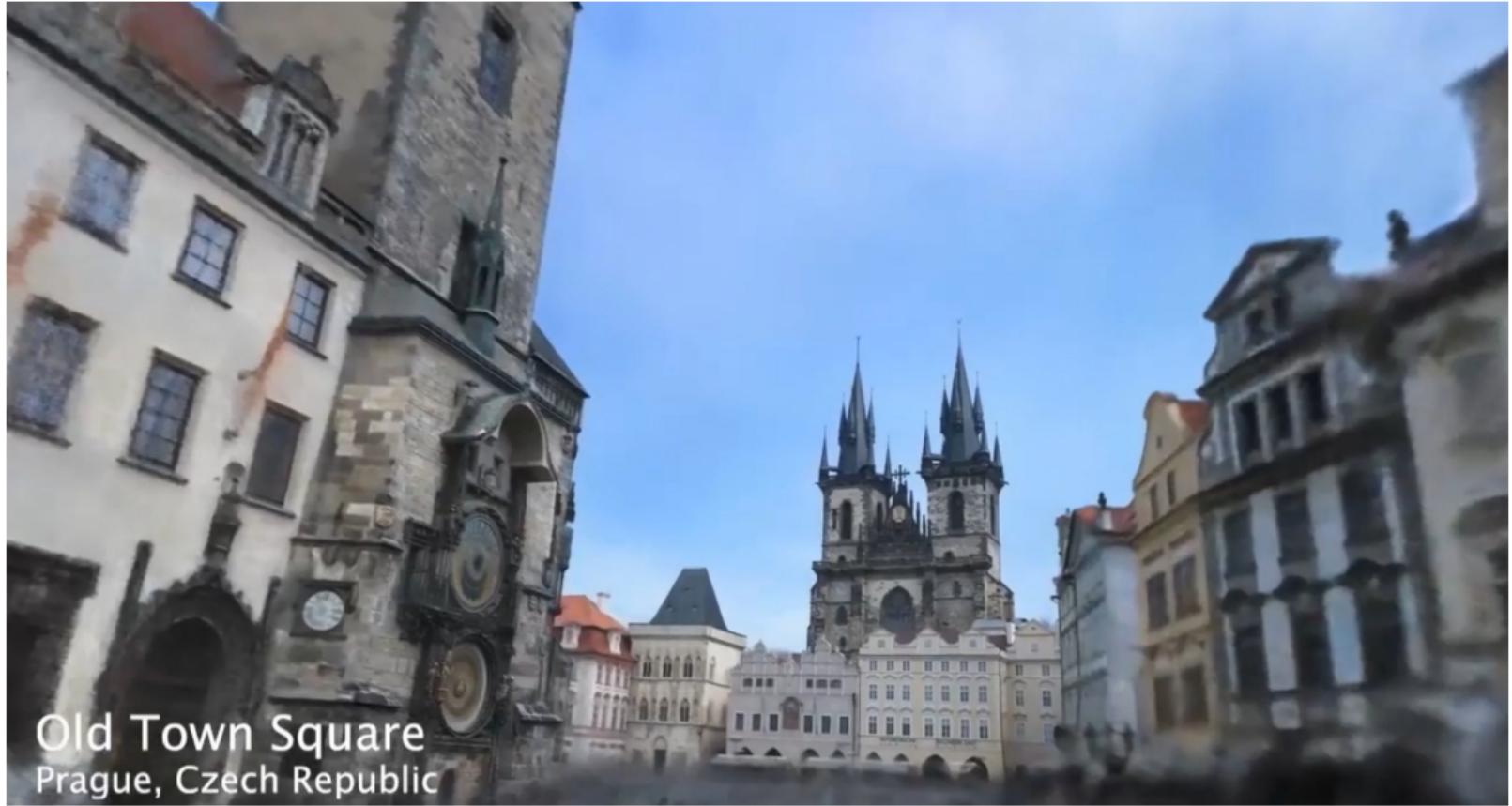
Computer Vision

Lecture 1 – Introduction

Prof. Dr.-Ing. Andreas Geiger
Autonomous Vision Group
University of Tübingen

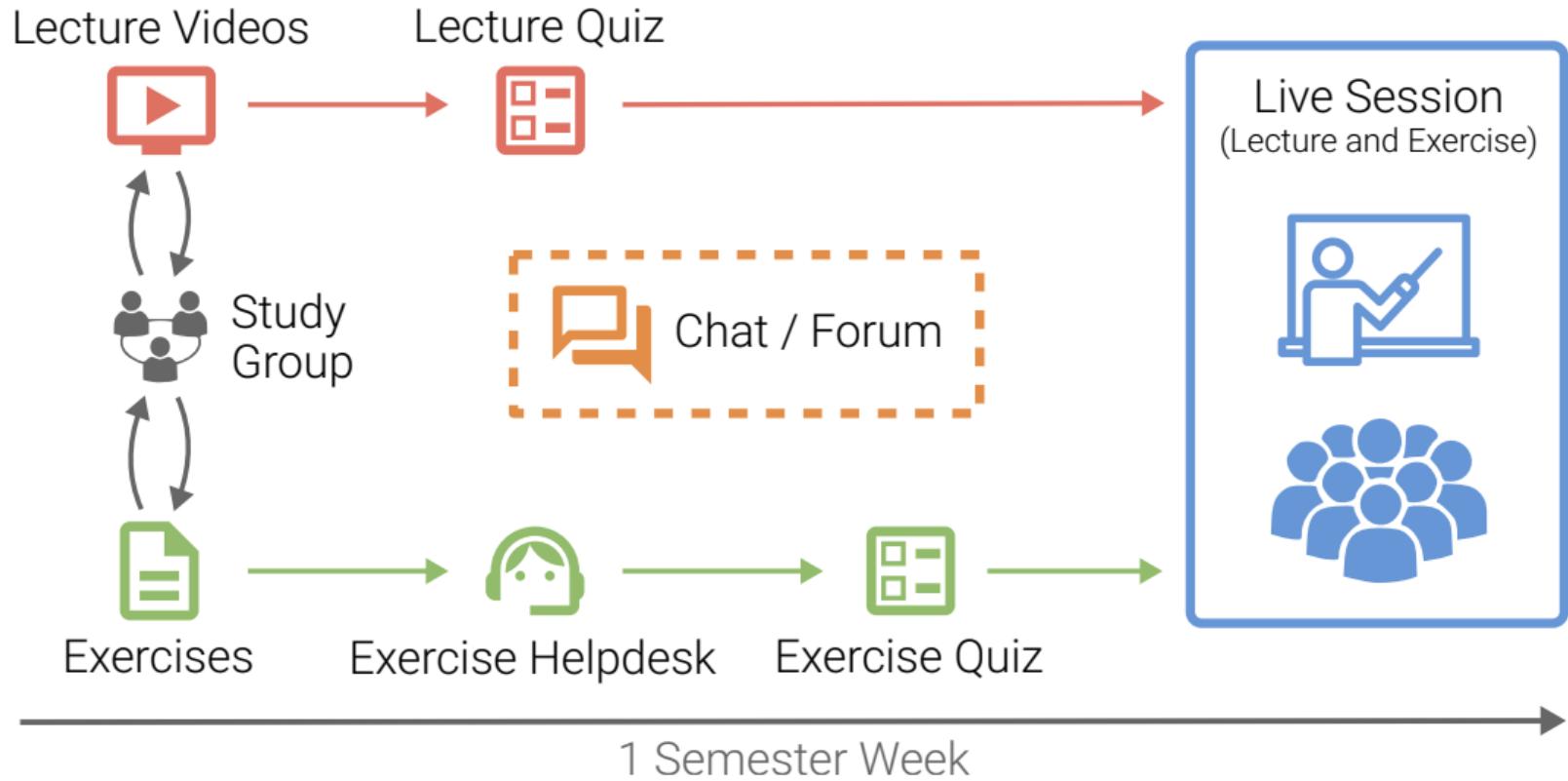


e l l i s
European Laboratory for Learning and Intelligent Systems



Old Town Square
Prague, Czech Republic

Flipped Classroom



Agenda

1.1 Organization

1.2 Introduction

1.3 History of Computer Vision

1.1

Organization

Team



Prof. Dr.-Ing.
Andreas Geiger



Markus
Flicke



Haiwen
Huang

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

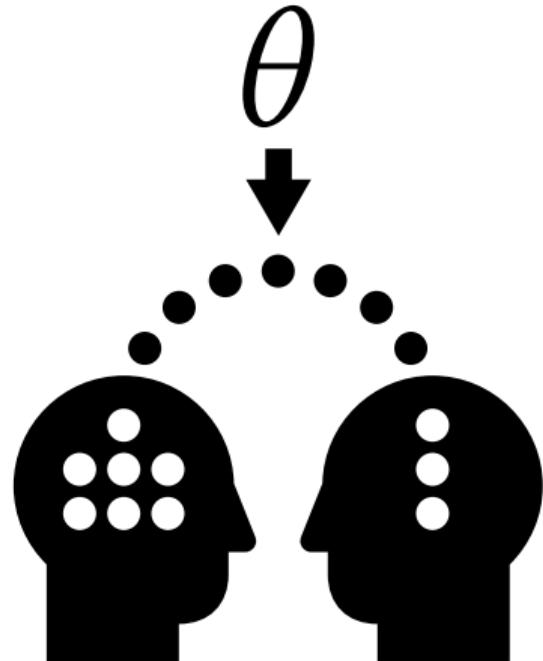
Goal: Students gain an understanding of the theoretical and practical concepts of computer vision using deep neural networks and graphical models. A strong focus is put on 3D vision. After this course, students should be able to develop and train computer vision models, reproduce research results and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

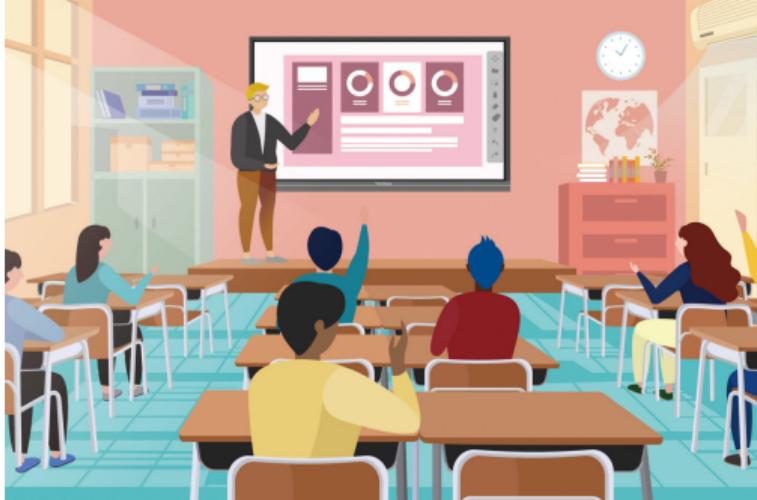
Teaching Philosophy

What is Teaching?

- ▶ Knowledge transfer (bidirectional)
- ▶ Understanding
- ▶ Learning to learn
- ▶ Critical thinking
- ▶ Engaging students
- ▶ Fostering discussion
- ▶ Apply knowledge in practice
- ▶ Having some fun along the way ..



Flipped Classroom



Regular Classroom

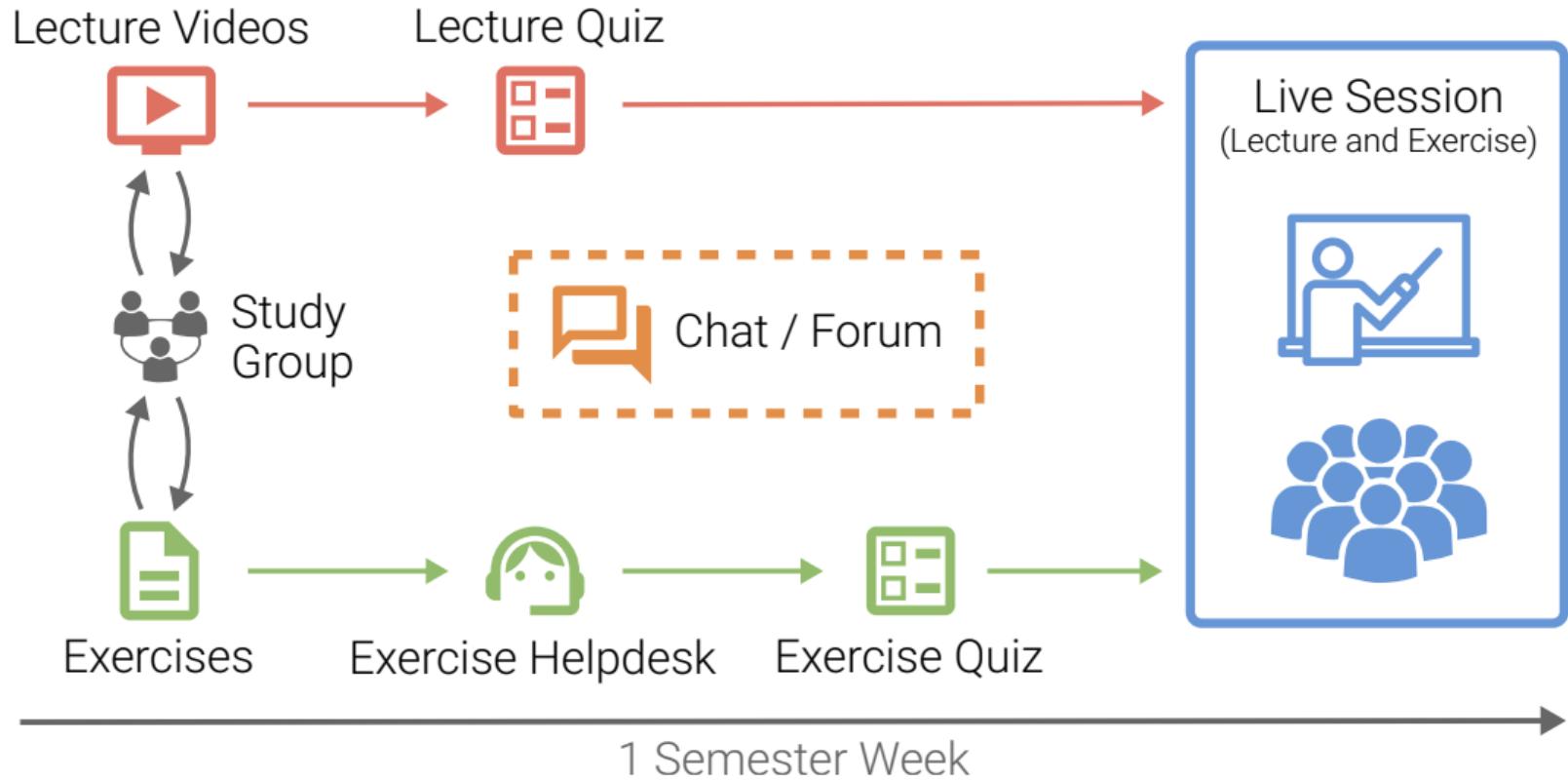
- ▶ Lecturer “reads” lecture in class
- ▶ Students digest content at home
- ▶ Little “quality time”, no fun



Flipped Classroom

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

Flipped Classroom



Flipped Classroom

More time for:

- ▶ Live interaction and discussion
- ▶ Answering questions online
- ▶ Understanding the learning progress
- ▶ Developing new formats
 - (e.g., Interactive sessions, quizzes ...)
- ▶ Implementing new tools
 - (e.g., Quiz server, chat, ...)

Flipped Classroom

Gradient Space Representation

How should we **parameterize** the normal \mathbf{n} ?

- ▶ While $\mathbf{n} \in \mathbb{R}^3$, it has only 2 degrees of freedom
- ▶ Parameterize \mathbf{n} by **neg. gradients of depth map**:

$$(p, q) = \left(-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y} \right)$$

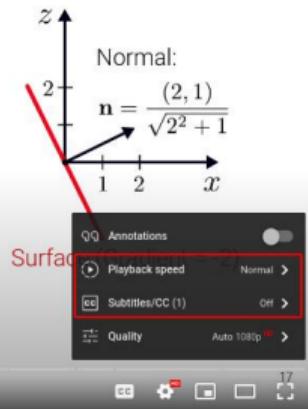
- ▶ The surface normal \mathbf{n} at pixel (x, y) is given by:

$$\mathbf{n} = \frac{(p, q, 1)^\top}{\sqrt{p^2 + q^2 + 1}}$$

Ikeuchi and Horn: Numerical Shape from Shading and Occluding Boundaries. AI, 1981



24:55 / 56:05 • Gradient Space Representation How should we parameterize the normal? >



Computer Vision — Andreas Geiger, 2021
Tübingen Machine Learning - 28 / 46

Lecture Number	Title	Duration
1	Computer Vision - Lecture 1.1 (Introduction: Organization)	3:38
2	Computer Vision - Lecture 1.2 (Introduction: Introduction)	26:53
3	Computer Vision - Lecture 1.3 (Introduction: History of...)	1:03:40
4	Computer Vision - Lecture 2.1 (Image Formation: Primitives...)	52:19
5	Computer Vision - Lecture 2.2 (Image Formation: Geometric...)	34:46
6	Computer Vision - Lecture 2.3 (Image Formation: Photometr...)	23:27
7	Computer Vision - Lecture 2.4 (Image Formation: Image...)	12:53
8	Computer Vision - Lecture 3.1 (Structure-from-Motion:...)	26:18
9	Computer Vision - Lecture 3.2 (Structure-from-Motion: Two-...)	33:55

- ▶ High quality lecture videos
- ▶ Available anytime, anywhere
- ▶ Slides available while watching
- ▶ Can be watched multiple times
- ▶ Adjustable speed, closed captions
- ▶ Questions ⇒ chat & live sessions

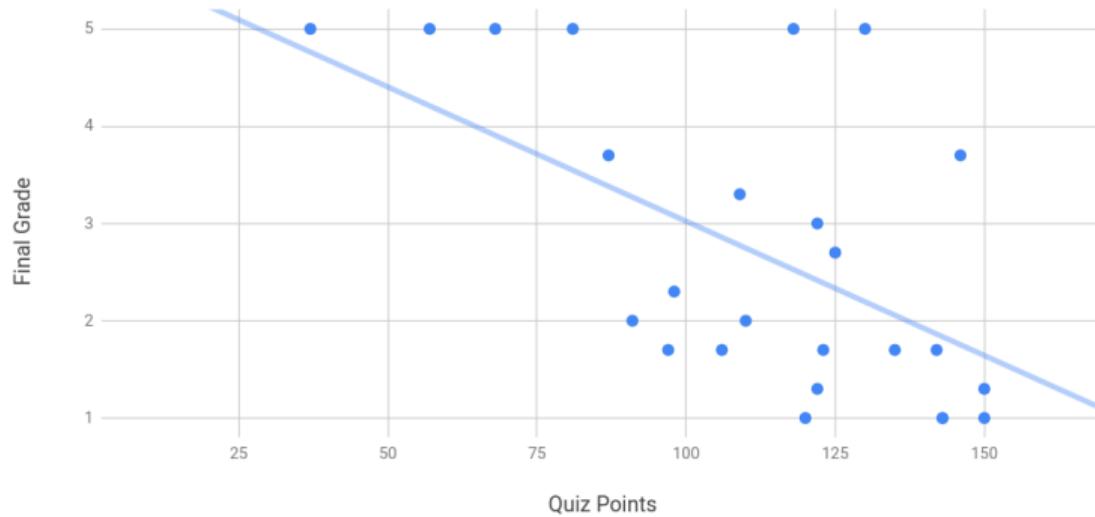
Organization

Organization

- ▶ Lectures provided in advance via YouTube
 - ▶ Students watch lecture before joining live session (mark time for this in calendar!)
 - ▶ Students are encouraged to ask questions via the lecture quiz and chat
- ▶ Live sessions (Fridays: 10:15-12:00)
 - ▶ Lecture Q&A, Exercise introduction and Q&A, live exercise solving
 - ▶ Will not be hybrid and not recorded
 - ▶ No tutorials by student assistants (this is a Master course), instead we provide:
 - ▶ Plenary feedback via Q&A sessions, individual help via weekly helpdesk (make use of it)
 - ▶ Fast (<24h) feedback via chat (make use of it)
- ▶ Exam
 - ▶ Written (date and time available on course website)
 - ▶ A bonus can be obtained upon participation in the quizzes
- ▶ Course Website with YouTube, Slide, Exercise, ILIAS, Quiz, Chat and Zoom links:
<https://uni-tuebingen.de/de/203146>
- ▶ **Register** to ILIAS and Lecture Quiz Server **today** to participate in this course!

Exercises

- Exercises are offered every 2 weeks, 6 assignments in total [!\[\]\(73b059f51b6349c47bd4937a81c667dc_img.jpg\) Slides / Exercises](#)
- 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** for the final exam ⇒ continuous participation



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 via zoom to answer your questions
- ▶ In the past, we graded exercises, but students rather preferred solutions
 ⇒ We will not grade exercises, but hand out and discuss the solutions instead
- ▶ 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
- ▶ Students gain a bonus for the exam
- ▶ Students collect AI generated Pokemon!
- ▶ Answers may not be shared
- ▶ Participation is voluntary
- ▶ Past quizzes only visible to participants
- ▶ Deadline: Thursdays, 9pm
- ▶ Ask if questions are ambiguous
- ▶ **Register today** with your student email

AVG Lecture Quiz Server

Andreas Geiger | a.geiger@uni-tuebingen.de | Admin | Edit | Logout

Edit Assignment

Each correct answer yields +1 point. Each incorrect answer yields -1 point. Meaningful student questions yield +1 point. If you wish, you may answer only a subset of the questions. The minimum overall score that can be attained in this assignment is 0 points.

Computer Vision | Summer 22 | Lecture 2

Opening: 21 Apr 2022 (21:00) - Closing: 28 Apr 2022 (21:00)

1. What is an ideal point?

- A nonhomogeneous point
- A point at infinity

2. What is the dimensionality of a 4D point represented in homogeneous coordinates?

3. How many degrees of freedom has a 3D line?

4. Affine transformations preserve

- parallel lines
- lengths
- angles

5. The diameter of a telecentric lens should be

- about as large as the object/scene
- significantly larger than the object
- significantly smaller than the object

6. Increasing the depth of a 3D point in perspective projection

- increases the (x,y) pixel coordinates
- decreases the (x,y) pixel coordinates
- does not affect the (x,y) pixel coordinates

7. Perspective projection

- is linear when using homogeneous coordinates
- is linear when using inhomogeneous coordinates

8. Assume a point light source. Diffuse materials

- scatter light uniformly in all directions at any given surface point
- scatter more light into the mirror direction than into other directions at any given surface point
- scatter the same amount of light independent of the surface normal





OpenAI ChatGPT 4.0

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

AVG Lecture Quiz Server: Student Questions

10. Do you have a question or comment about this lecture which you like to discuss during the live session?

Slide number:

(required field; enter 0 for a general question)

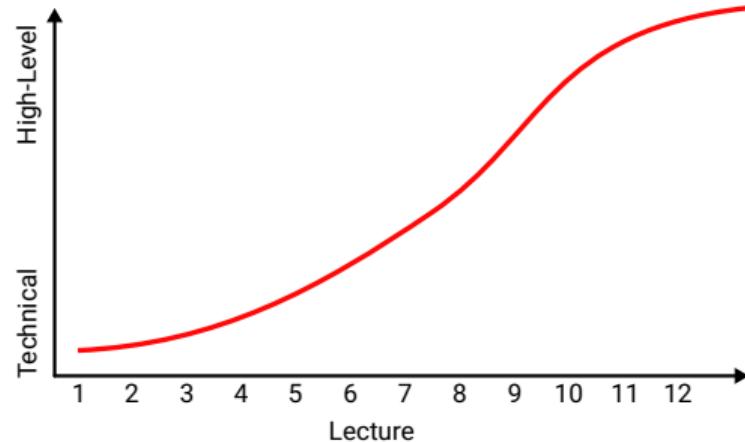
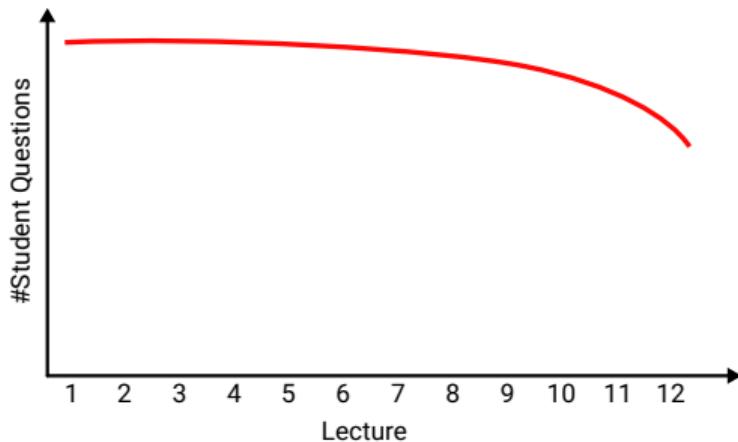
Hint: Question must comprise between 20 and 200 characters.

0/200

We appreciate your feedback as well:

- ▶ Questions stipulate thinking and provide valuable feedback to us
- ▶ Students are encouraged to ask a question regarding the lecture
- ▶ We will select some of the questions for discussion during the live session
- ▶ Also dare to ask “easy” questions, many others will be wondering as well
- ▶ Our goal is to build a relationship of trust and openness in this class!

AVG Lecture Quiz Server: Student Questions



- ▶ Dare to ask any kind of question
- ▶ All questions are valuable: technical and high-level
- ▶ But do not only ask high-level questions
- ▶ Dare to ask “easy” and “technical” questions - they are not easy
- ▶ Your questions will not negatively influence your exam grade

Helpdesk

- ▶ We offer a weekly zoom helpdesk where our TAs will provide individual support
- ▶ Ask any question about the exercise
- ▶ Share your screen to show a problem
- ▶ Start working on your exercise early!
- ▶ Let's do a quick time poll ..

Chat



Andreas Geiger | 14 Apr. 2023 (17:16)

This is the chatroom for our lecture.

1
1



Andreas Geiger | 14 Apr. 2023 (17:18)

It supports multiple threads which can be associated with specific lectures or quiz questions.

It also supports **markdown** and latex:

$$x = \int_0^1 \frac{5}{20} x \, dx$$



Send Message (Ctrl+Enter)

Live Sessions

The “Hybrid Lecture Paradox”

- ▶ Hybrid lectures are the future!
Students can join from anywhere!
- ▶ However, as it turns out:
 - ▶ Students only **join online**
 - ▶ But they prefer to **join live**
 - ▶ ???
- ▶ Result: Severely reduced interaction
(bad audio, no faces, no spontaneity)
- ▶ Conclusion:
Live sessions will be live. Period.

Live Sessions

1. Lecture Q&A (45 min)

- ▶ Discussion of hard quiz questions
- ▶ Discussion of student questions
(join to hear your answer!)
- ▶ Discussion of live questions

2. Exercise Q&A (45 min)

- ▶ Exercise introduction
- ▶ Exercise Q&A and live solving
(start solving before joining!)
- ▶ Exercise discussion



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 on ILIAS



Computer Vision

Exam, Summer 2021

Tübingen, August 10, 2021

Student Number:	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Seat Number:	<input type="text"/>				
First Name:	<input type="text"/>				
Last Name:	<input type="text"/>				
Date of Birth:	<input type="text"/>				
Program of Study (e.g., "Informatik"):	<input type="text"/>				
Master	<input type="checkbox"/>	Bachelor	<input type="checkbox"/>		

Task	1	2	3	4	5	Σ
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**.

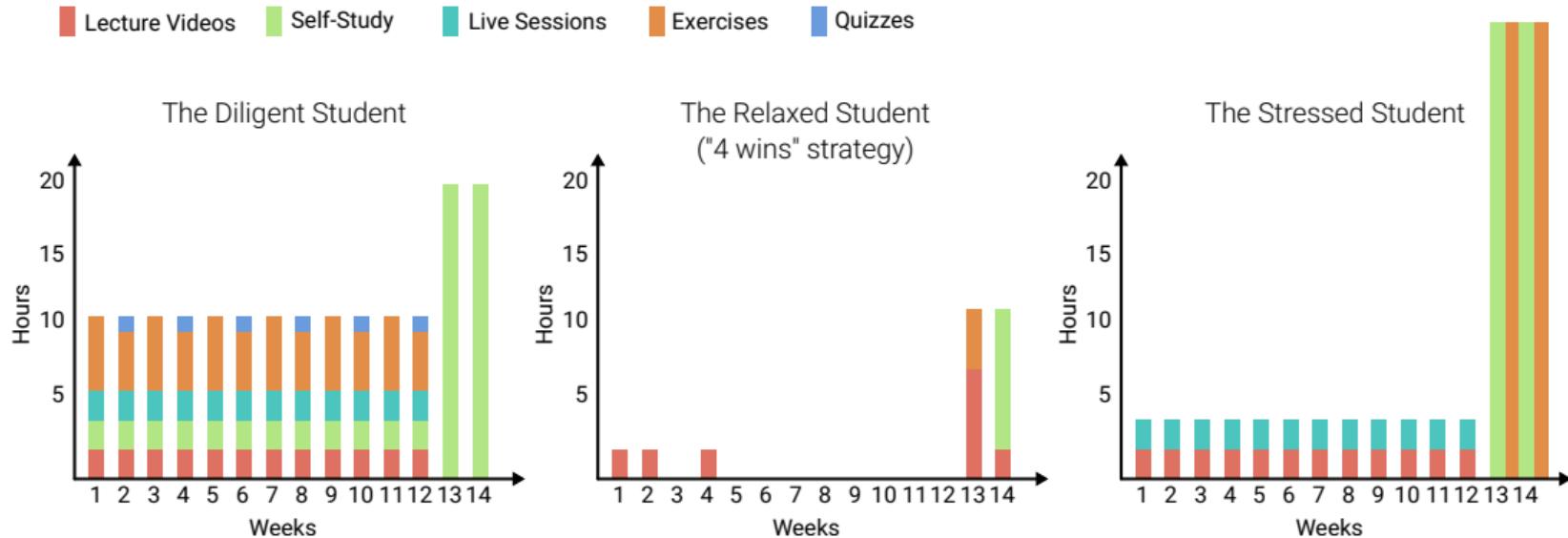
Good Luck!

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 = $\left\lfloor \frac{\text{lec. points}}{36} \right\rfloor + \left\lfloor \frac{\text{ex. points}}{20} \right\rfloor$
- ▶ 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>
- ▶ Inofficial 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: $\mathbf{A}^T, \mathbf{A}^{-1}, \text{Tr}(\mathbf{A}), \det(\mathbf{A}), \mathbf{A} + \mathbf{B}, \mathbf{AB}, \mathbf{Ax}, \mathbf{x}^\top \mathbf{y}$
- ▶ 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: $x \perp\!\!\!\perp y | z \Leftrightarrow p(x, y|z) = p(x|z)p(y|z)$
- ▶ Expectation: $\mathbb{E}_{x \sim p} [f(x)] = \int_x p(x)f(x)dx$
- ▶ Variance: $\text{Var}(f(x)) = \mathbb{E} [(f(x) - \mathbb{E}[f(x)])^2]$
- ▶ Distributions: Bernoulli, Categorical, Gaussian, Laplace
- ▶ Entropy: $H(x)$, KL Divergence: $D_{KL}(p\|q)$

Prerequisites



$$\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \quad \Rightarrow \quad \mathbf{A} \mathbf{B} = \begin{pmatrix} 6 & 7 \\ 5 & 5 \end{pmatrix}$$

$$\hat{\mathbf{w}}_{ML} = \operatorname{argmax}_{\mathbf{w}} \underbrace{\sum_{i=1}^N \log p_{model}(y_i | \mathbf{x}_i, \mathbf{w})}_{\text{Log-Likelihood}}$$

$$\mathbf{A}^{-1} \mathbf{A} \mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$$

$$\mathbf{I}_N \mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$$

$$\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$$

$$\sum_{i=1}^M \frac{\partial f}{\partial g_i} \frac{dg_i}{dx}$$

$$\mathbf{x}^T \mathbf{y} = (\mathbf{x}^T \mathbf{y})^T = \mathbf{y}^T \mathbf{x} = \mathbf{x}^T \mathbf{y}$$

Math for Deep Learning

Prof. Dr. Andreas Geiger

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!

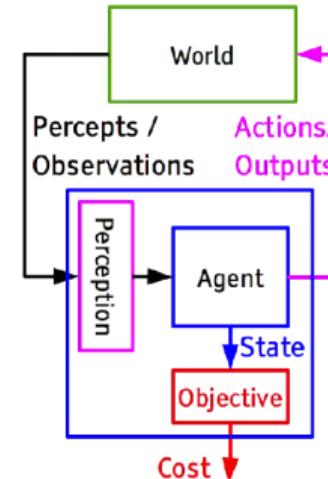
Looking forward to our discussions

Artificial Intelligence

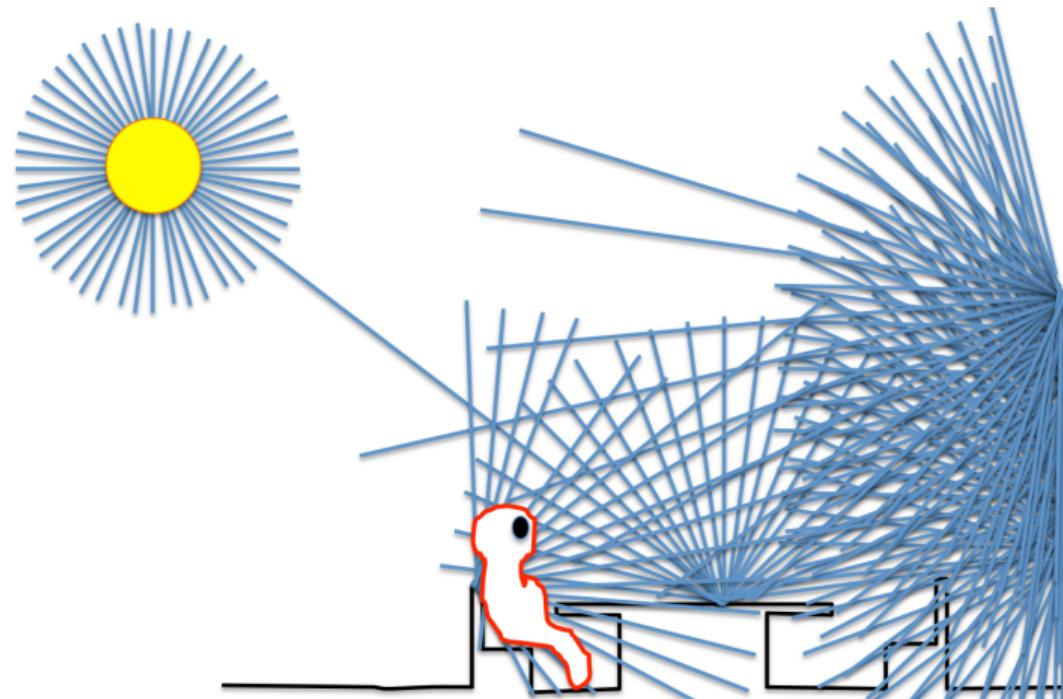
"An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves."

[John McCarthy]

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



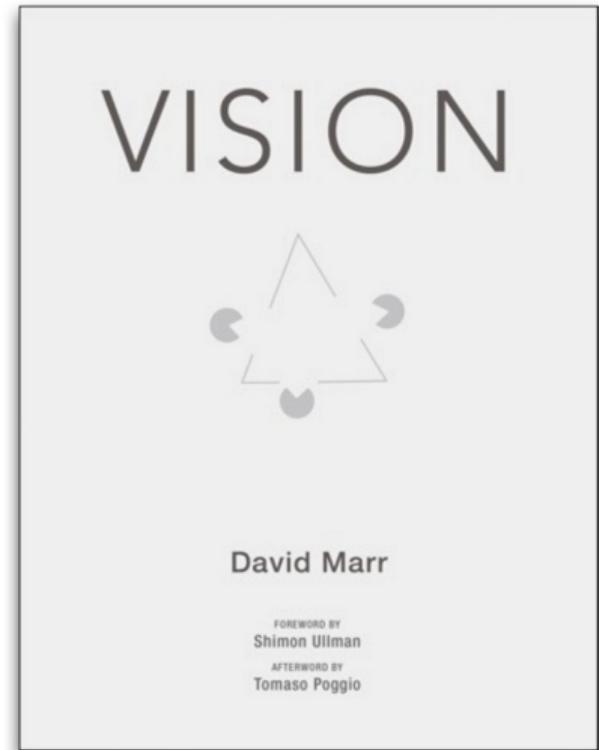
Computer Vision



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

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.

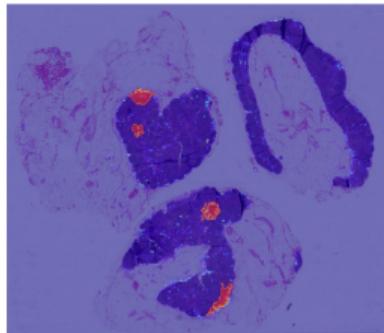


Computer Vision Applications

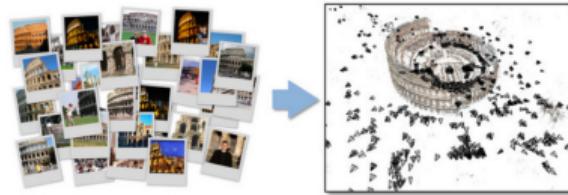
Robotics



Medical applications



3D modeling



Driving



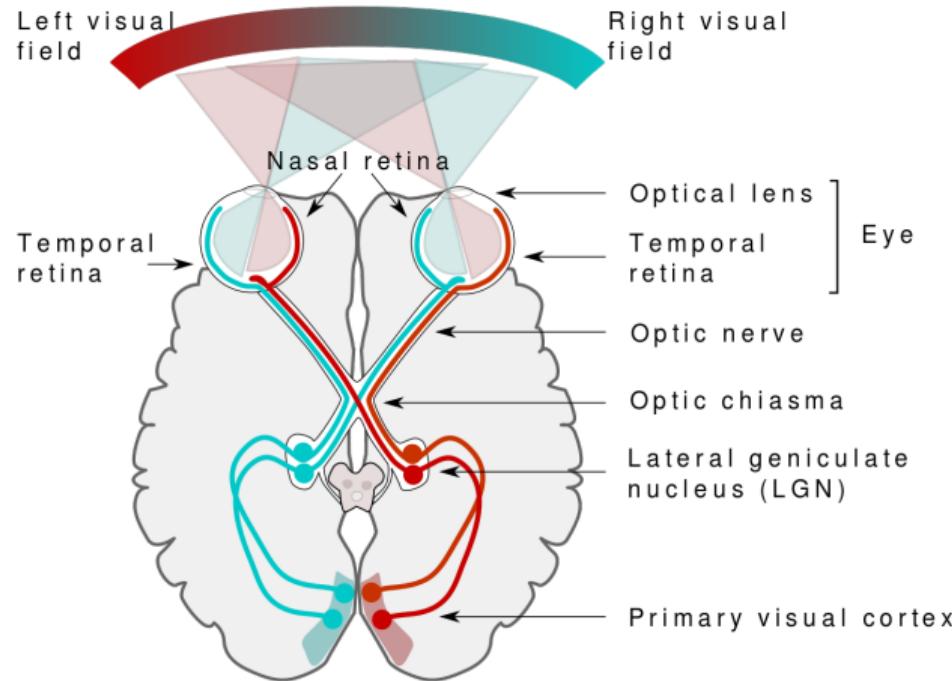
Mobile devices



Accessibility

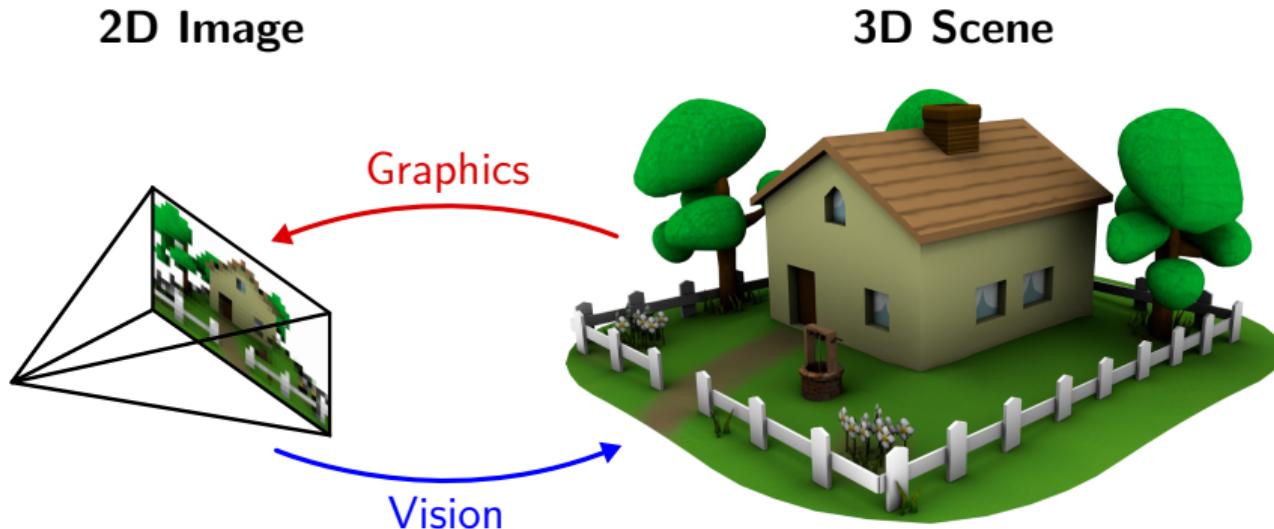


Computer Vision vs. Biological Vision



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

Computer Vision vs. Computer Graphics



Pixel Matrix

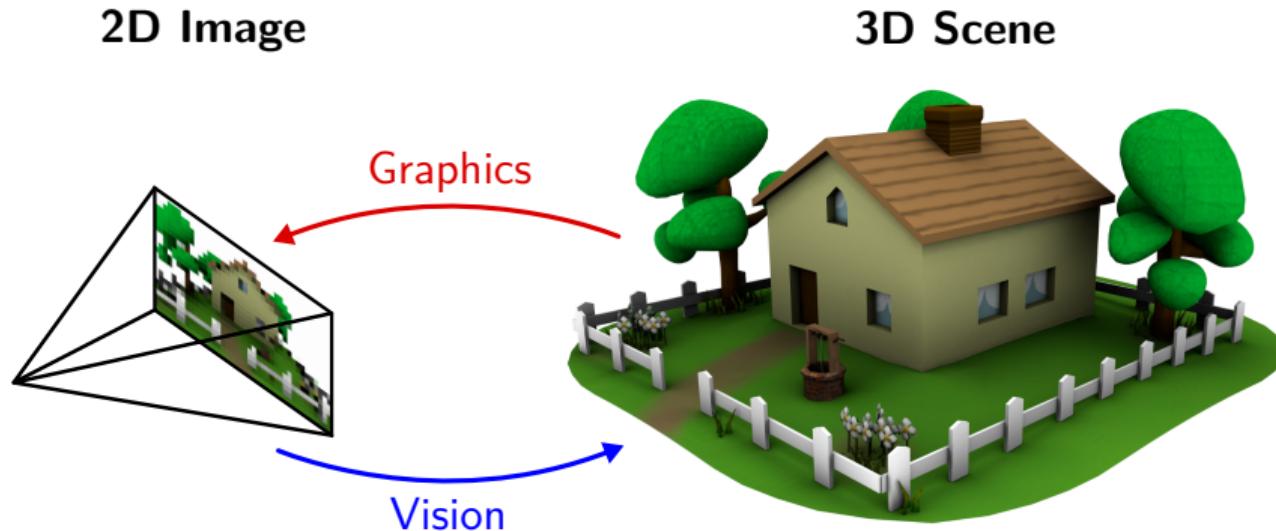
217	191	252	255	239
102	80	200	146	138
159	94	91	121	138
179	106	136	85	41
115	129	83	112	67
94	114	105	111	89

Objects Material

Shape/Geometry Motion

Semantics 3D Pose

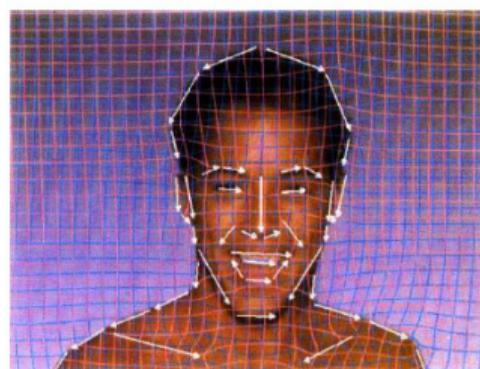
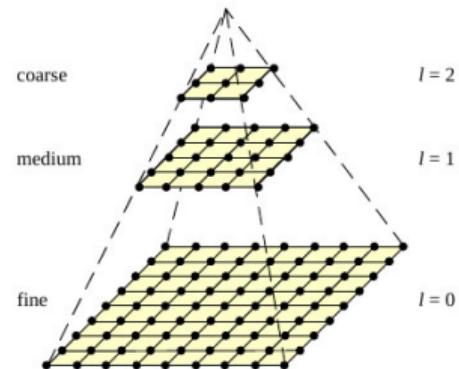
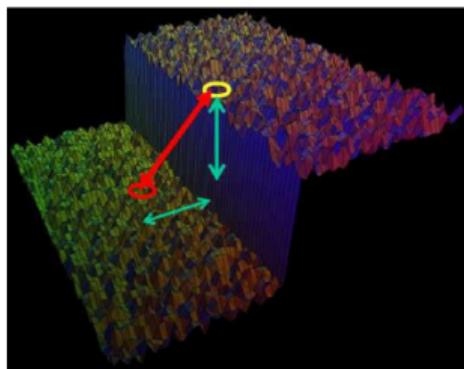
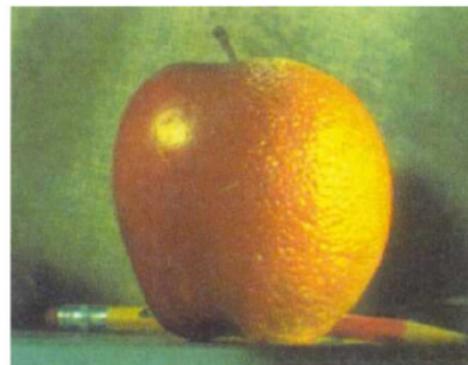
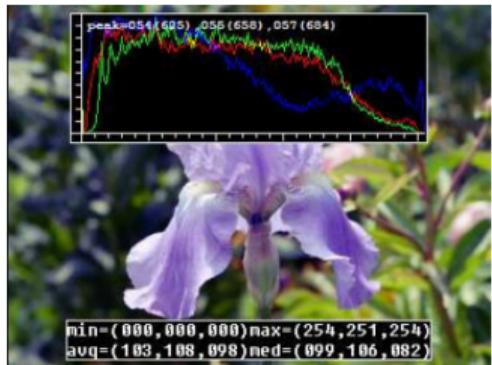
Computer Vision vs. Computer Graphics



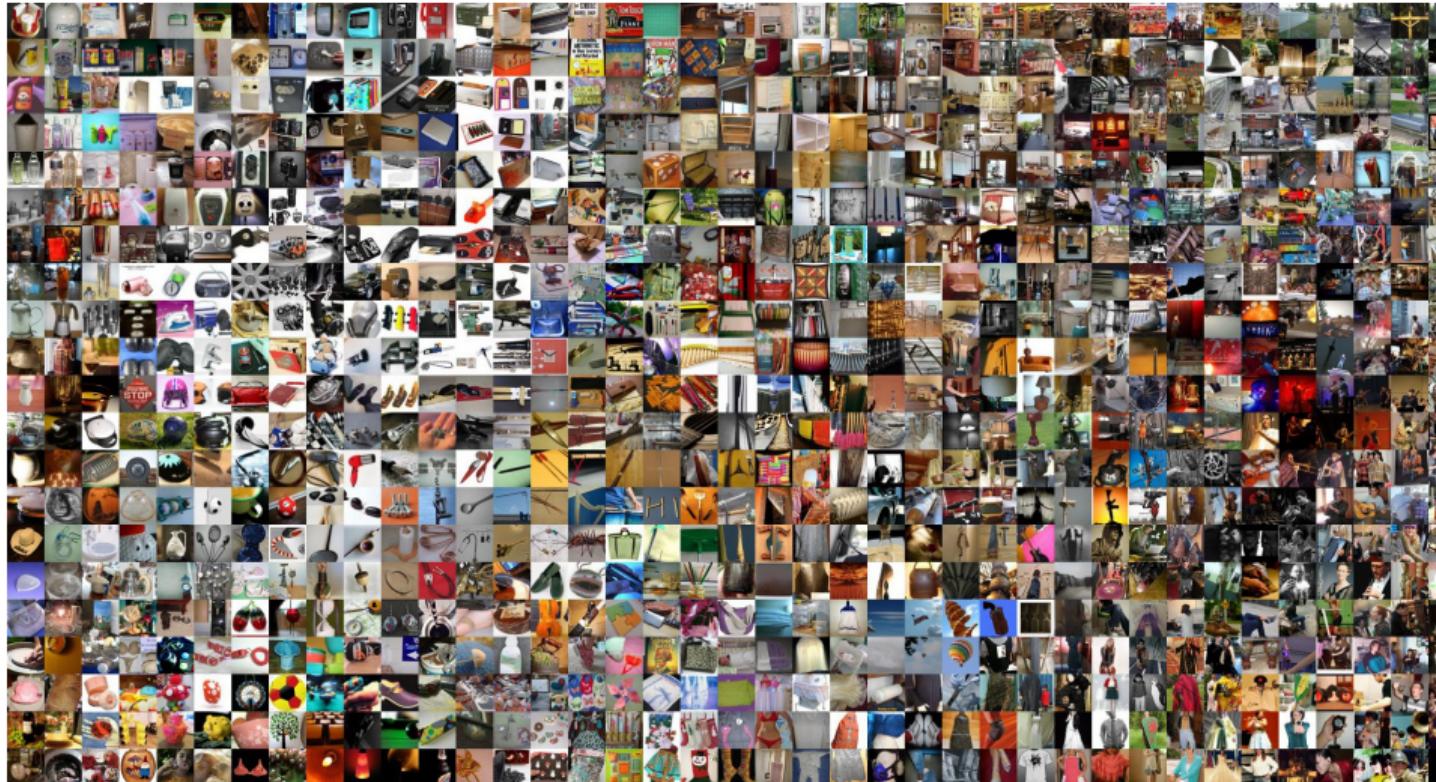
Computer Vision is an ill-posed inverse problem:

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

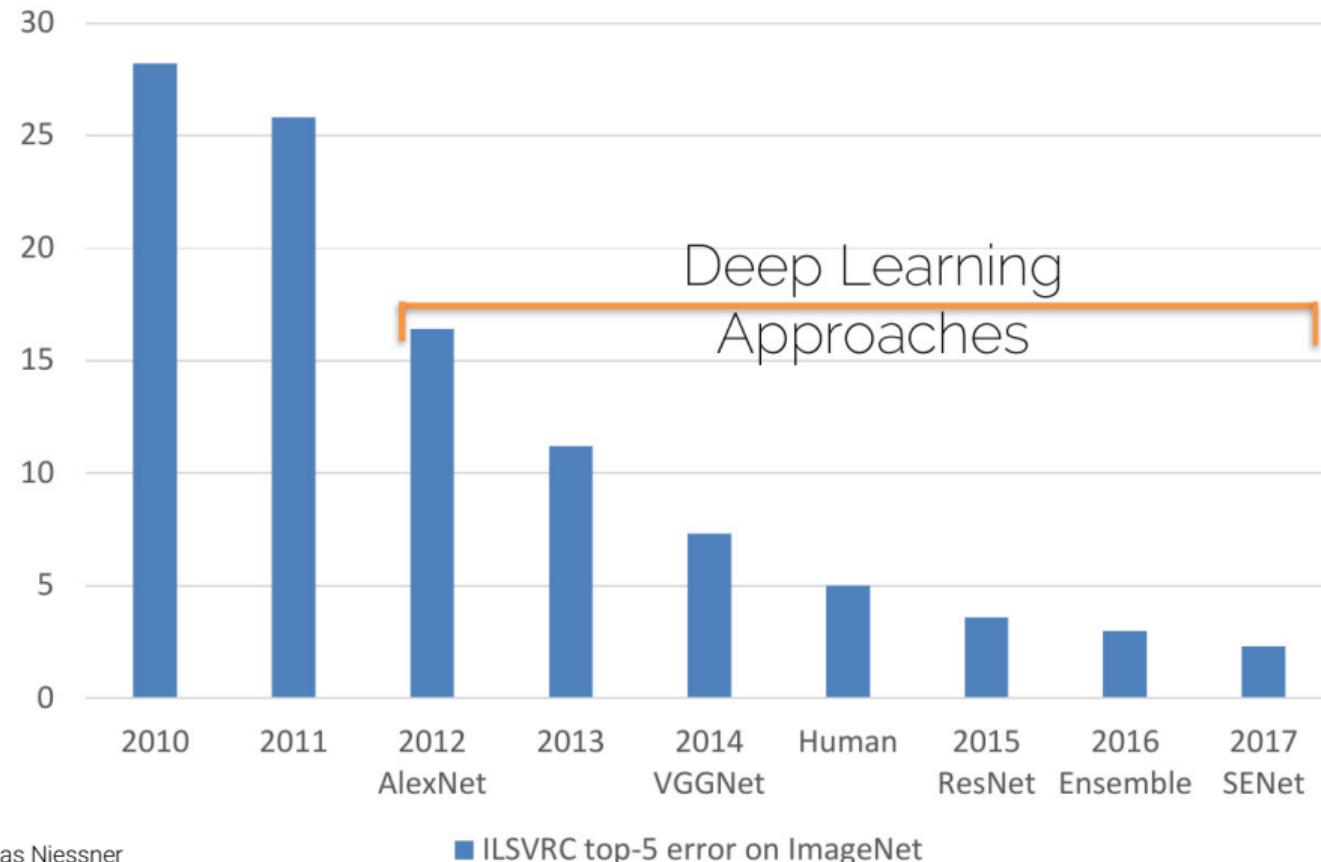
Computer Vision vs. Image Processing



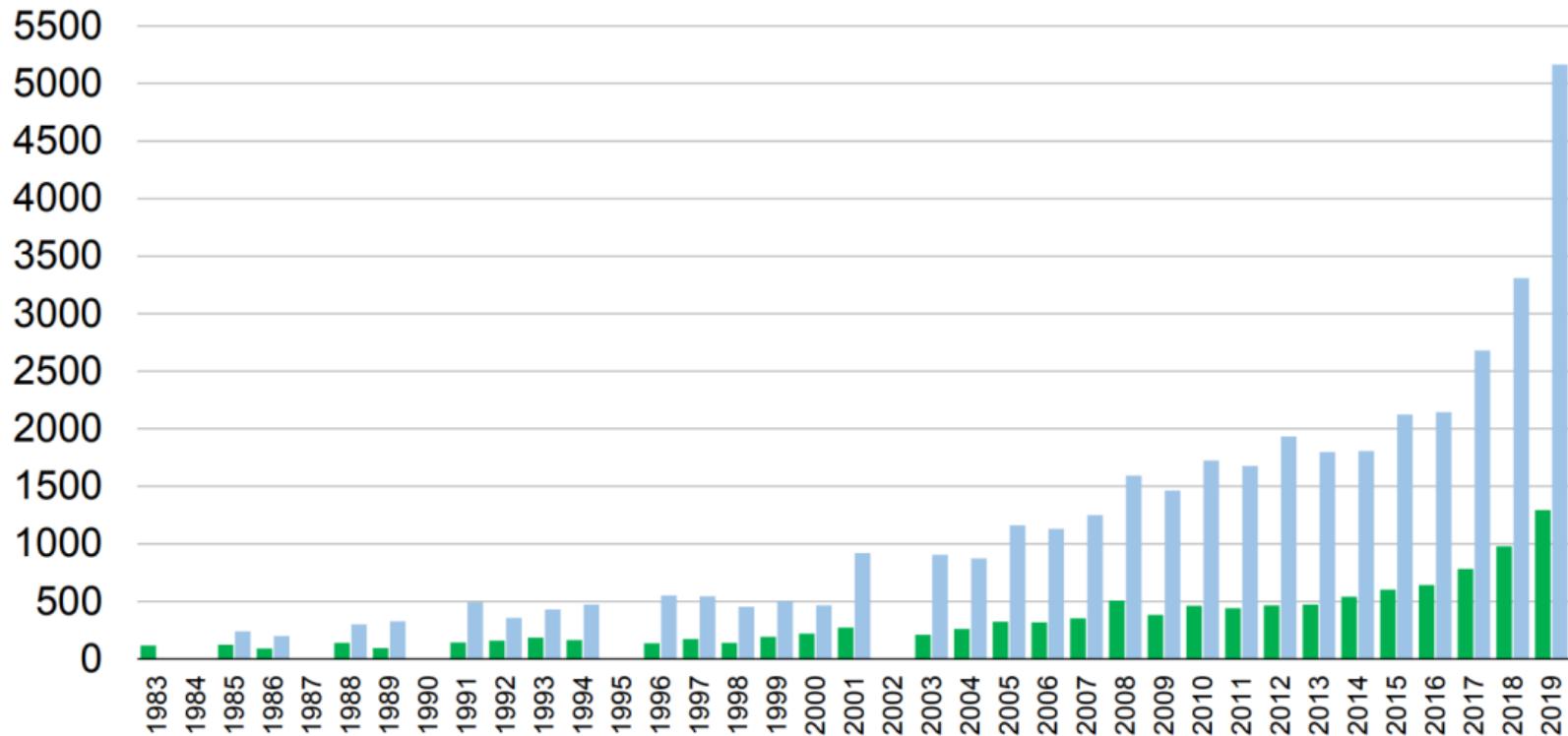
Computer Vision vs. Machine Learning



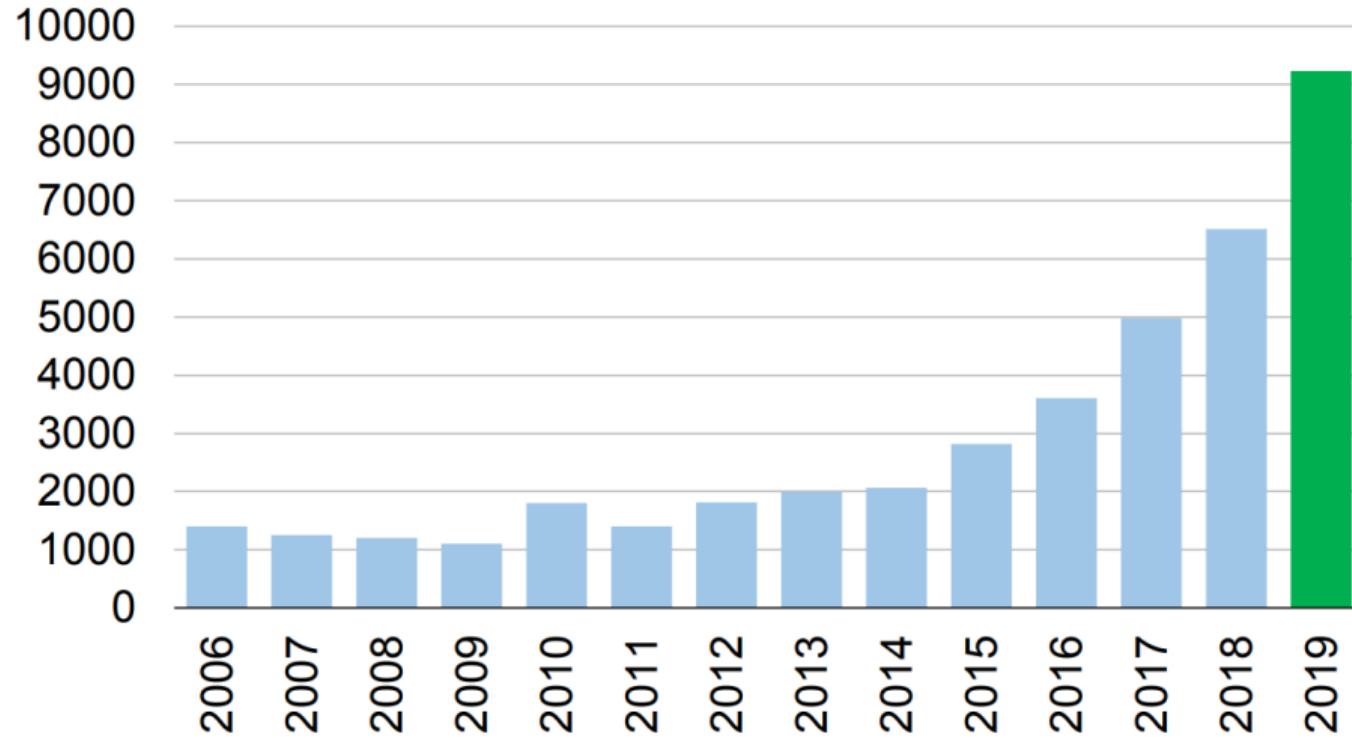
The Deep Learning Revolution



CVPR Submitted and Accepted Papers



CVPR Attendance



CVPR 2019 Sponsors

Diamond Sponsors



Platinum Sponsors



Gold Sponsors



Silver Sponsors



Why is Visual Perception hard?

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".

Why is Visual Perception hard?



Why is Visual Perception hard?

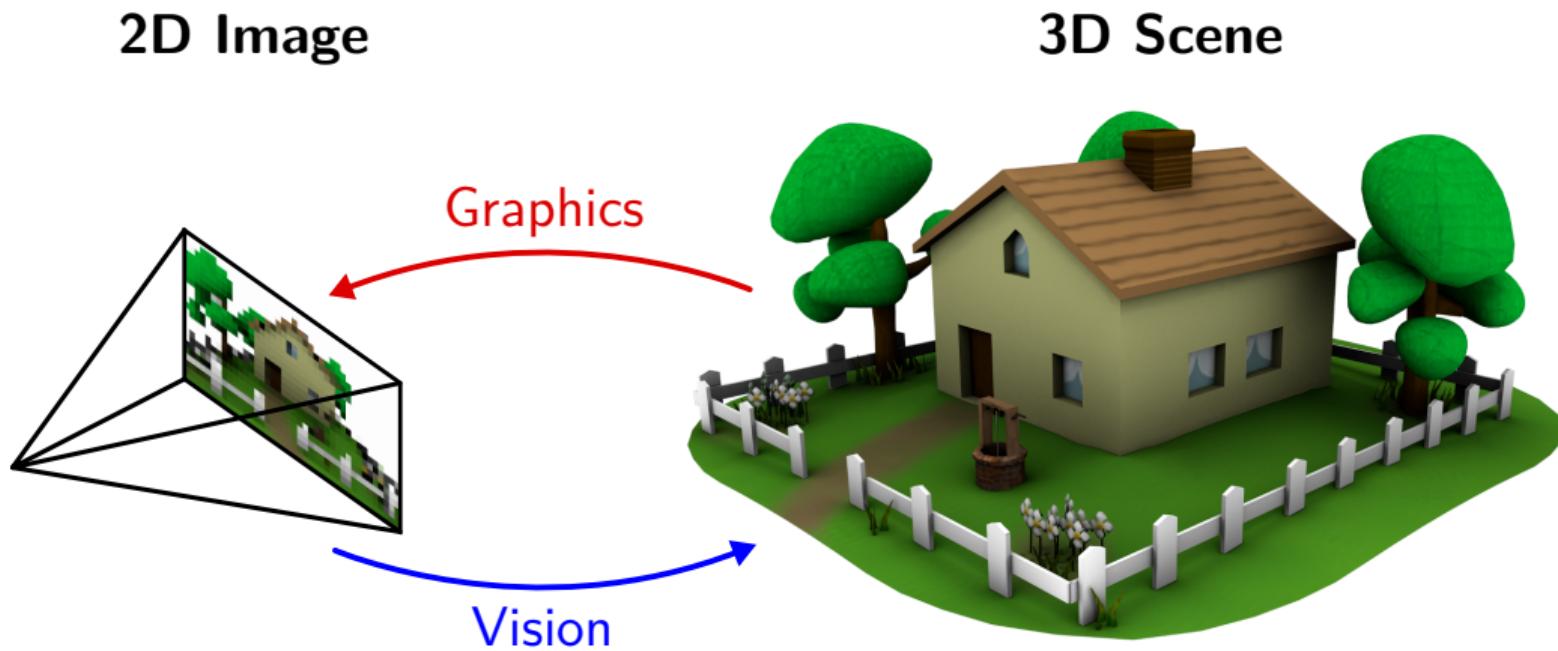


What we see

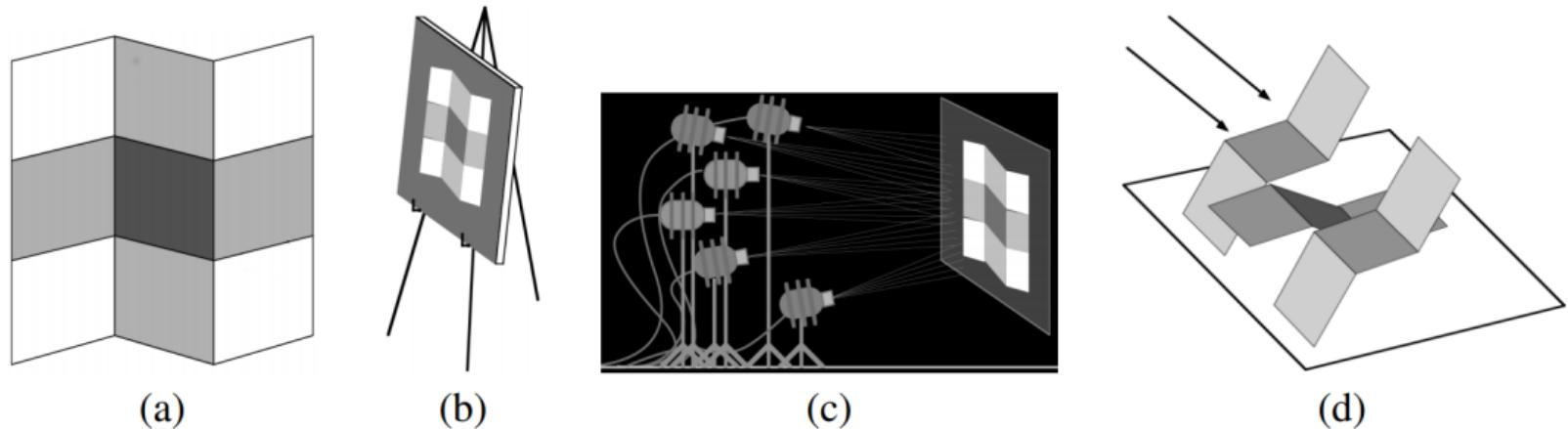
200	133	110	103	117	90	47	30	32	79	66	65
197	122	123	138	98	100	46	45	22	11	43	55
140	116	165	159	90	56	58	47	26	13	54	102
132	148	119	108	123	57	64	46	21	22	79	94
125	121	80	143	101	55	61	38	20	21	81	65
50	71	74	63	52	39	41	39	32	26	97	66
51	59	62	44	40	40	36	28	27	31	29	44
59	62	70	50	48	35	34	35	26	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	163	156	63	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	60	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	69	132	169	183	128
31	25	62	54	47	44	81	190	227	231	206	155
44	66	99	72	67	63	69	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	96	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	162	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 the computer sees

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



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



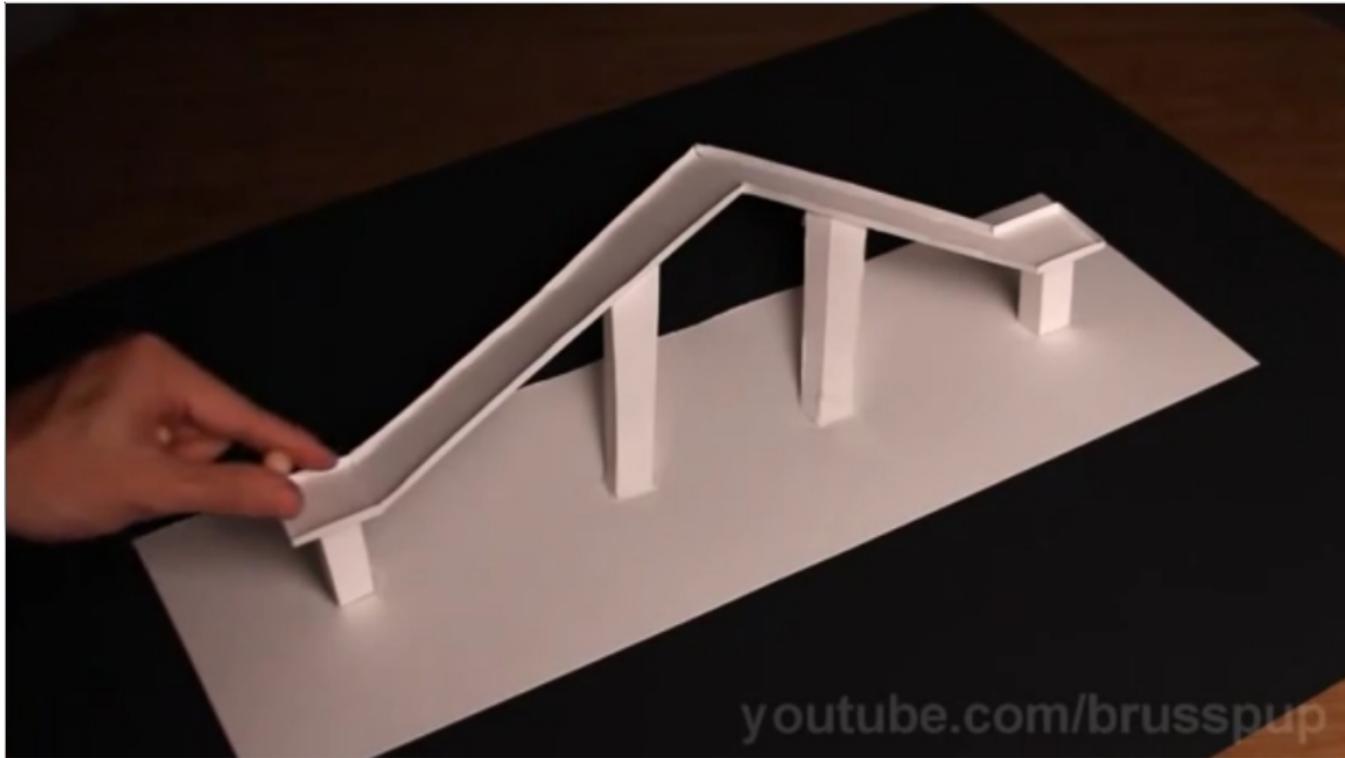
Adelson and Pentland's workshop metaphor:

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.

Ames Room Illusion

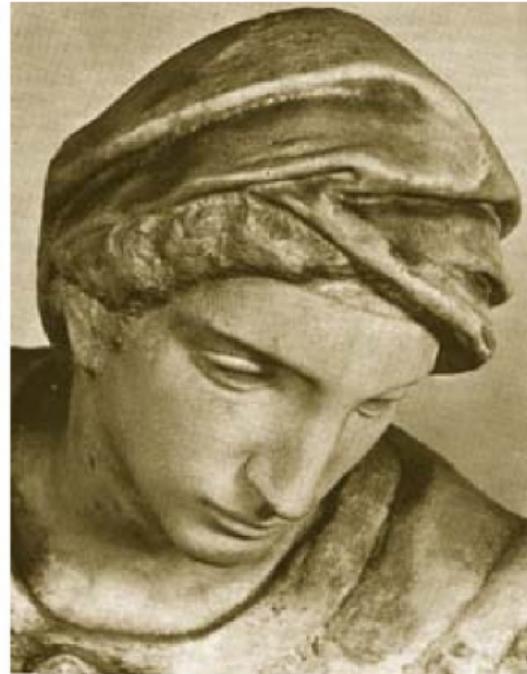


Perspective Illusion



youtube.com/brusspup

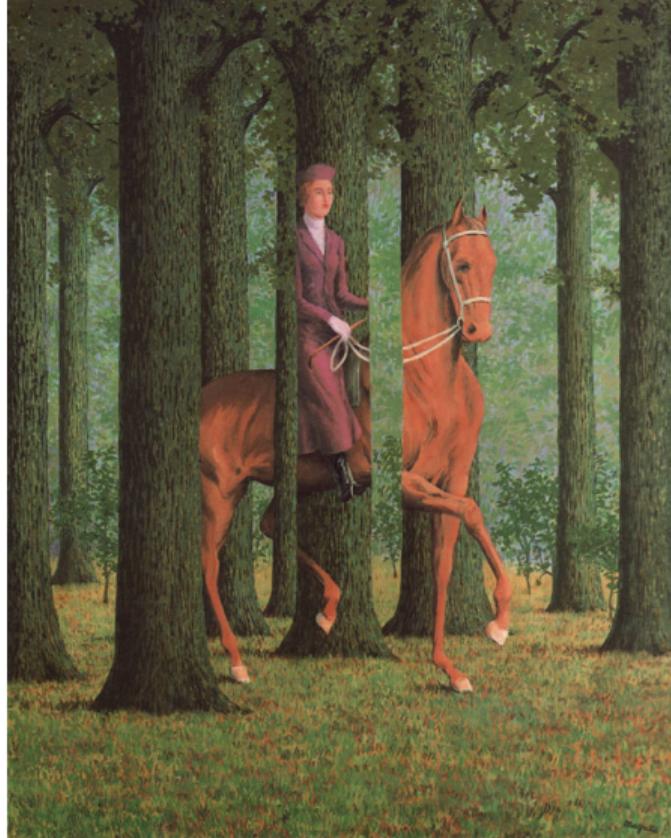
Challenges: Viewpoint Variation



Challenges: Deformation

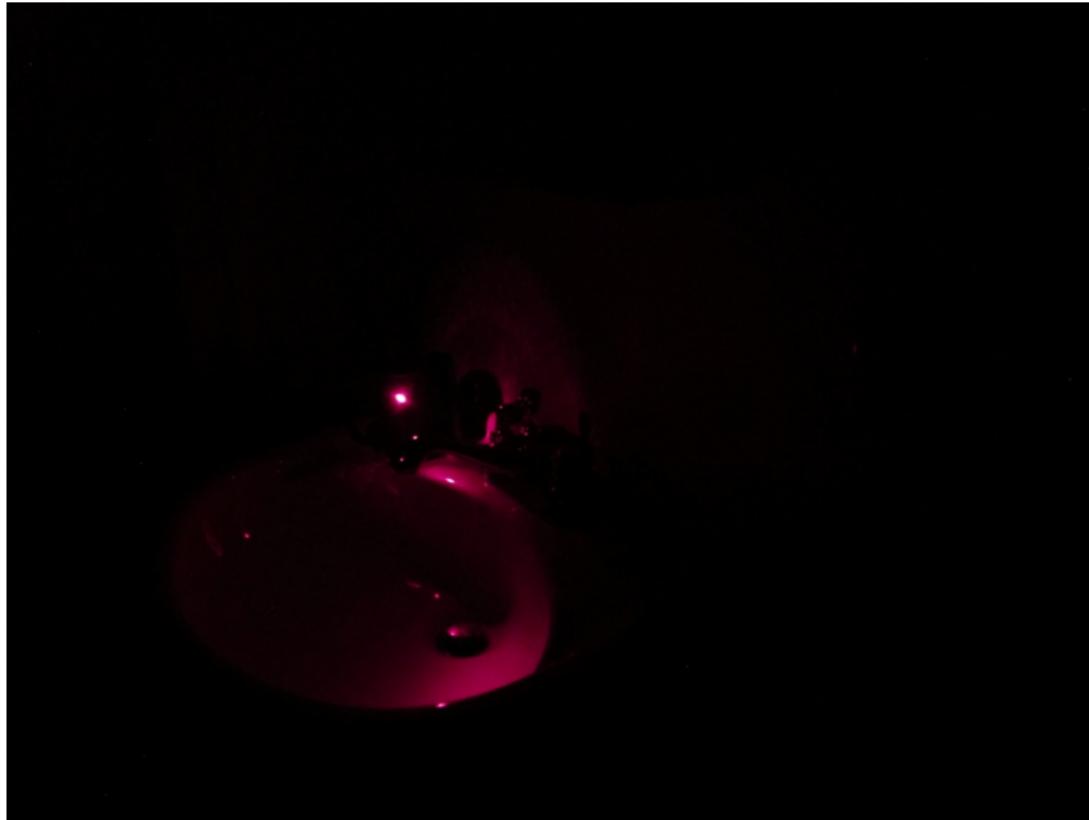


Challenges: Occlusion



René Magritte (1957)

Challenges: Illumination



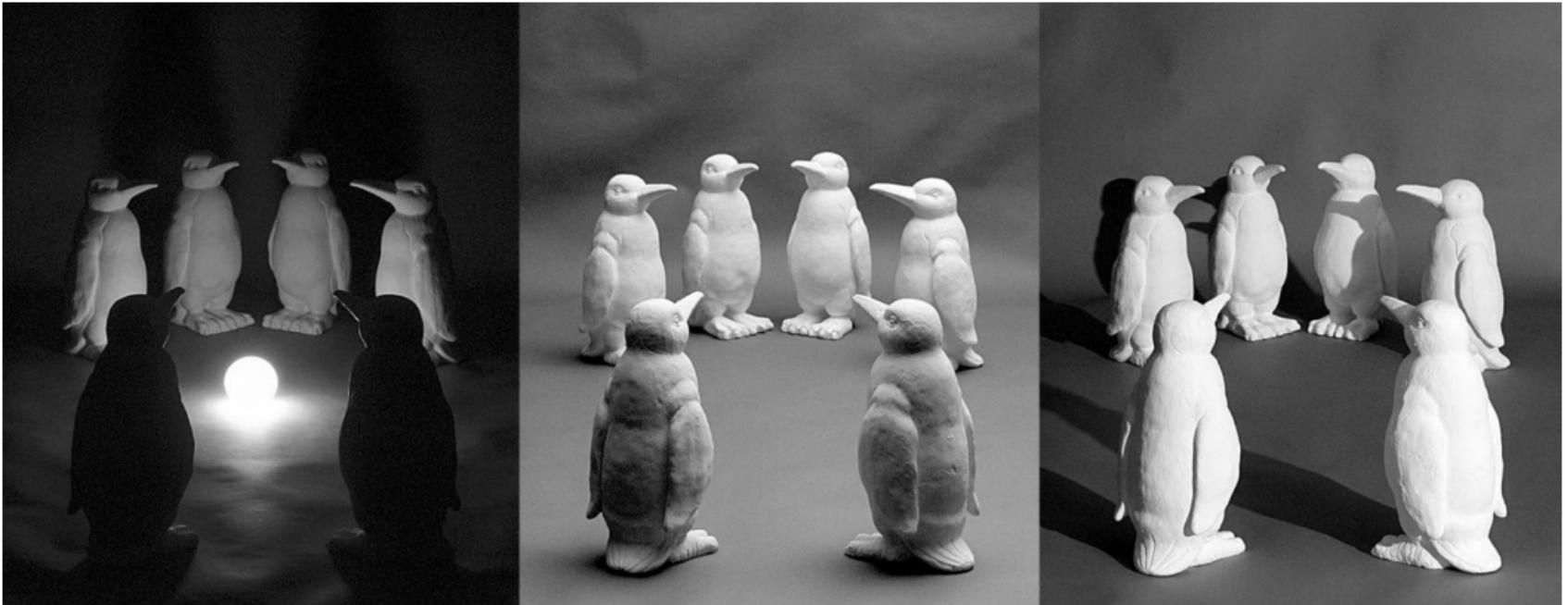
Challenges: Illumination



Challenges: Illumination



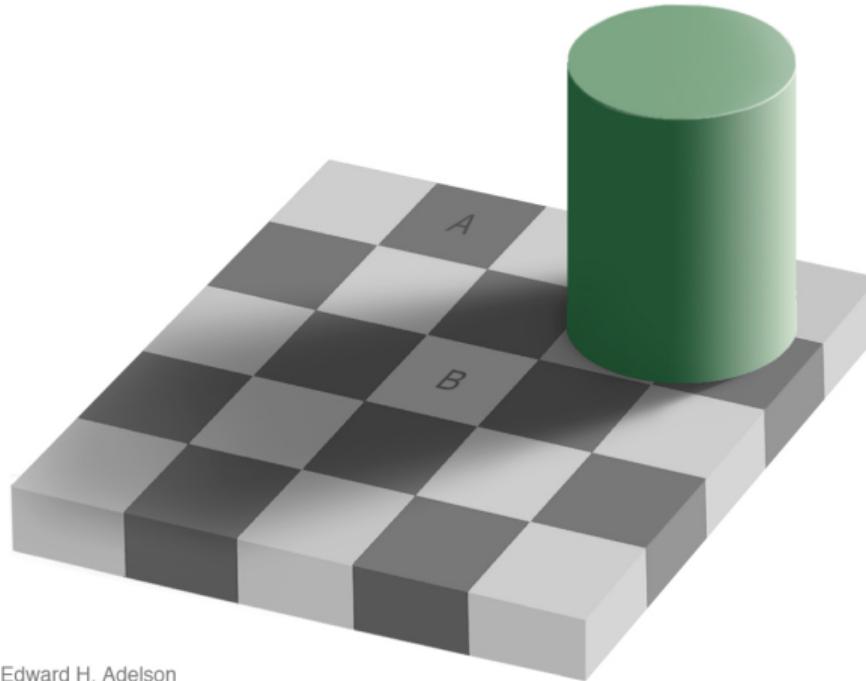
Challenges: Illumination



Challenges: Motion



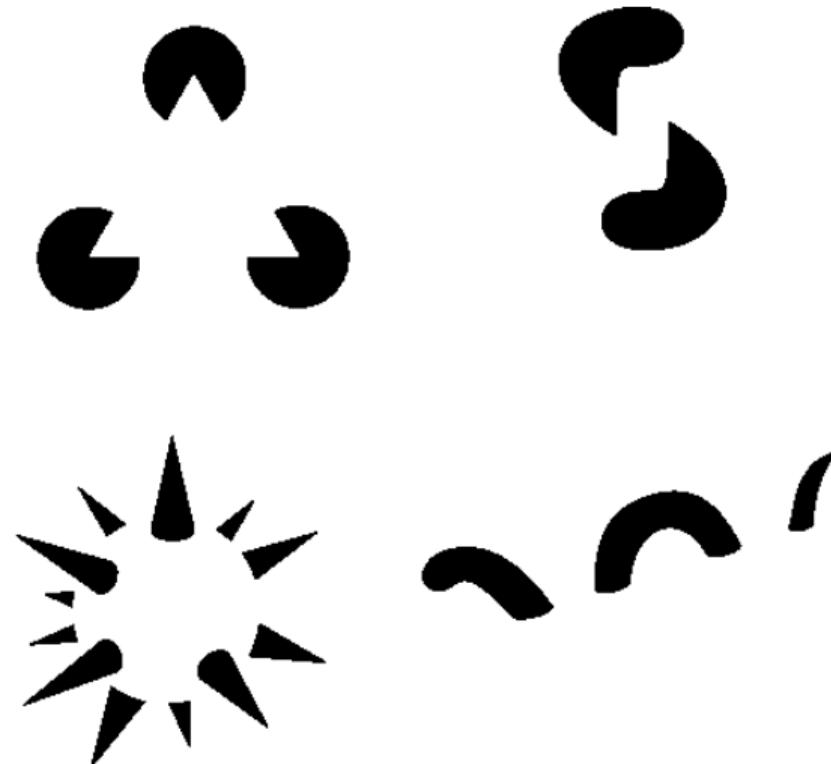
Challenges: Perception vs. Measurement



Edward H. Adelson

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

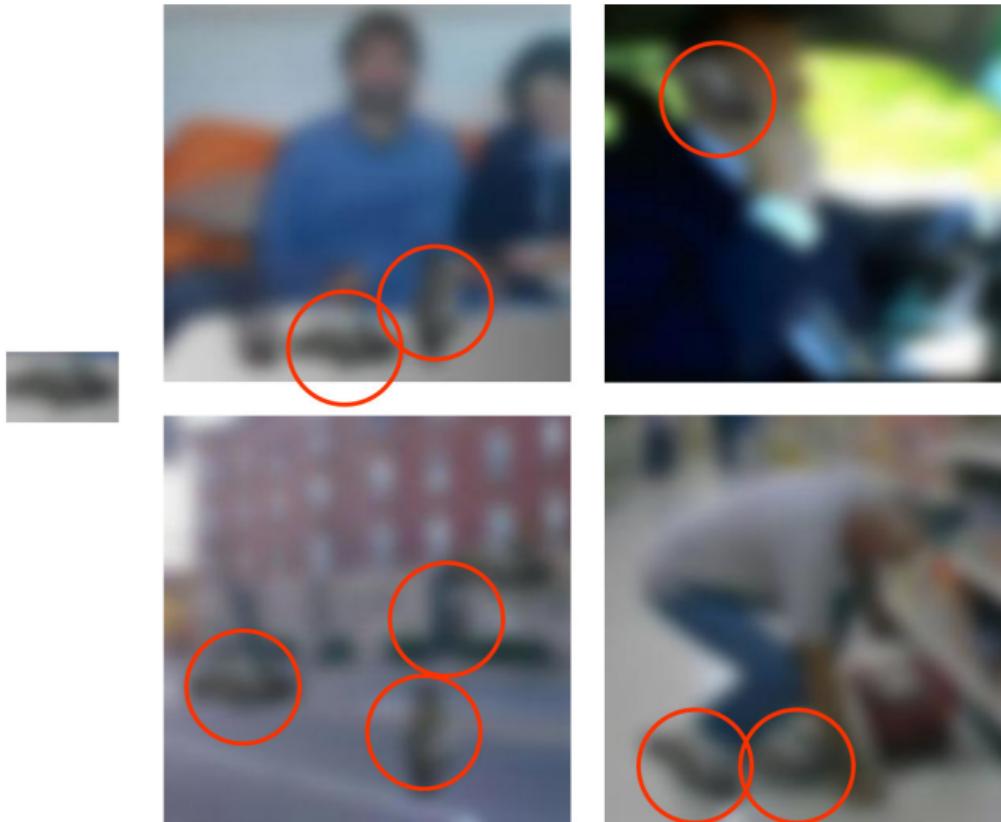
Challenges: Perception vs. Measurement



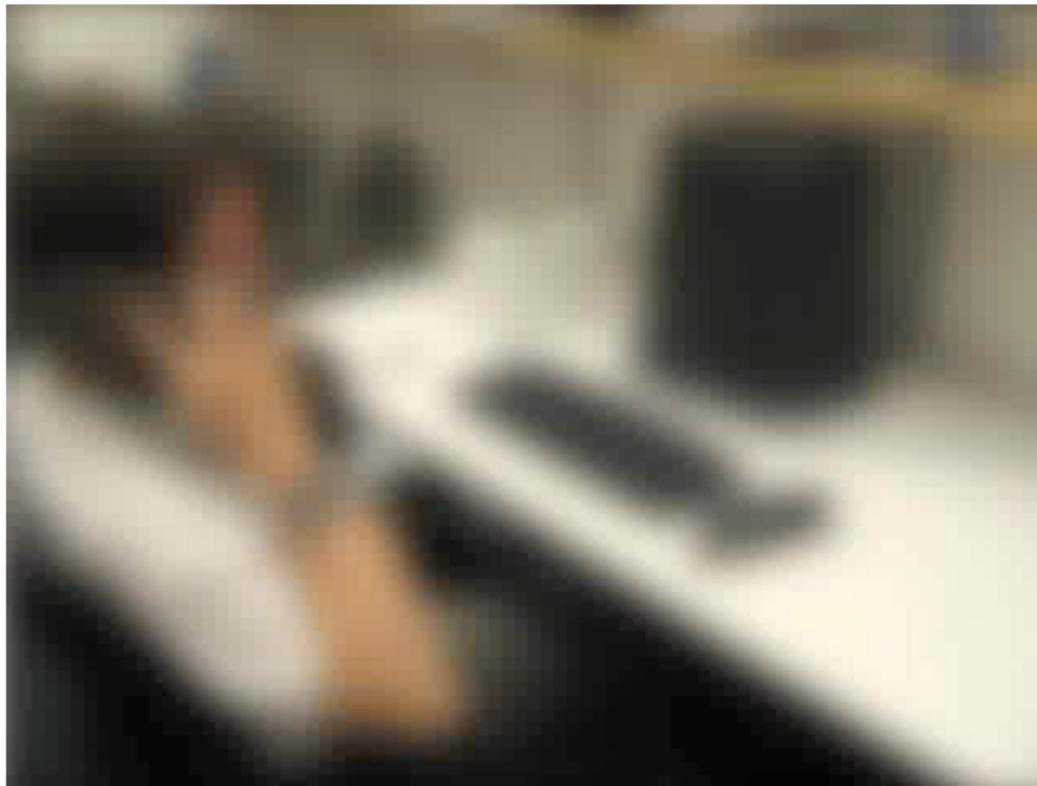
Challenges: Perception vs. Measurement



Challenges: Local Ambiguities



Challenges: Local Ambiguities



Challenges: Local Ambiguities



Challenges: Intra Class Variation



<http://www.homeworkshop.com/>

Challenges: Number of Object Categories



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

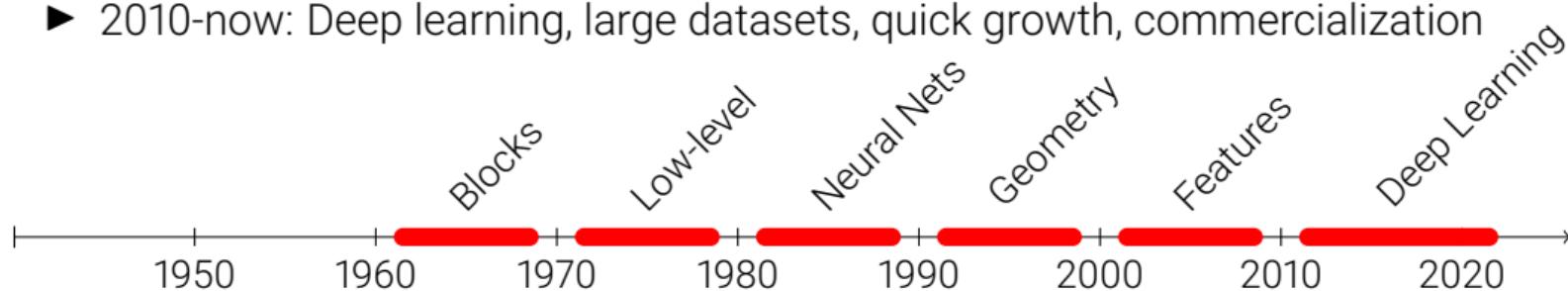
- ▶ Multi-decade project to measure the entire Indian subcontinent with scientific precision
- ▶ Under the leadership of George Everest, the project was made responsible of the Survey of India
- ▶ Manual bundle adjustment proves Mt. Everest highest mountain on earth mountain on earth



Overview

Waves of development:

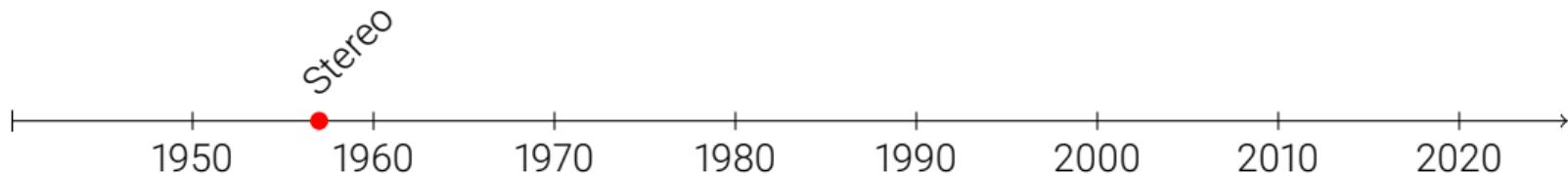
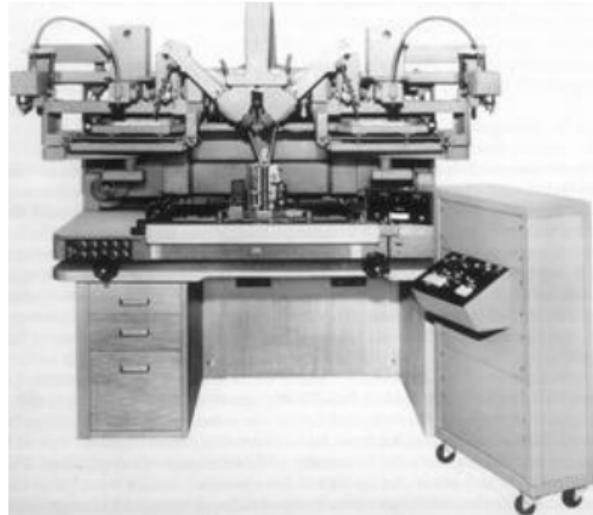
- ▶ 1960-1970: Blocks Worlds, Edges and Model Fitting
- ▶ 1970-1981: Low-level vision: stereo, flow, shape-from-shading
- ▶ 1985-1988: Neural networks, backpropagation, self-driving
- ▶ 1990-2000: Dense stereo and multi-view stereo, MRFs
- ▶ 2000-2010: Features, descriptors, large-scale structure-from-motion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



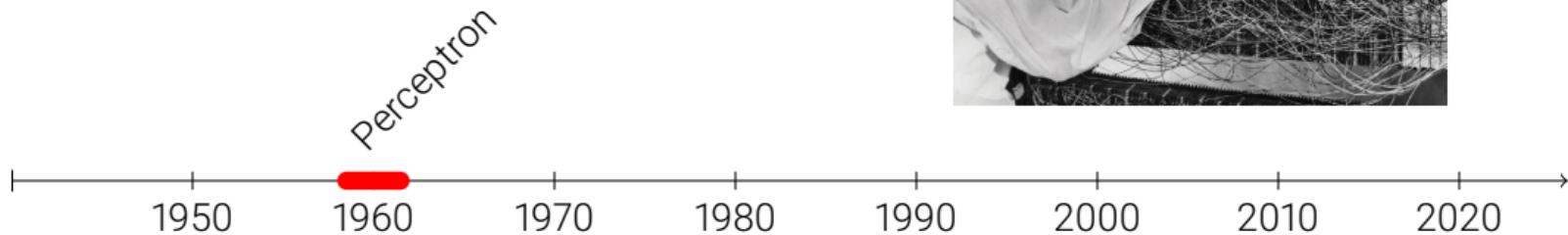
A Brief History of Computer Vision

1958-1962: Rosenblatt's Perceptron

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

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

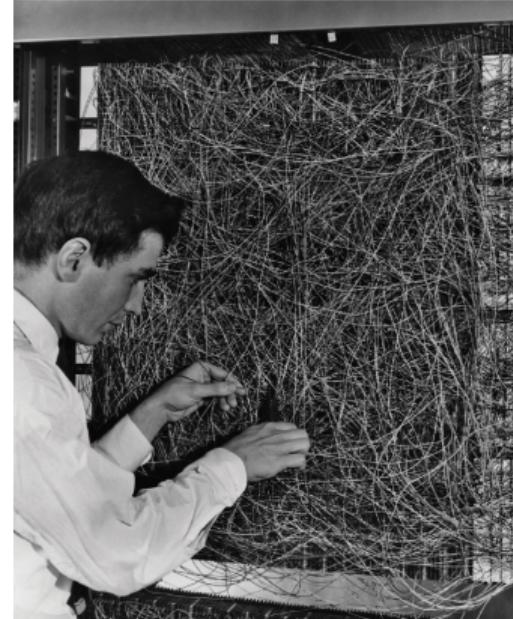
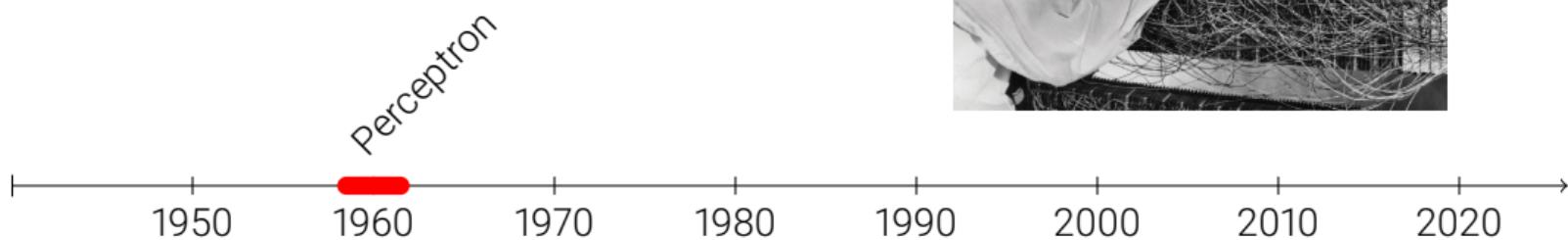
- ▶ Novikoff proved convergence



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

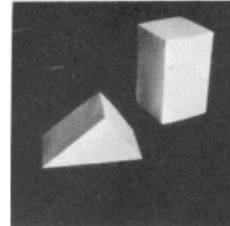
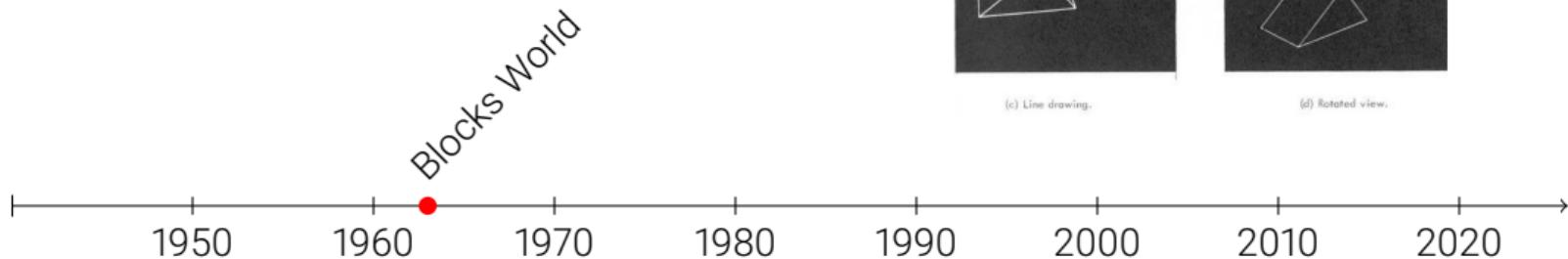


Rosenblatt: The perceptron - a probabilistic model for information storage and organization in the brain. Psychological Review, 1958.

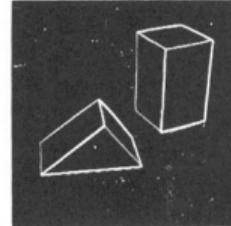
A Brief History of Computer Vision

1963: Larry Robert's Blocks World

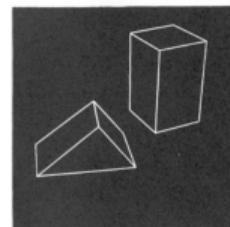
- ▶ Scene understanding for robotics
- ▶ Extracts edges as primitives
- ▶ Infers 3D structure of an object from topological structure of the 2D lines
- ▶ Interpret images as projections of 3D scenes, not 2D pattern recognition



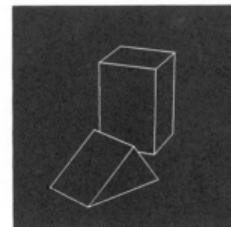
(a) Original picture.



(b) Differentiated picture.



(c) Line drawing.

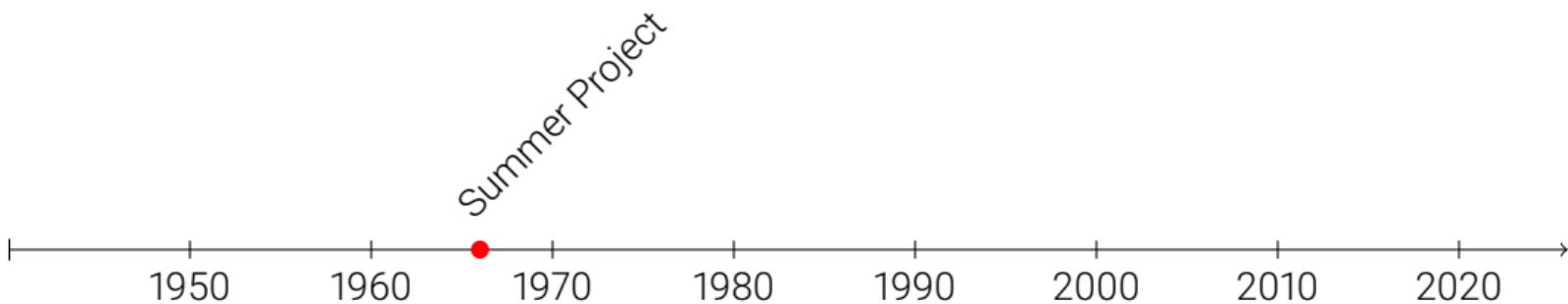


(d) Rotated view.

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
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT
Seymour Papert

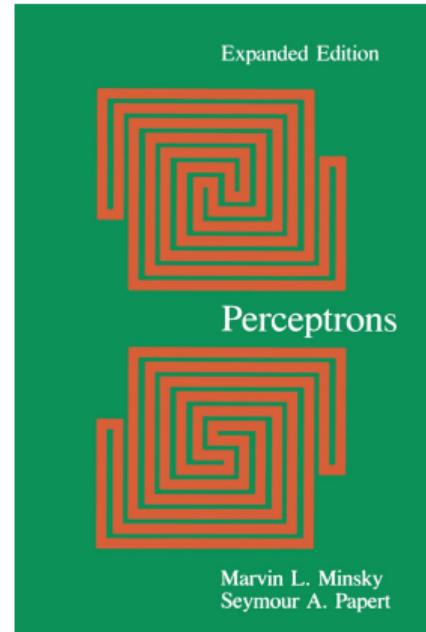
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

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

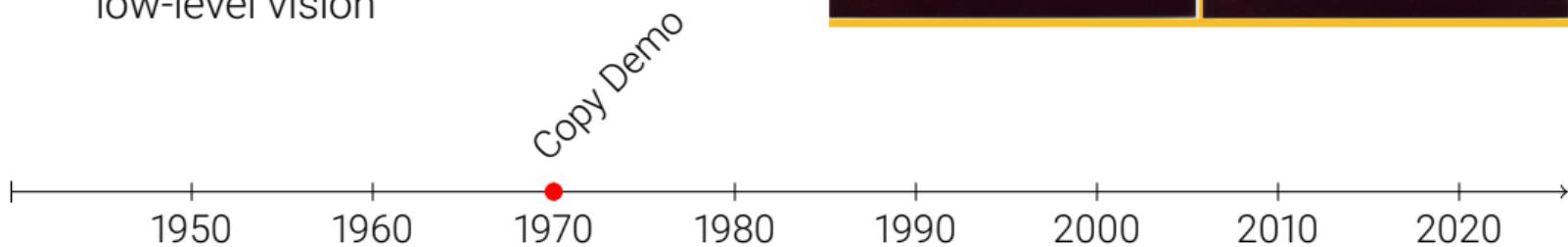
Minsky/Papert



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



A Brief History of Computer Vision

1970: Shape from Shading

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

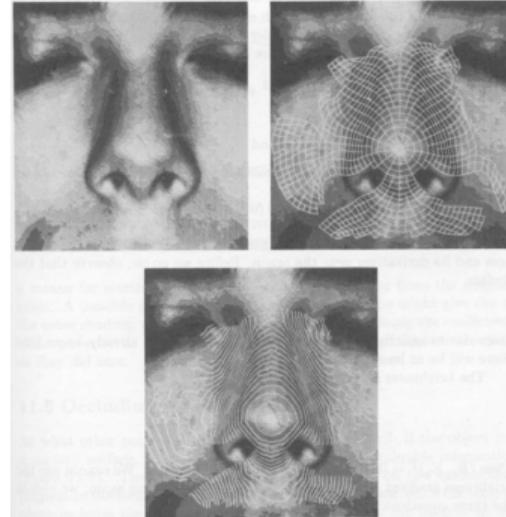
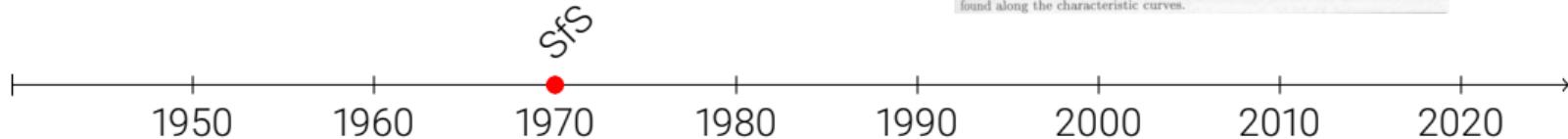


Figure 11-7. The shape-from-shading method is applied here to the recovery of the shape of a nose. The first picture shows the (crudely quantized) gray-level image available to the program. The second picture shows the base characteristics superimposed, while the third shows a contour map computed from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

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

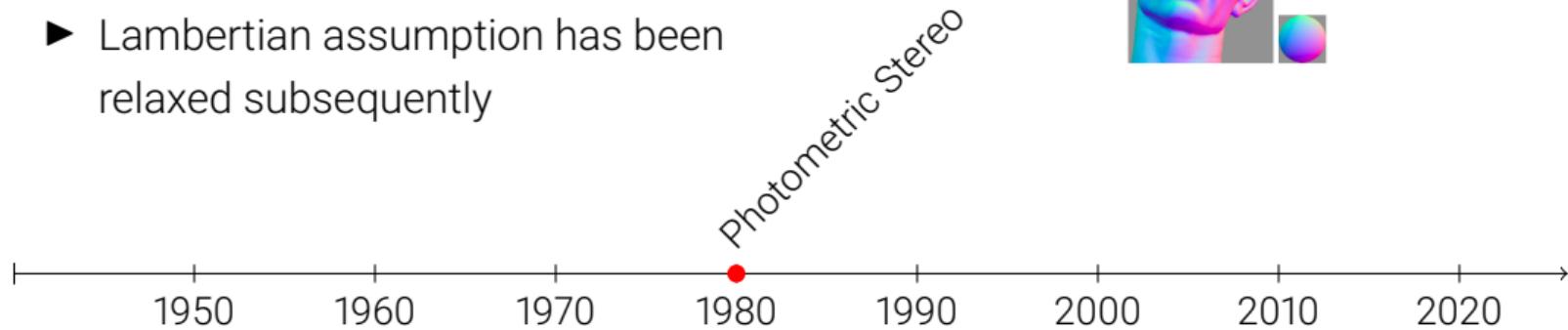
Intrinsic Images



A Brief History of Computer Vision

1980: Photometric Stereo

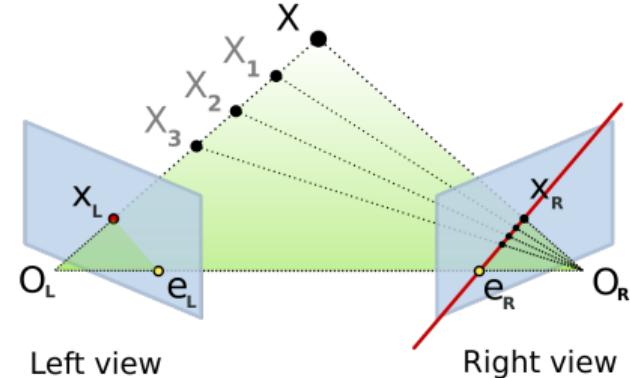
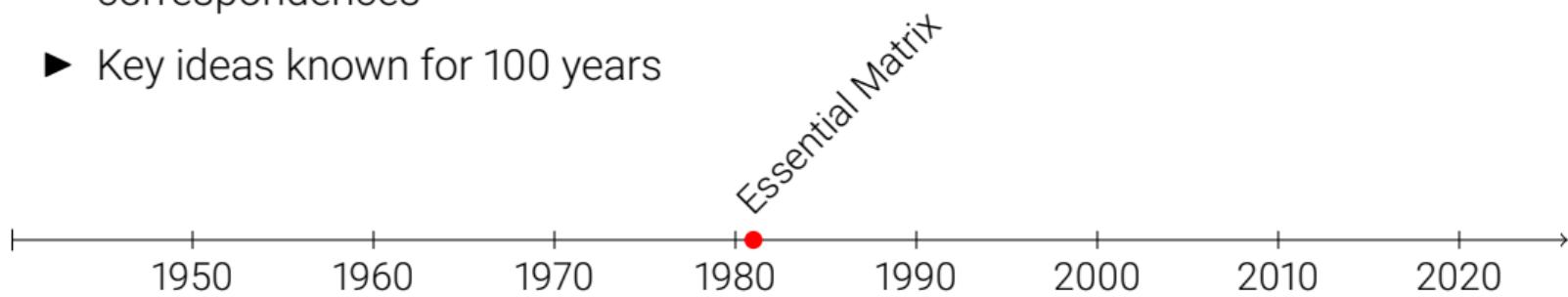
- ▶ Recover 3D from multiple 2D images, taken from the same viewpoint with different lighting conditions
- ▶ Requires at least 3 images
- ▶ Unprecedented detail and accuracy
- ▶ Lambertian assumption has been relaxed subsequently



A Brief History of Computer Vision

1981: Essential Matrix

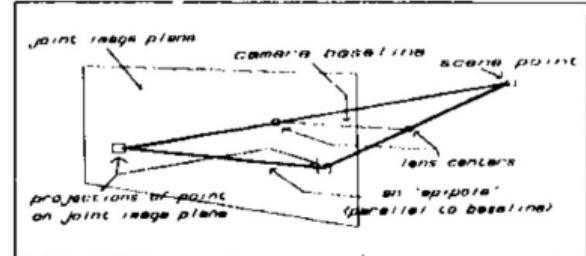
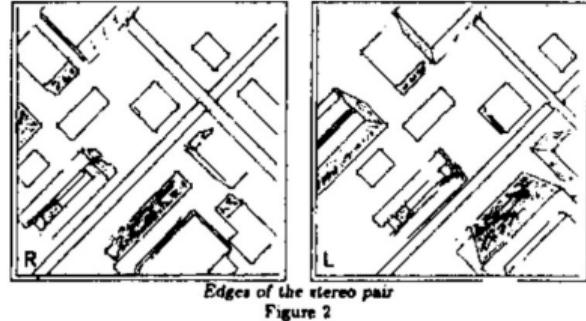
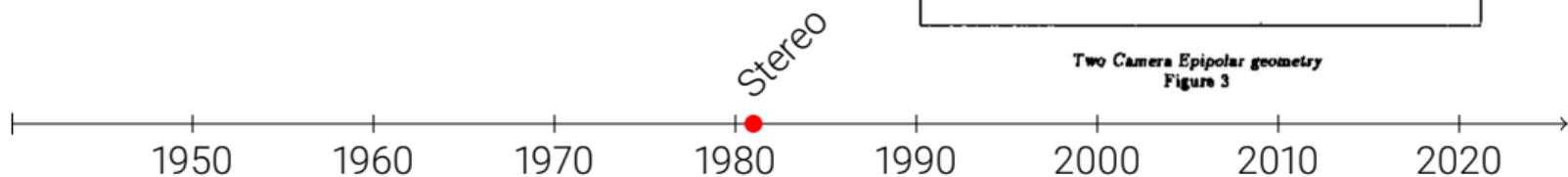
- ▶ Defines two-view geometry as matrix mapping points to epipolar lines
- ▶ Reduces correspondence search to a 1D problem
- ▶ Can be estimated from a set of 2D correspondences
- ▶ Key ideas known for 100 years



A Brief History of Computer Vision

1981: Binocular Scanline Stereo

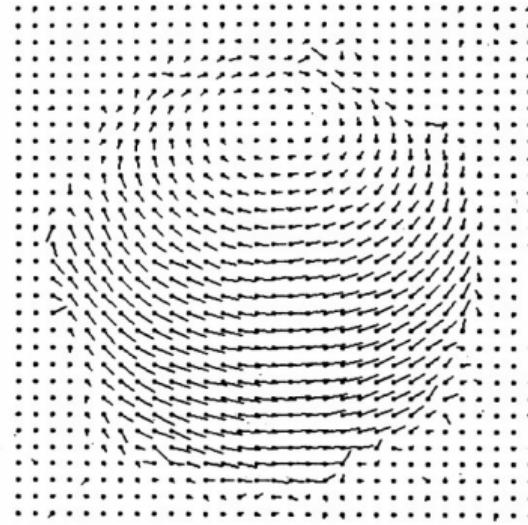
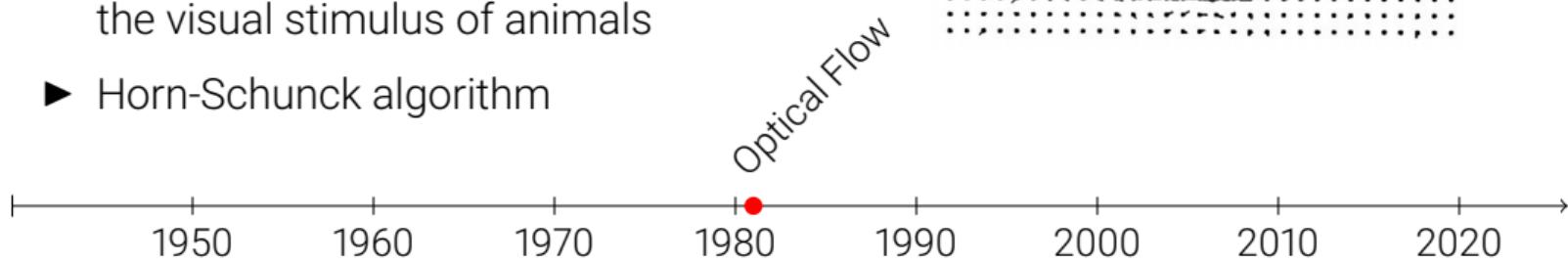
- ▶ Correlate points along epipolar lines
- ▶ Use dynamic programming to introduce constraints along scanlines (image rows)
- ▶ Allows for overcoming ambiguities, but streaking artifacts between rows



A Brief History of Computer Vision

1981: Dense Optical Flow

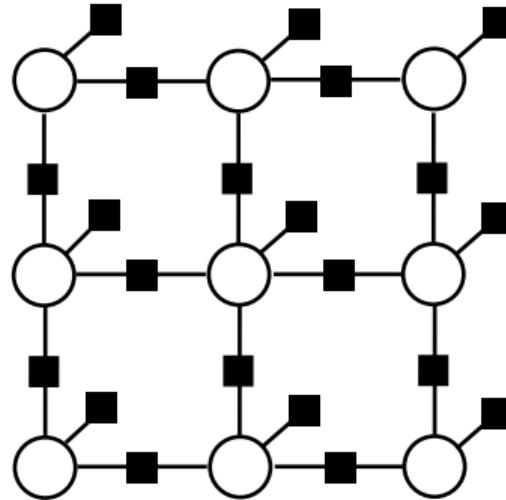
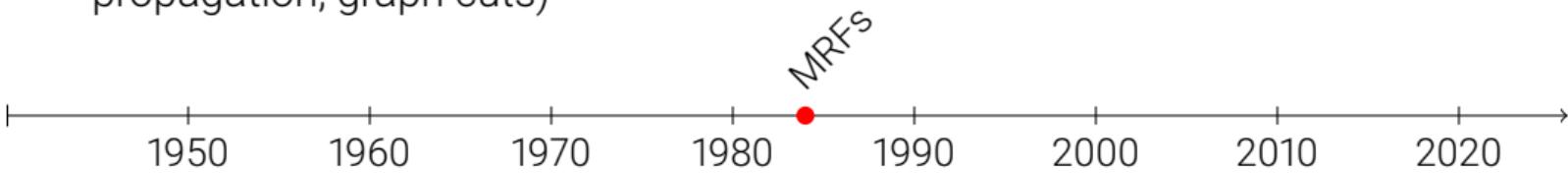
- ▶ Pattern of apparent motion of objects, surfaces, and edges in a visual scene
- ▶ Measured by (densely) tracking pixels between two frames
- ▶ Investigated by Gibson to describe the visual stimulus of animals
- ▶ Horn-Schunck algorithm



A Brief History of Computer Vision

1984: Markov Random Fields

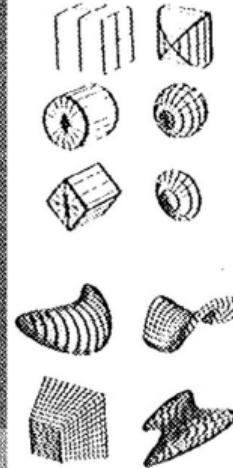
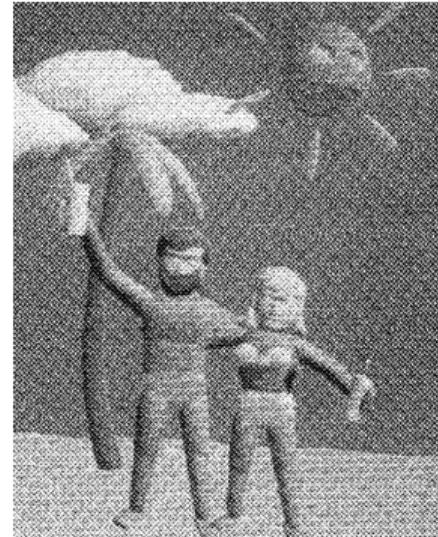
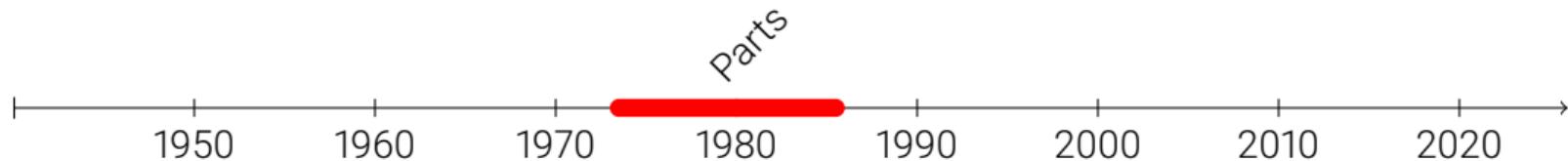
- ▶ MRFs for encoding prior knowledge (e.g., about smoothness)
- ▶ Resolves ambiguities in many ill-posed vision problems (e.g., stereo, flow, denoising)
- ▶ Global optimization (e.g., variational inference, sampling, belief propagation, graph cuts)



A Brief History of Computer Vision

1980s: Part-based Models

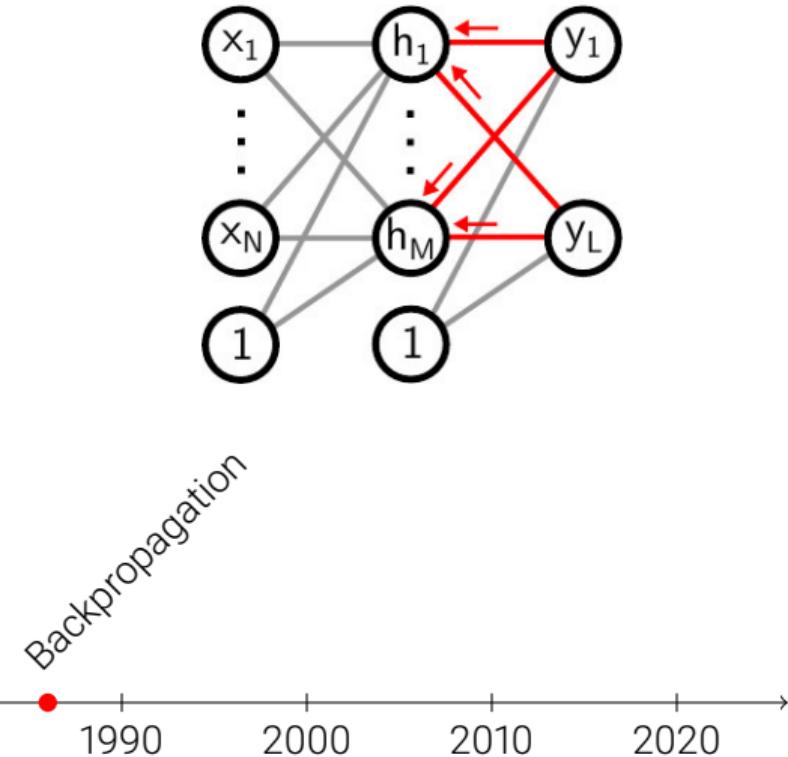
- ▶ 1973: Pictorial Structures
- ▶ 1976: Generalized Cylinders
(solids of revolution, swept curves)
- ▶ 1986: Superquadrics
(generalization of quadric surfaces)
- ▶ Express complex relationships
- ▶ Compact representation



A Brief History of Computer Vision

1986: Backpropagation Algorithm

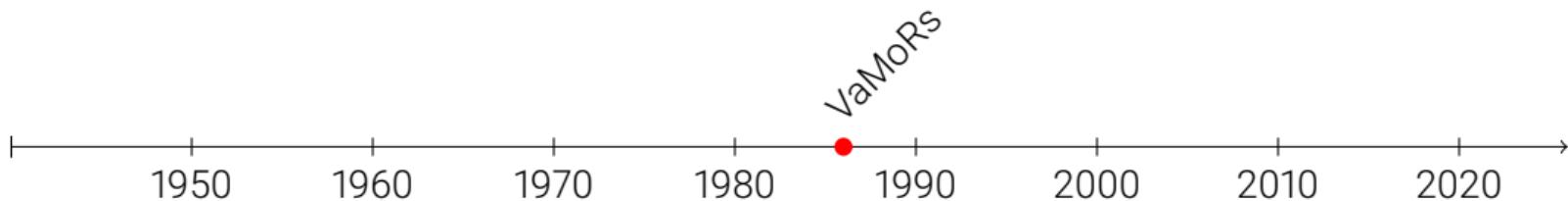
- ▶ Efficient calculation of gradients in a deep network wrt. network weights
- ▶ Enables application of gradient based learning to deep networks
- ▶ Known since 1961, but first empirical success in 1986
- ▶ Remains main workhorse today



A Brief History of Computer Vision

1986: Self-Driving Car VaMoRs

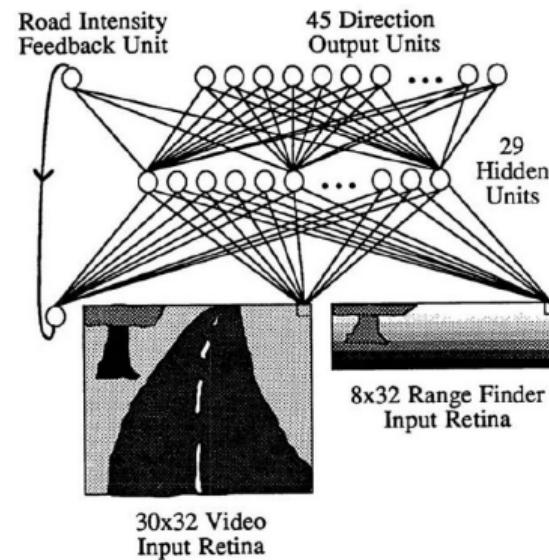
- ▶ Developed by Ernst Dickmanns in context of EUREKA-Prometheus
- ▶ Demonstration to Daimler-Benz Research 1986 in Stuttgart
- ▶ Longitudinal & lateral guidance with lateral acceleration feedback
- ▶ Speed: 0 to 36 km/h



A Brief History of Computer Vision

1988: Self-Driving Car ALVINN

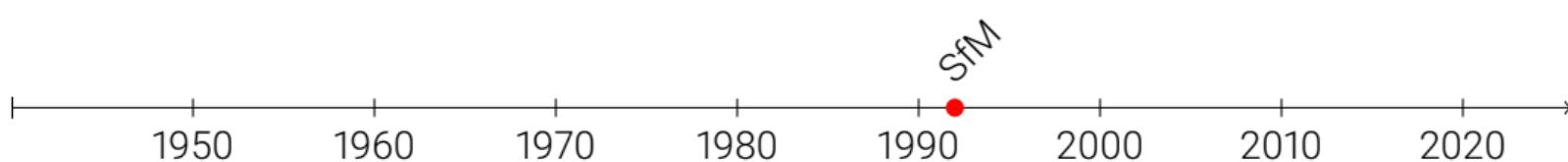
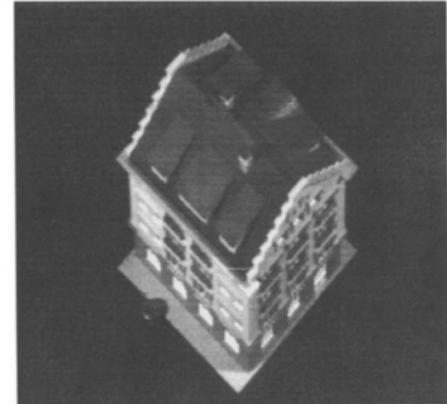
- ▶ Forward-looking, vision based driving
- ▶ Fully connected neural network maps road images to vehicle turn radius
- ▶ Trained on simulated road images
- ▶ Tested on unlined paths, lined city streets and interstate highways
- ▶ 90 consecutive miles at up to 70 mph



A Brief History of Computer Vision

1992: Structure-from-Motion

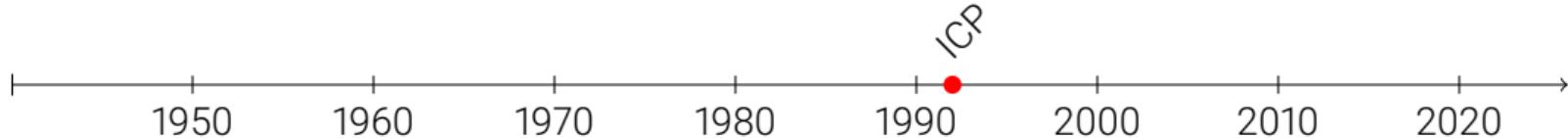
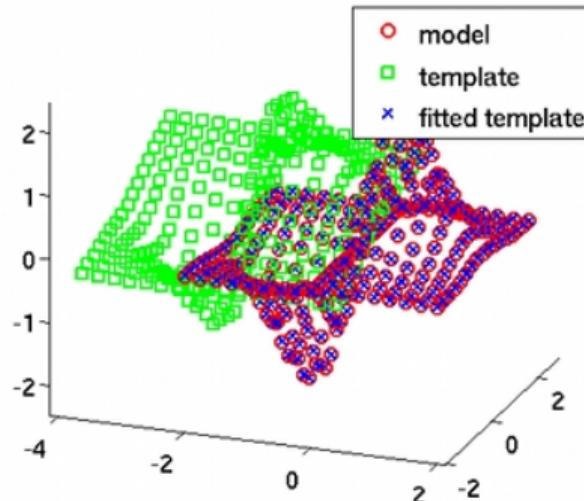
- ▶ Estimating 3D structures from 2D image sequences of static scenes
- ▶ Requires only a single camera
- ▶ Tomasi-Kanade factorization provides closed-form (SVD-based) solution for orthographic case
- ▶ Today: non-linear least squares



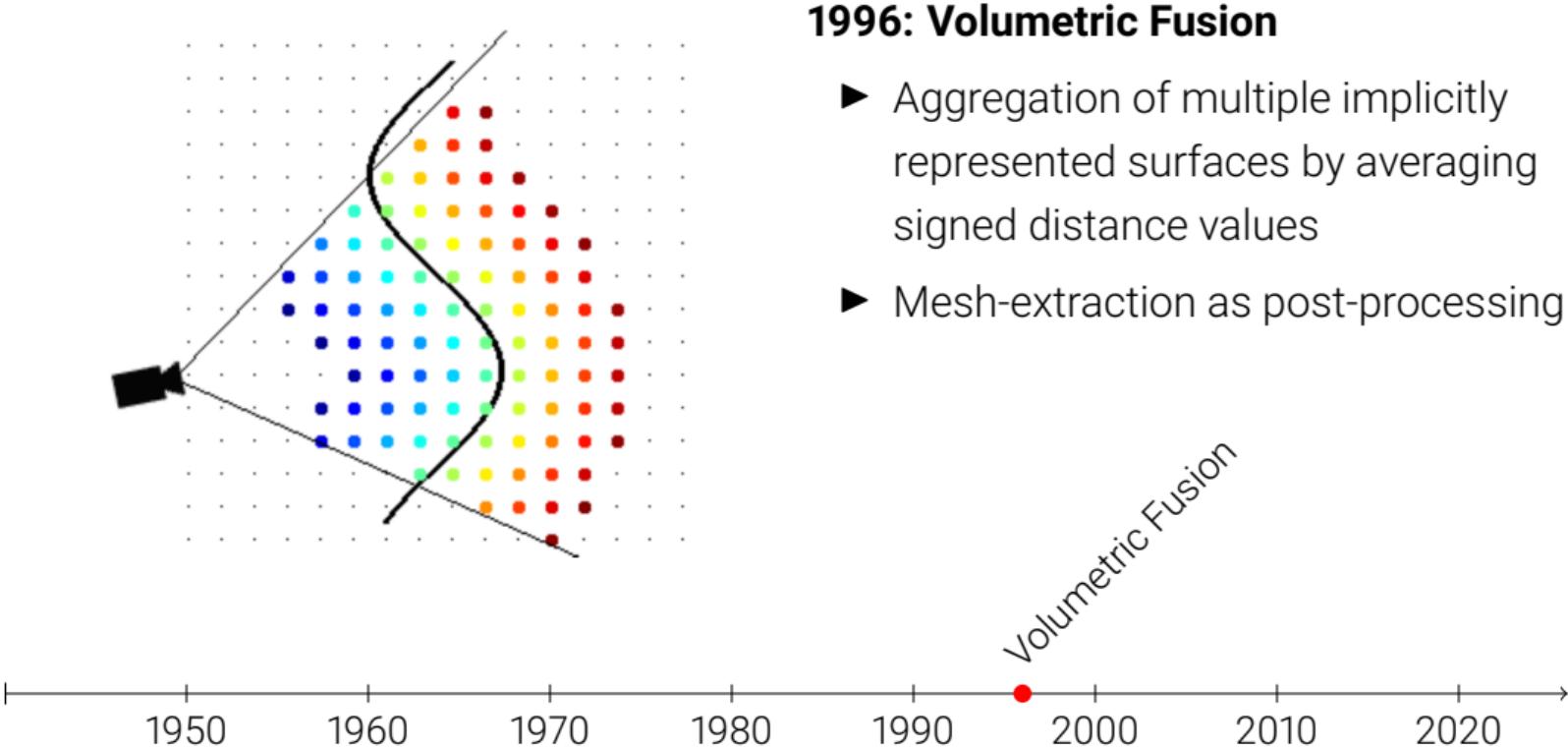
A Brief History of Computer Vision

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



A Brief History of Computer Vision



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

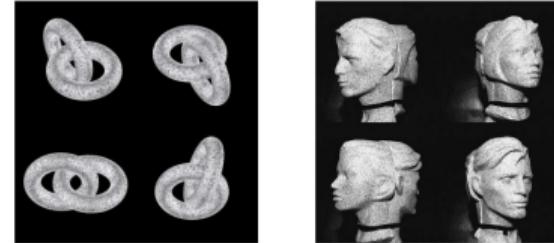


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

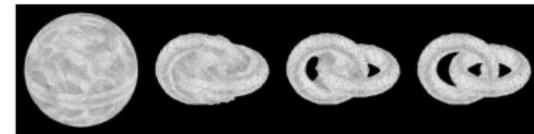
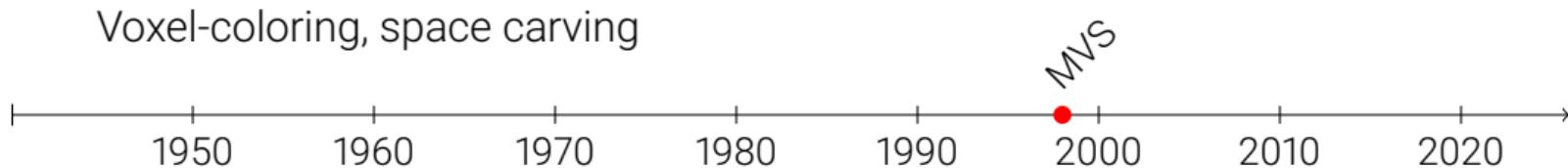


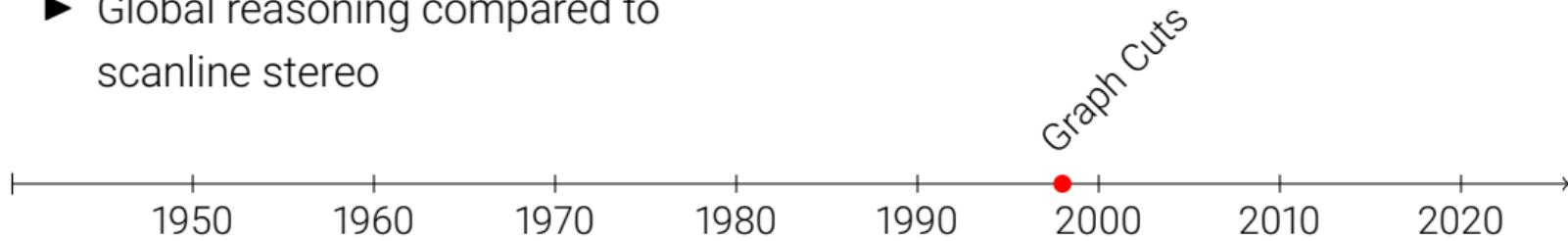
Fig. 4. Evolution of the surface for the two tori.



A Brief History of Computer Vision

1998: Stereo with Graph Cuts

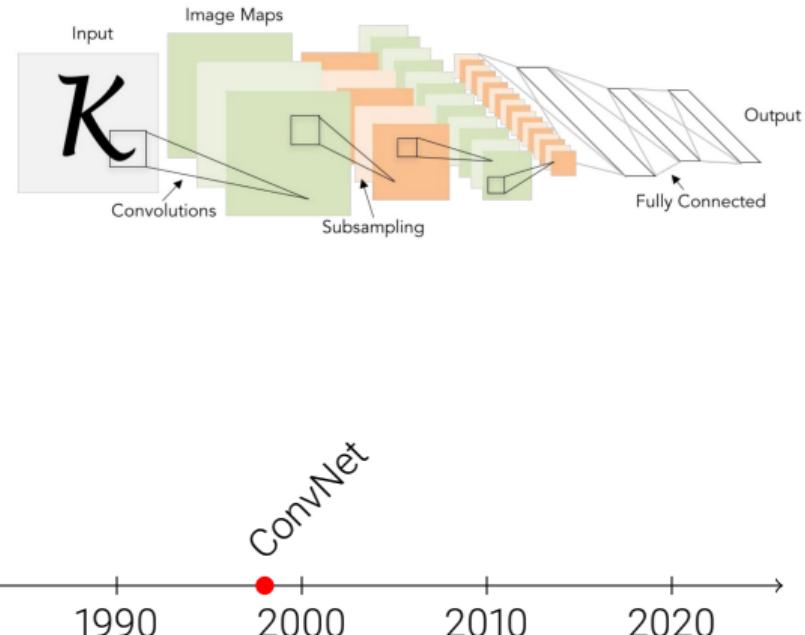
- ▶ Popular discrete MAP inference algorithm for Markov Random Fields
- ▶ First versions included unary and pairwise terms
- ▶ Later versions also included specific forms of higher-order potentials
- ▶ Global reasoning compared to scanline stereo



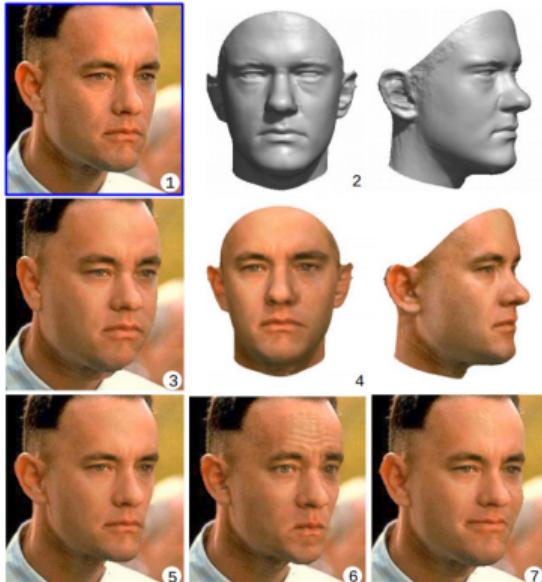
A Brief History of Computer Vision

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

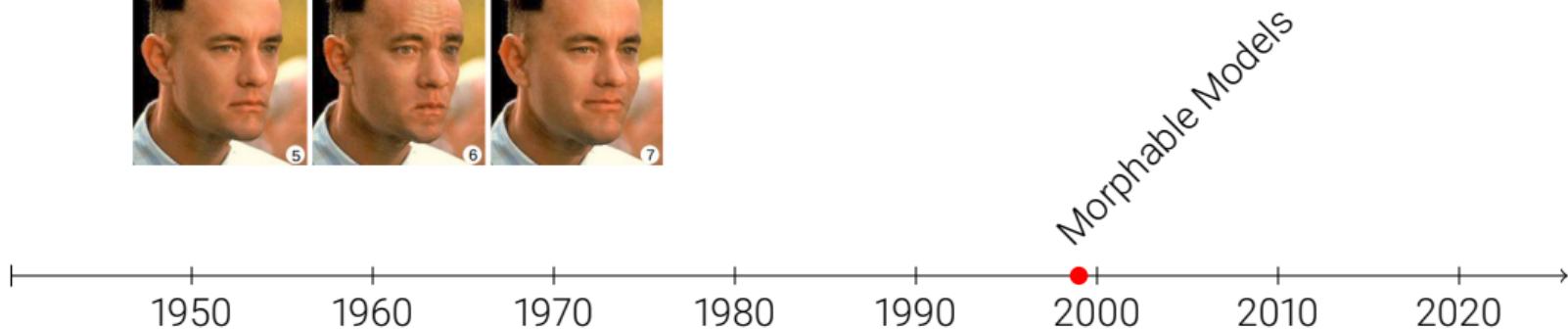


A Brief History of Computer Vision

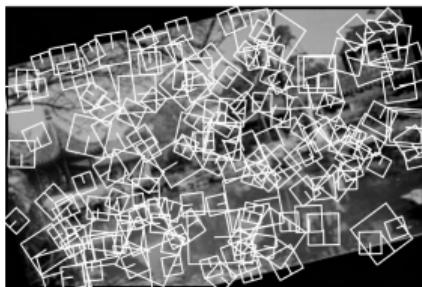
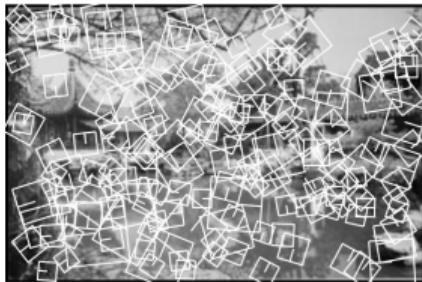


1999: Morphable Models

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

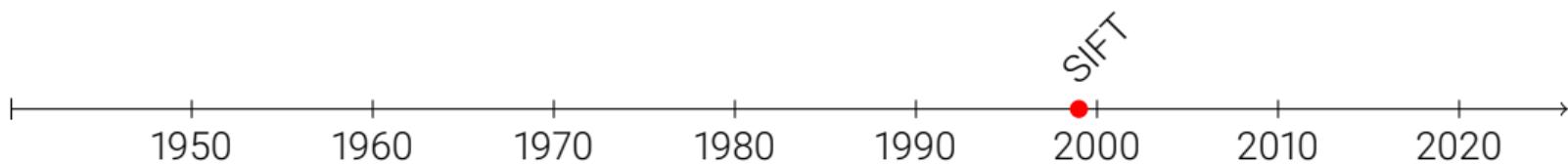


A Brief History of Computer Vision



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, ...)

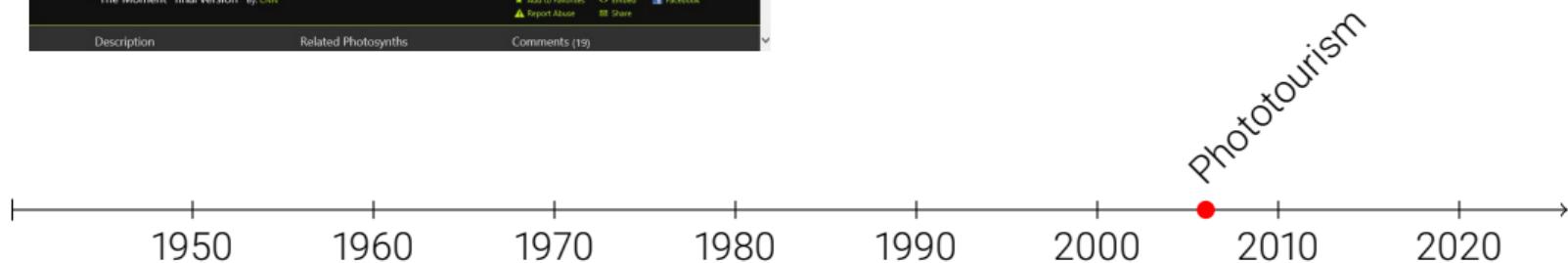


A Brief History of Computer Vision



2006: Photo Tourism

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

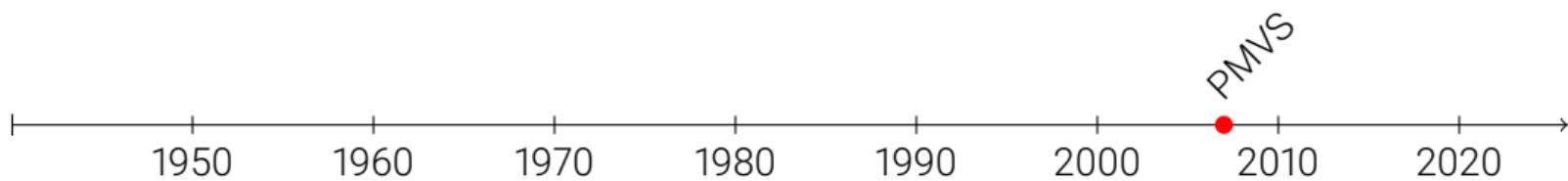


A Brief History of Computer Vision



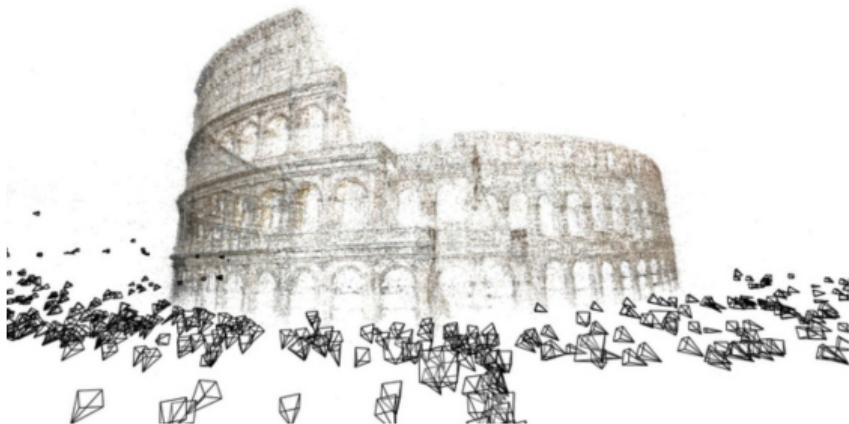
2007: PMVS

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

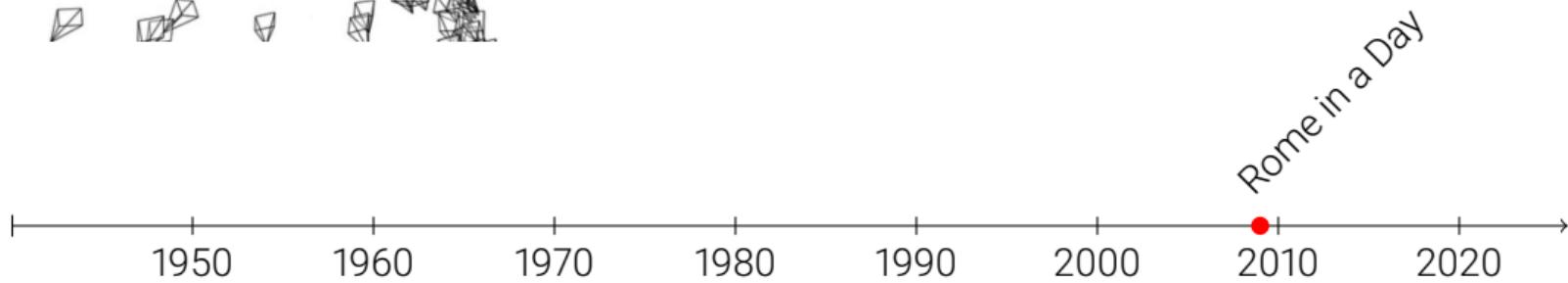


A Brief History of Computer Vision

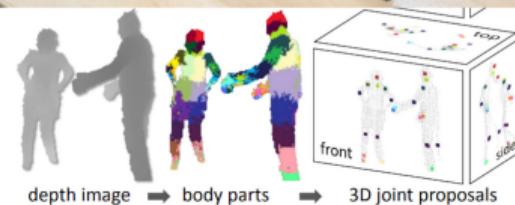
2009: Building Rome in a Day



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

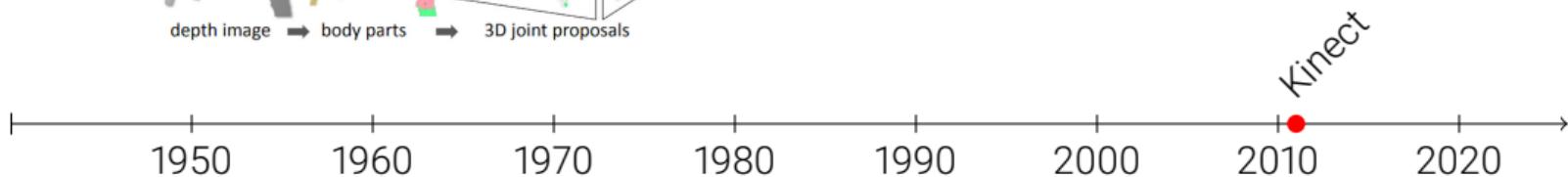


A Brief History of Computer Vision



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



A Brief History of Computer Vision

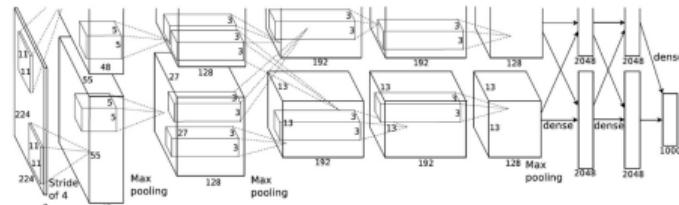
2009-2012: ImageNet and AlexNet

ImageNet

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

AlexNet

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



Image/AlexNet



A Brief History of Computer Vision

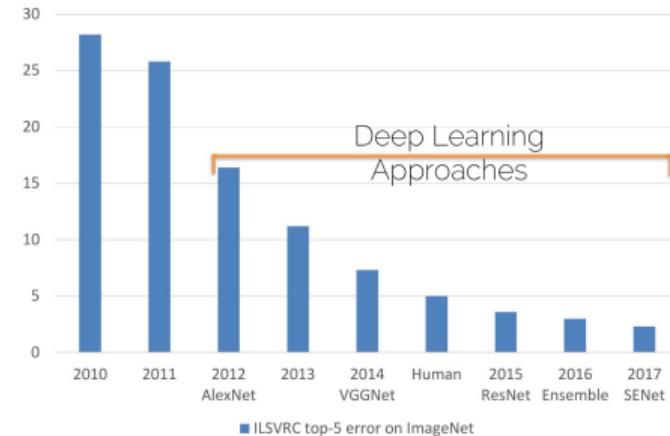
2009-2012: ImageNet and AlexNet

ImageNet

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

AlexNet

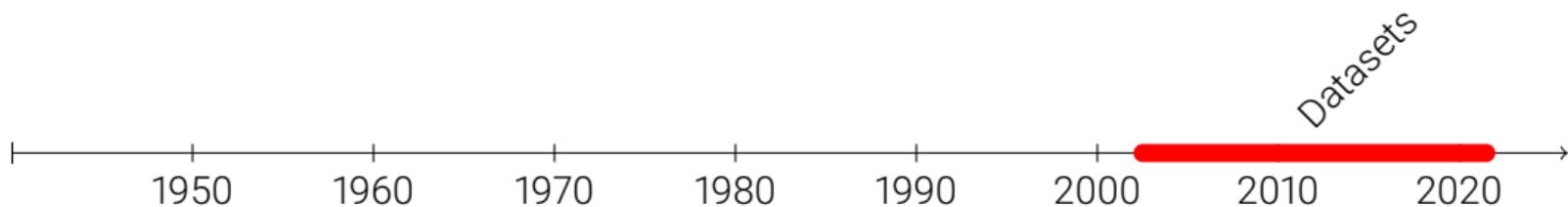
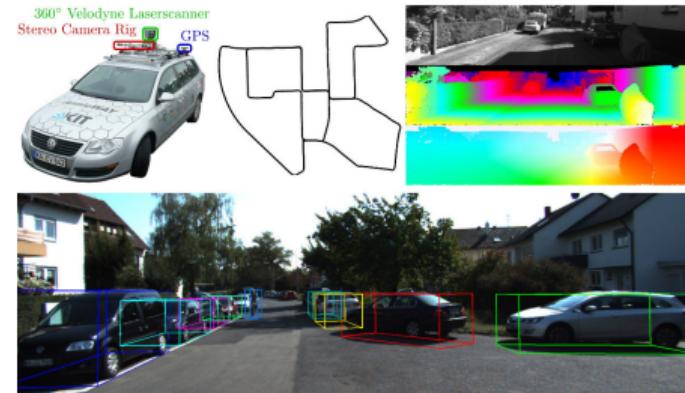
- ▶ First neural network to win ILSVRC via **GPU training, deep models, data**
- ▶ Sparked deep learning revolution



A Brief History of Computer Vision

2002-now: Golden Age of Datasets

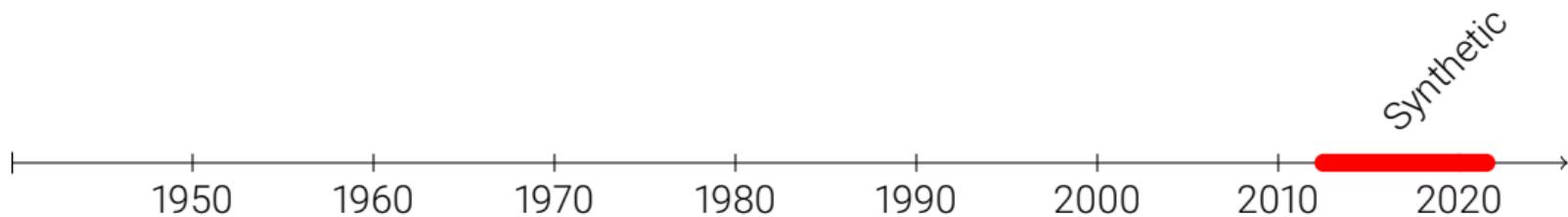
- ▶ Middlebury Stereo and Flow
- ▶ KITTI, Cityscapes: Self-driving
- ▶ PASCAL, MS COCO: Recognition
- ▶ ShapeNet, ScanNet: 3D DL
- ▶ Visual Genome: Vision/Language
- ▶ MITOS: Breast cancer



A Brief History of Computer Vision

2012-now: Synthetic Data

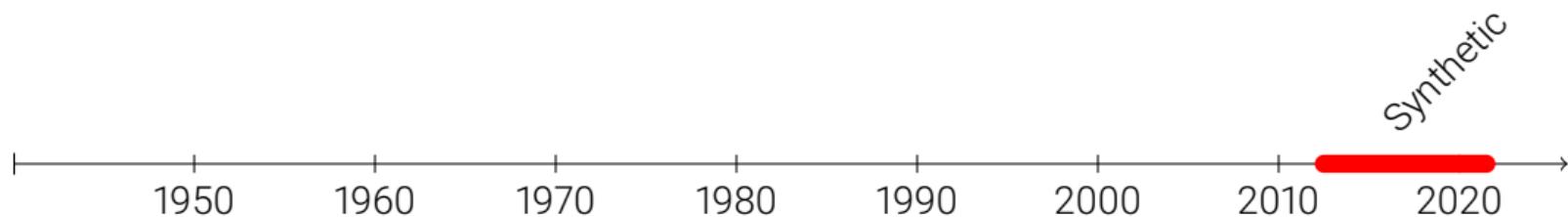
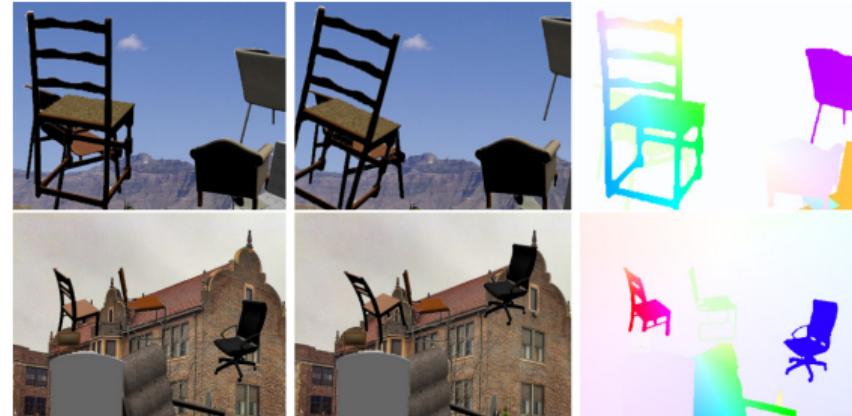
- ▶ Annotating real data is expensive
- ▶ Led to surge of synthetic datasets
- ▶ Creating 3D assets is also costly



A Brief History of Computer Vision

2012-now: Synthetic Data

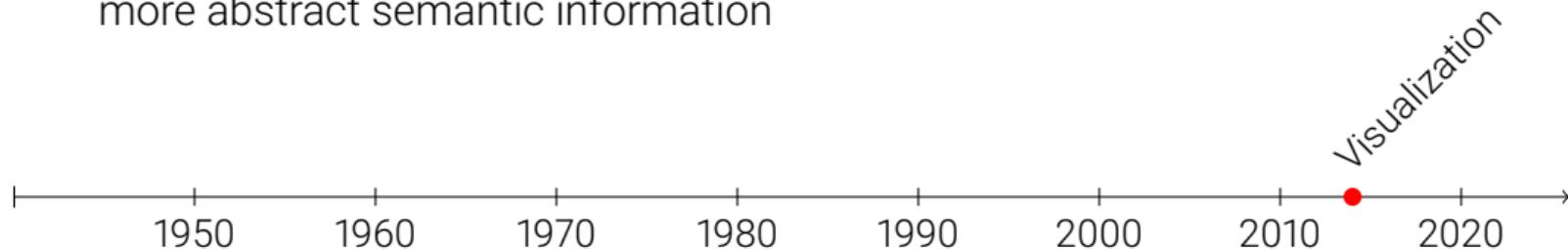
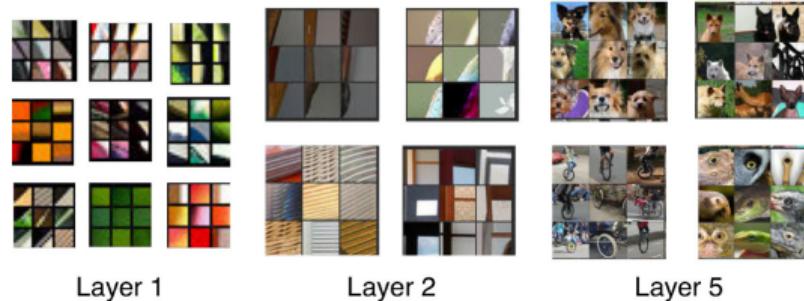
- ▶ Annotating real data is expensive
- ▶ Led to surge of synthetic datasets
- ▶ Creating 3D assets is also costly
- ▶ But even very simple 3D datasets proved tremendously useful for pre-training (e.g., in optical flow)



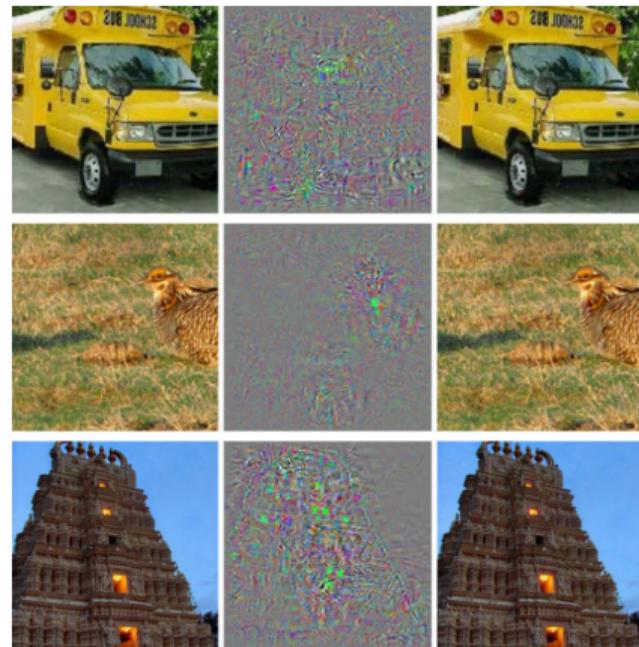
A Brief History of Computer Vision

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

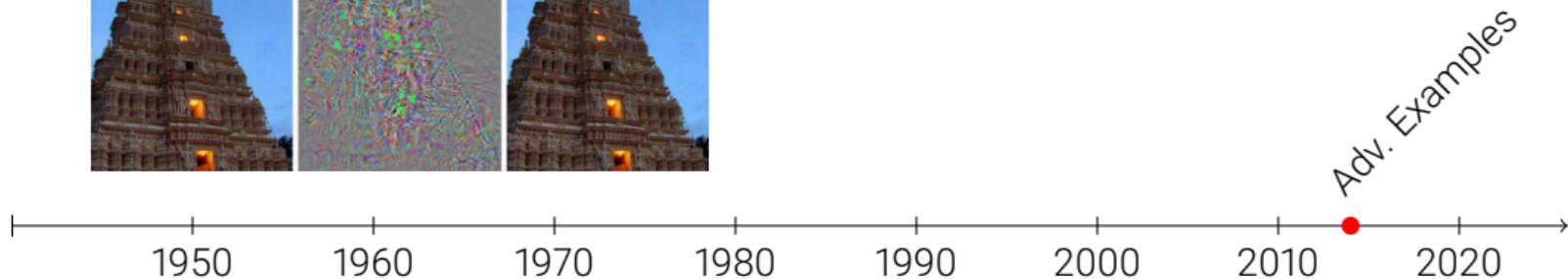


A Brief History of Computer Vision



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”

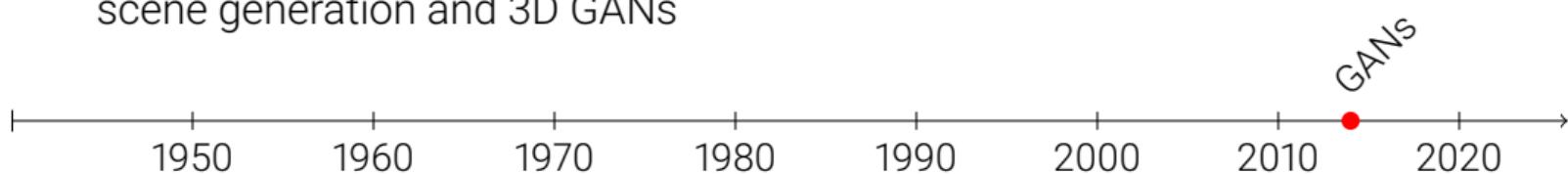


A Brief History of Computer Vision

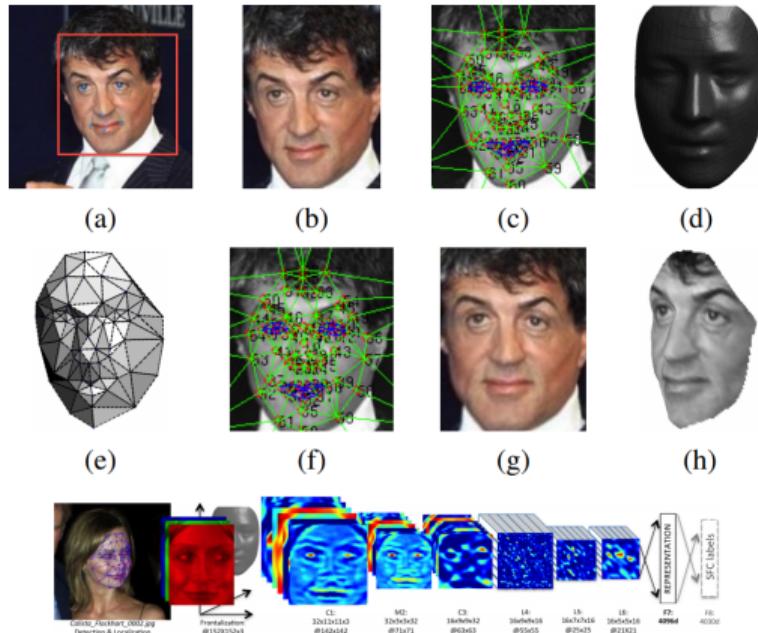
2014: Generative Adversarial Networks

- 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

Moore's Law of AI
4.5 years of progress on faces

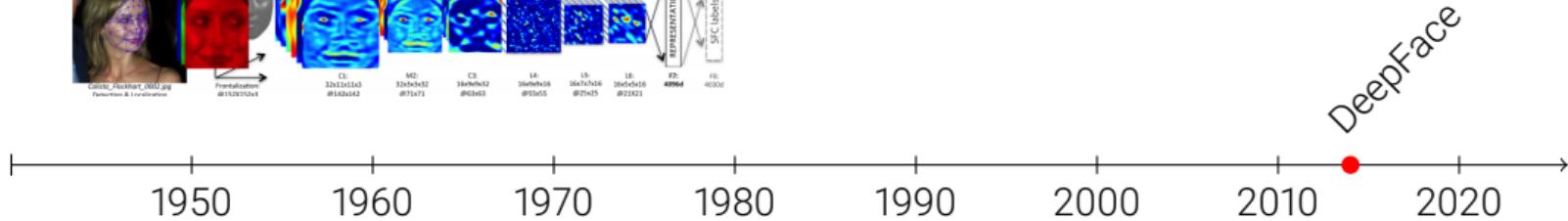


A Brief History of Computer Vision

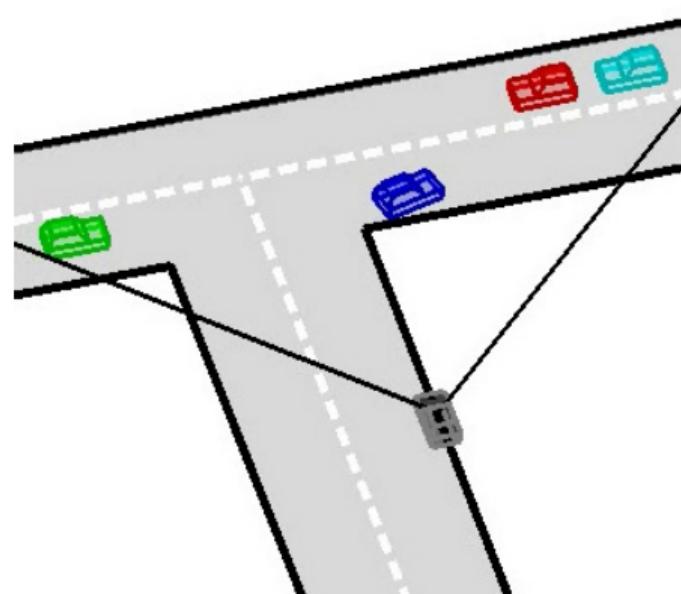


2014: DeepFace

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



A Brief History of Computer Vision



2014: 3D Scene Understanding

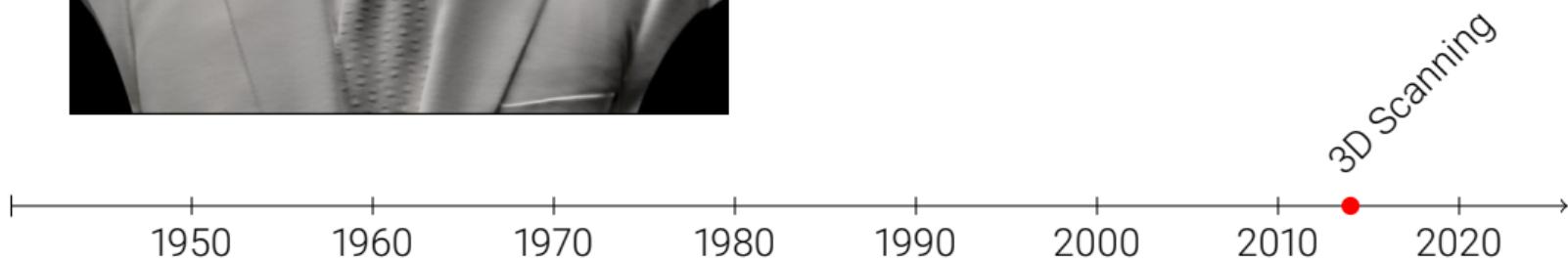
- ▶ Parsing RGB and RGB-D images into holistic 3D scene representations
- ▶ Methods for indoors and outdoors

A Brief History of Computer Vision



2014: 3D Scanning

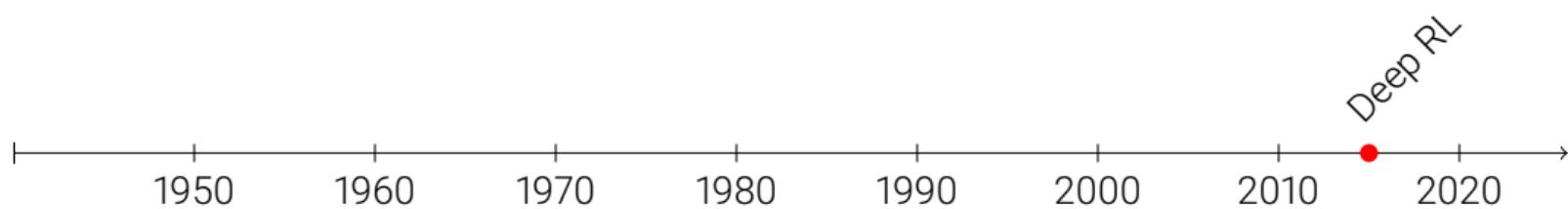
- ▶ 3D scanning techniques allow for creating accurate replicas
- ▶ Debevec's team scans Obama
- ▶ Exhibition in Smithsonian
- ▶ <https://dpo.si.edu/blog/>



A Brief History of Computer Vision

2015: Deep Reinforcement Learning

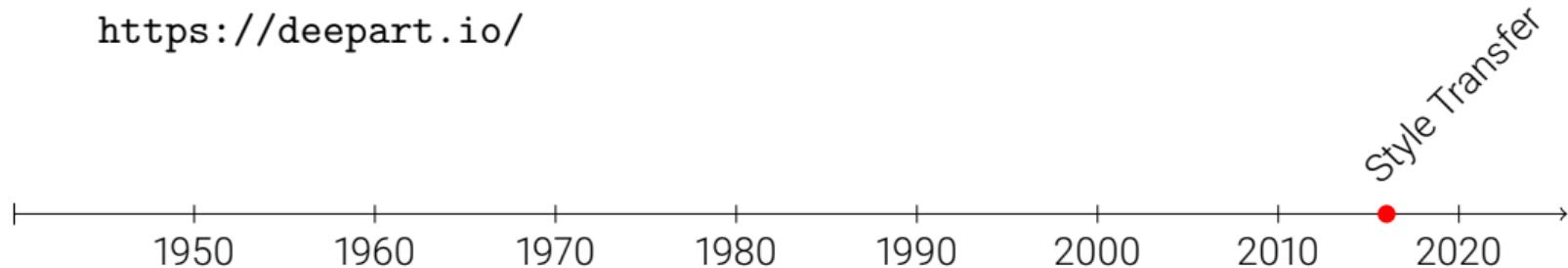
- ▶ Learning a policy (state→action) through random exploration and reward signals (e.g., game score)
- ▶ No other supervision
- ▶ Success on many Atari games
- ▶ But some games remain hard



A Brief History of Computer Vision

2016: Style Transfer

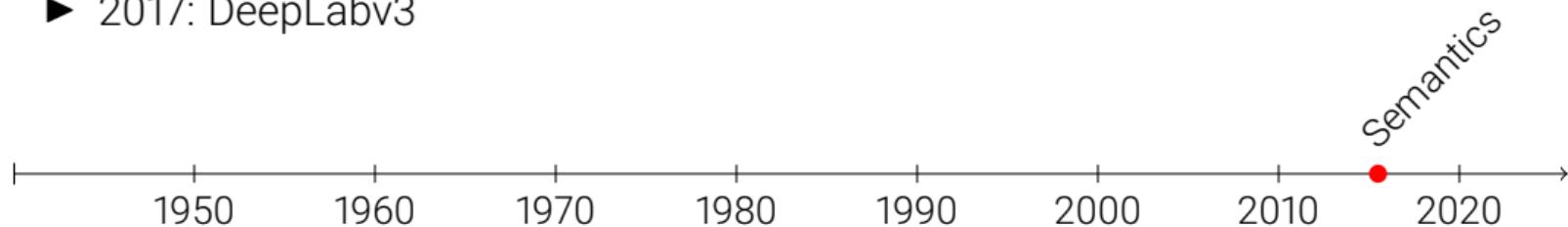
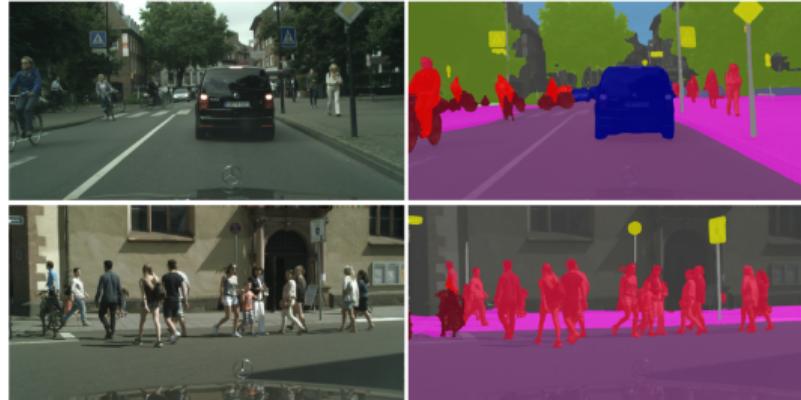
- ▶ Manipulate photograph to adopt style of another image (painting)
- ▶ Uses deep network pre-trained on ImageNet for disentangling content from style
- ▶ It is fun! Try yourself:
<https://deepart.io/>



A Brief History of Computer Vision

2015-2017: Semantic Segmentation

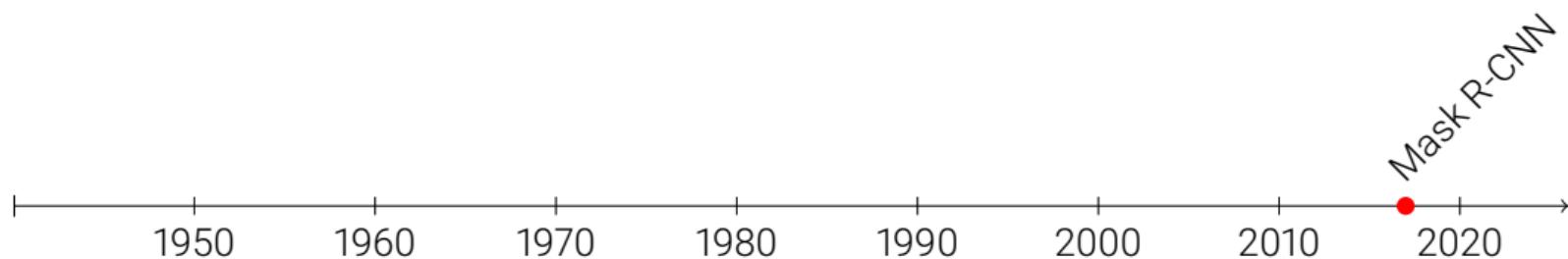
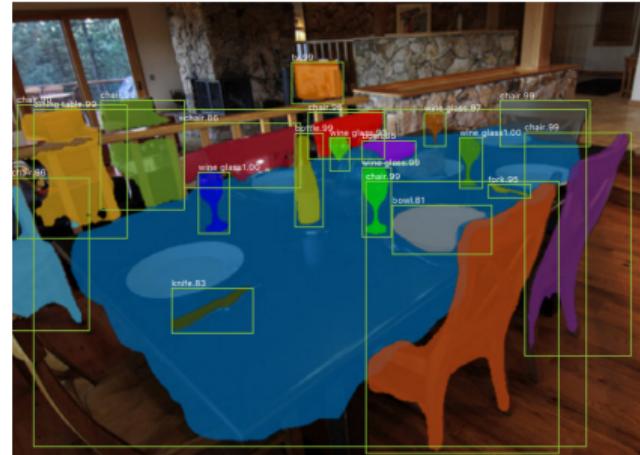
- ▶ Assign semantic class to every pixel
- ▶ Semantic segmentation starts to work on challenging real-world datasets (e.g., CityScapes)
- ▶ 2015: FCN, SegNet
- ▶ 2016: DeepLab, FSO
- ▶ 2017: DeepLabv3



A Brief History of Computer Vision

2017: Mask R-CNN

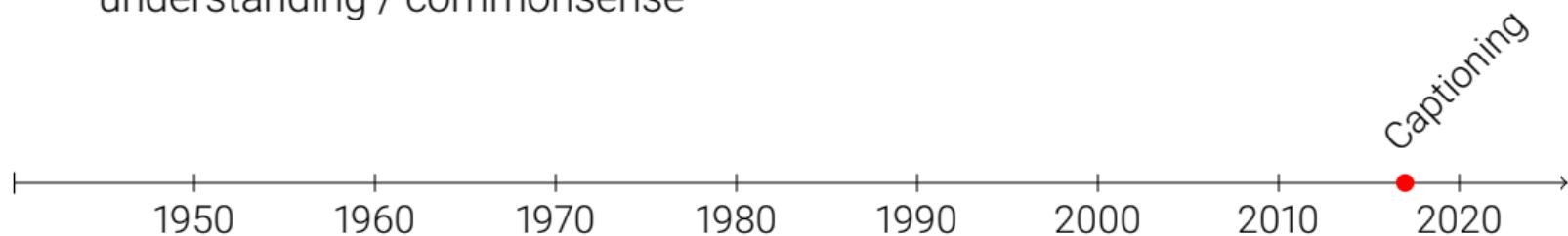
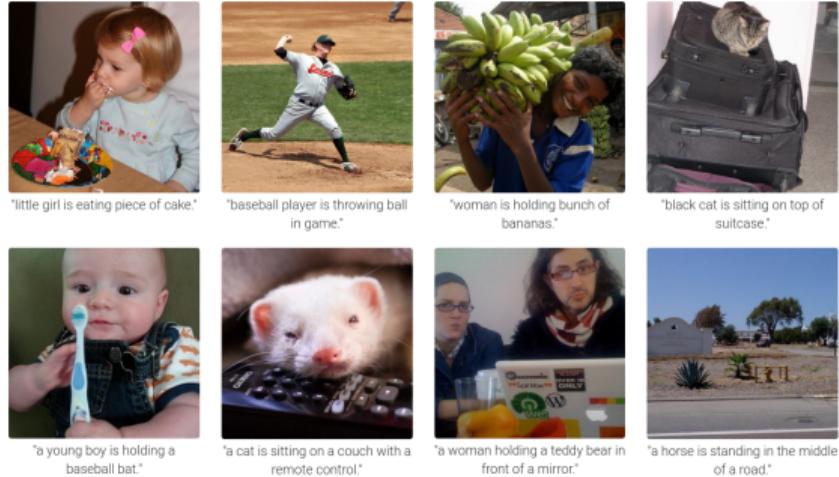
- Deep neural network for joint object detection and instance segmentation
- Outputs “structured object”, not only a single number (class label)
- State-of-the-art on MS-COCO



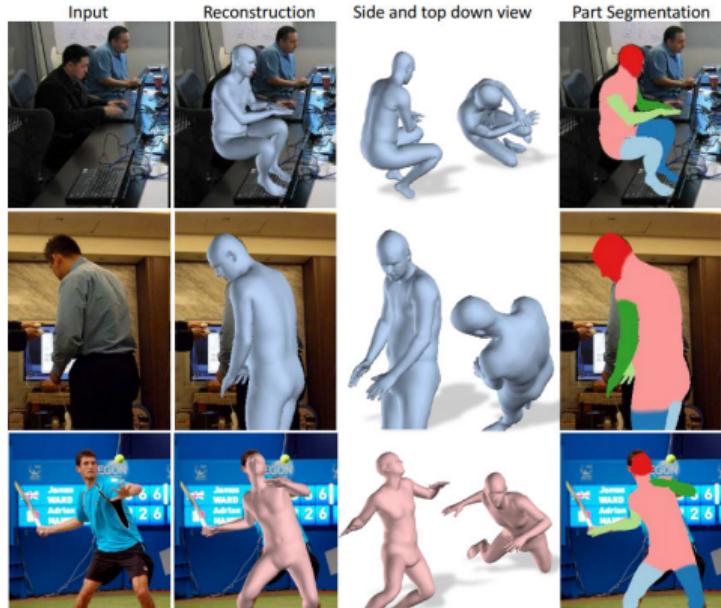
A Brief History of Computer Vision

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

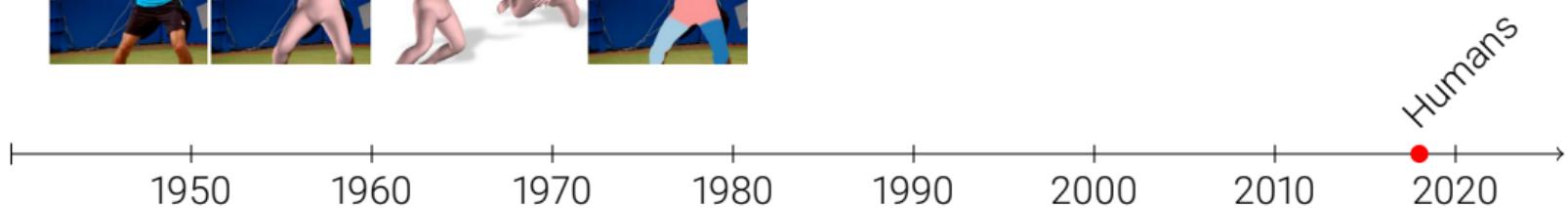


A Brief History of Computer Vision



2018: Human Shape and Pose

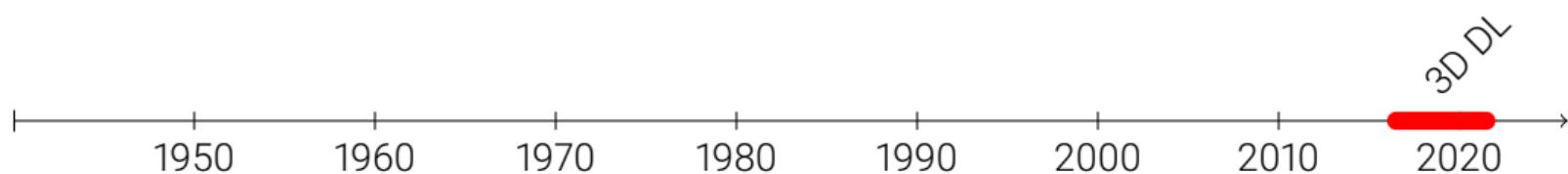
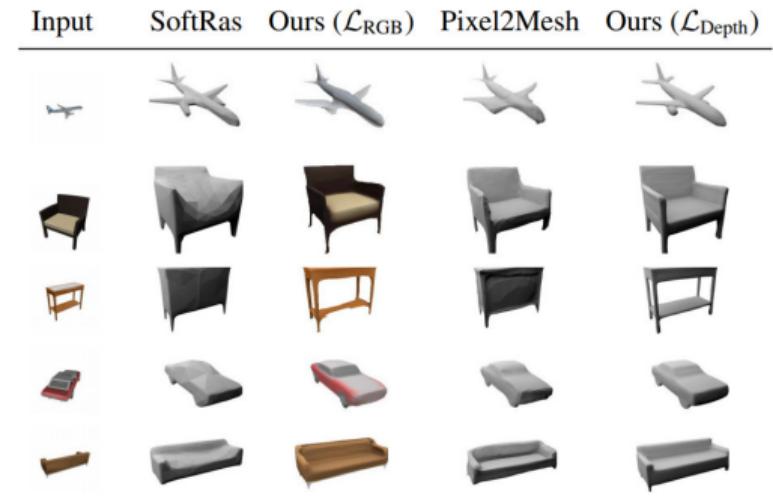
- ▶ Human pose/shape models mature
- ▶ Rich parametric models (e.g., SMPL, STAR)
- ▶ Regression from RGB images only
- ▶ Models of pose-dependent deformation and clothing



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



A Brief History of Computer Vision

Applications and Commercial Products



Google Portrait Mode



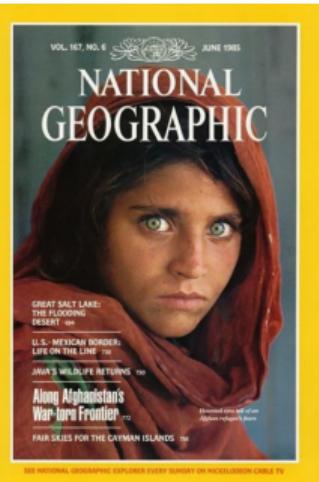
Skydio 2 Drone



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

