Computer Vision with Deep Learning

Introduction

Fatma Güney

COMP411/511 - Fall 2024

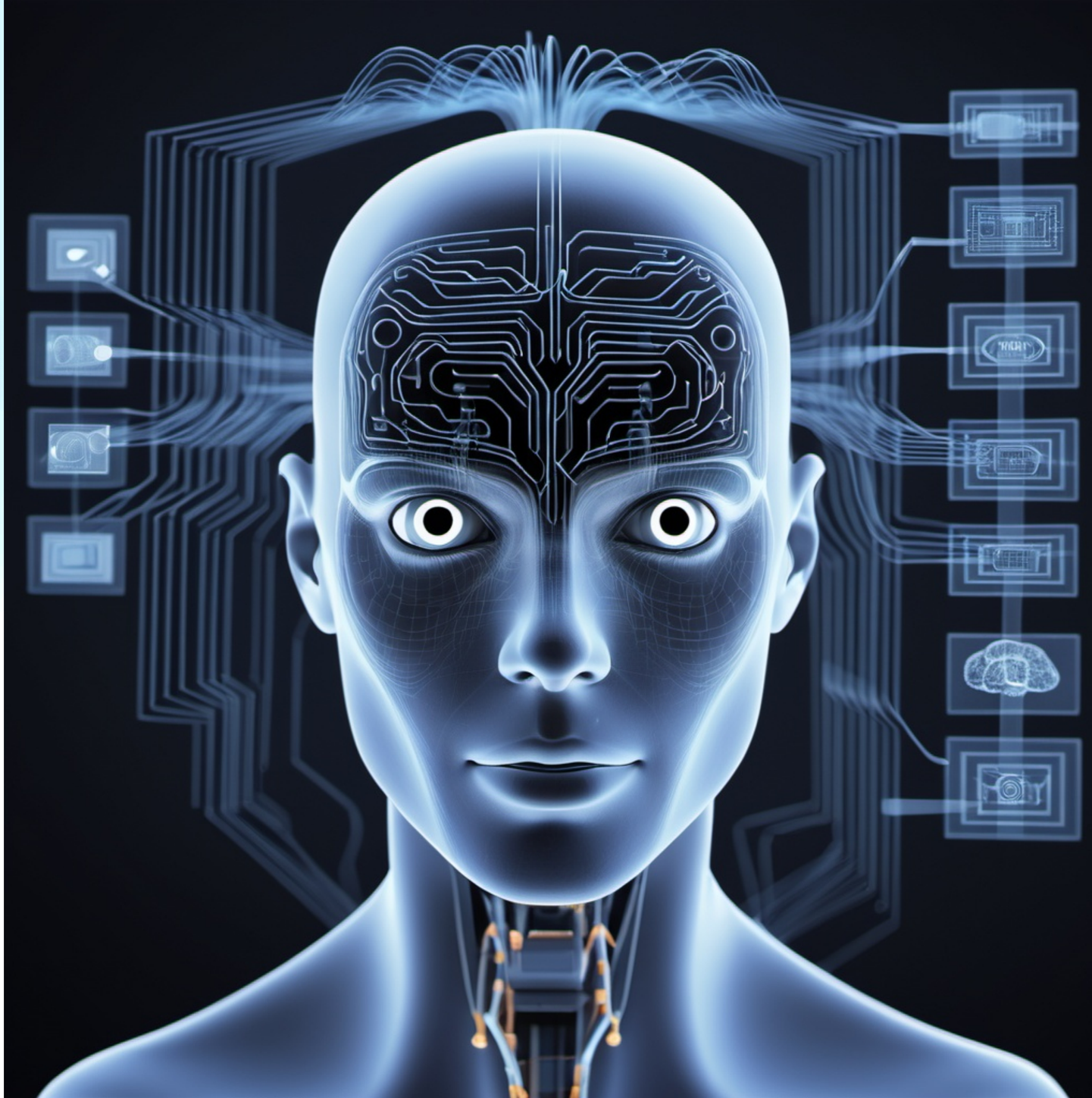
PLAN

Organization

Introduction

History

Prompt: *An introductory image for a course called "Computer Vision with Deep Learning"*



TEAM

Fatma Güney

Th, 16:15-17:15
ENG203A

Sadra Safadoust

ENG123

Gökay Aydemir

ENG123

COURSE MATERIALS

- 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/
- Goodfellow, Bengio, Courville: Deep Learning
  - http://www.deeplearningbook.org
- Deisenroth, Faisal, Ong: Mathematics for Machine Learning
  - https://mml-book.github.io

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`

RELATED COURSES

- Andreas Geiger: Computer Vision, Deep Learning
  - https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/autonomous-vision/lectures/

- Stanford CS231n: Deep Learning for Computer Vision
  - https://cs231n.stanford.edu/schedule.html

- Justin Johnson: Deep Learning for Computer Vision
  - https://web.eecs.umich.edu/~justincj/teaching/eecs498/WI2022/

PREREQUISITES

Math:

- Linear algebra (MATH107), Probability (ENGR200). 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 (COMP100)
- Introduction to Machine Learning (ENGR421)

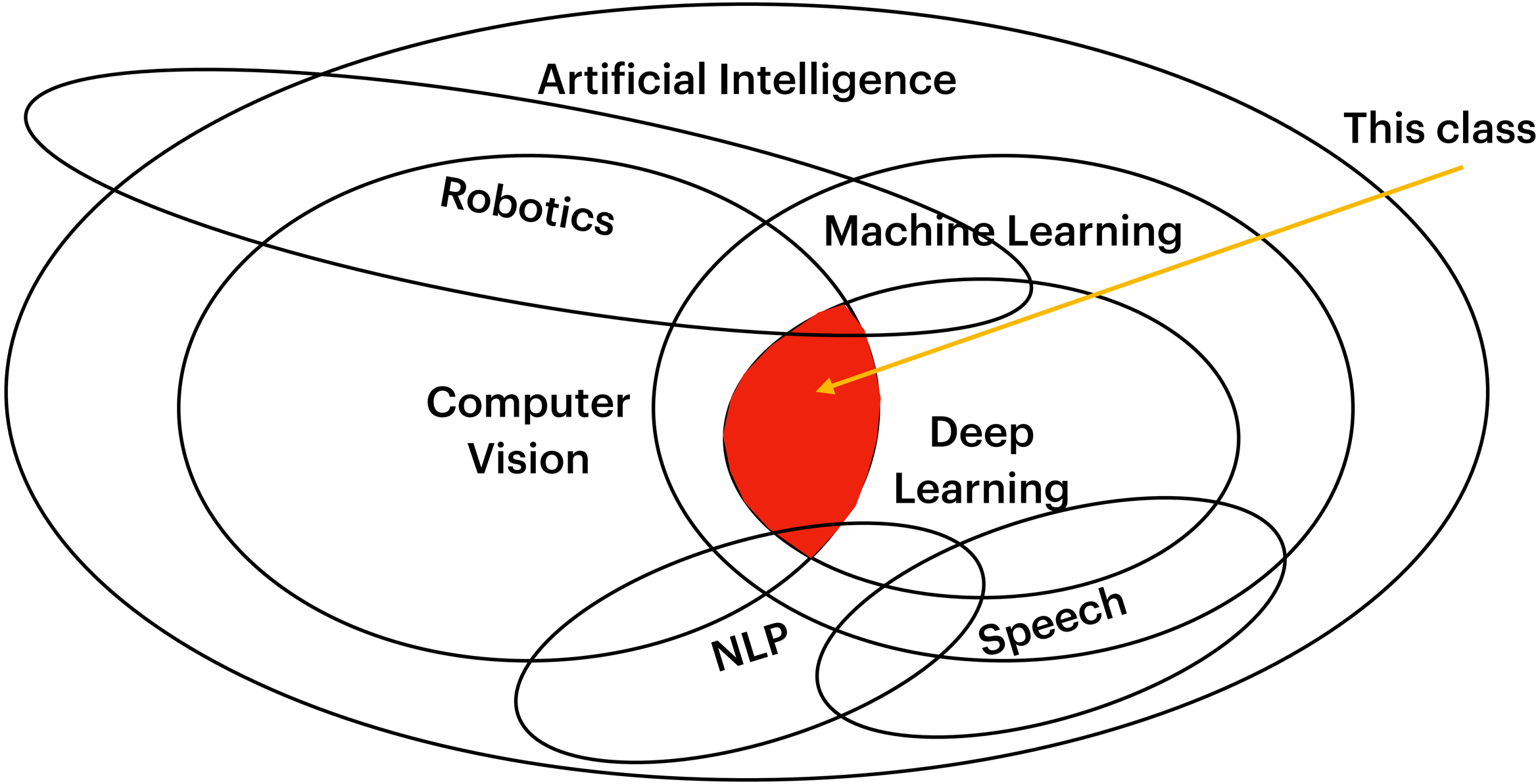
AGENDA

Goal: Understanding, implementing, training and debugging deep end-to-end neural network architectures for various tasks of computer vision. Image classification. Loss functions and optimization. Backpropagation. Convolutional neural networks. Recurrent neural networks for video and image analysis. Object detection and segmentation. Generative vision models.

- Deep Neural Networks
- Regularization and Optimization
- Convolutional Neural Networks
- Sequence Models
- Attention and Transformers
- Object Detection and Segmentation
- 3D Vision
- Deep Generative Models
- Self-Supervised Learning
- Visualizing and Understanding

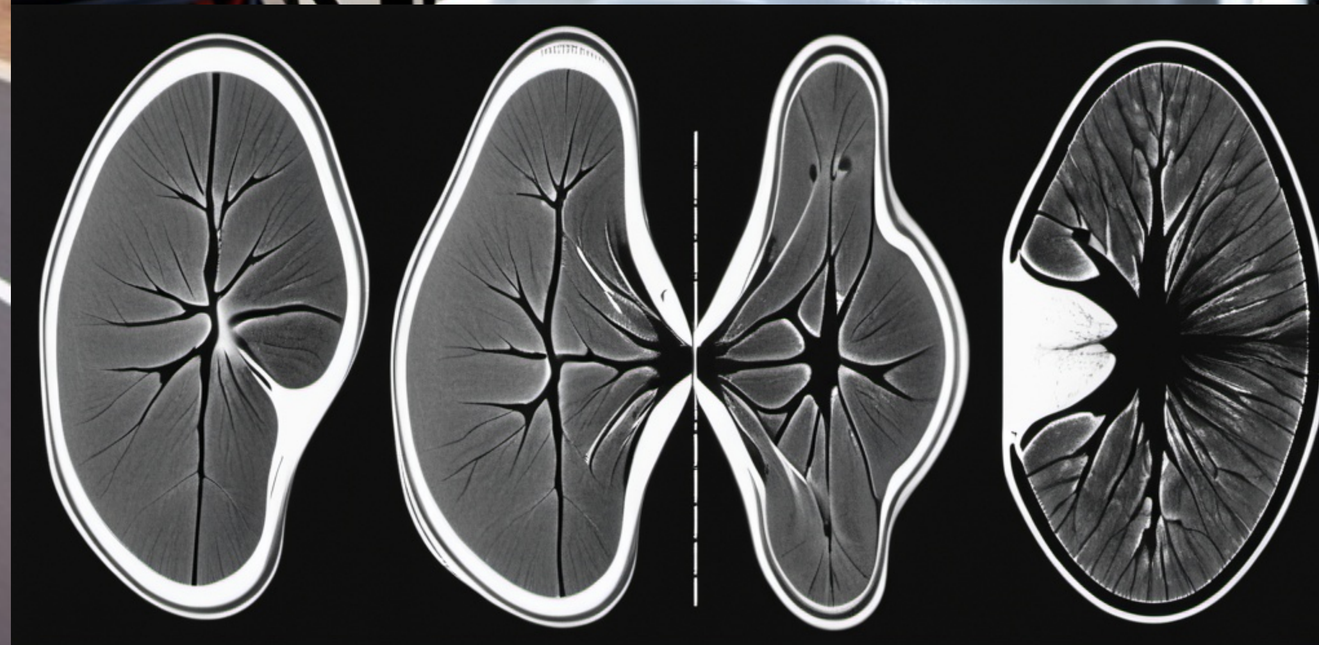
GRADING

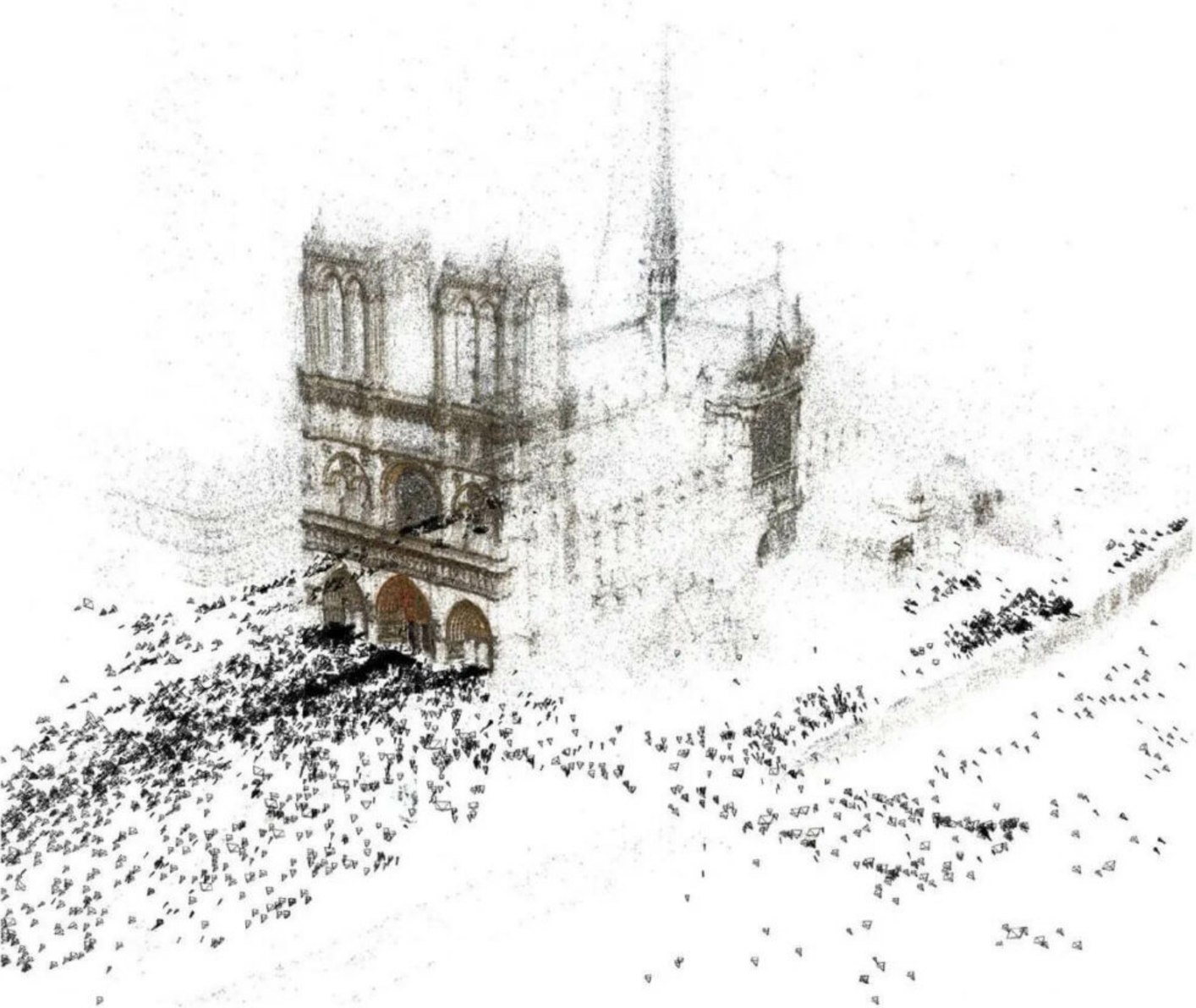
	Description	%
Homework	Programming Assignments	35
Quiz	In-class	15
Written Exam	Final	20
Project	Term Project	30


Artificial Intelligence
This class
Robotics
Machine Learning
Computer Vision
Deep Learning
NLP
Speech

Goal of Computer Vision

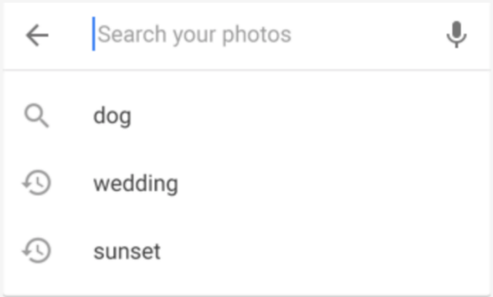
is to convert light into meaning
(geometric, semantic, . . .)

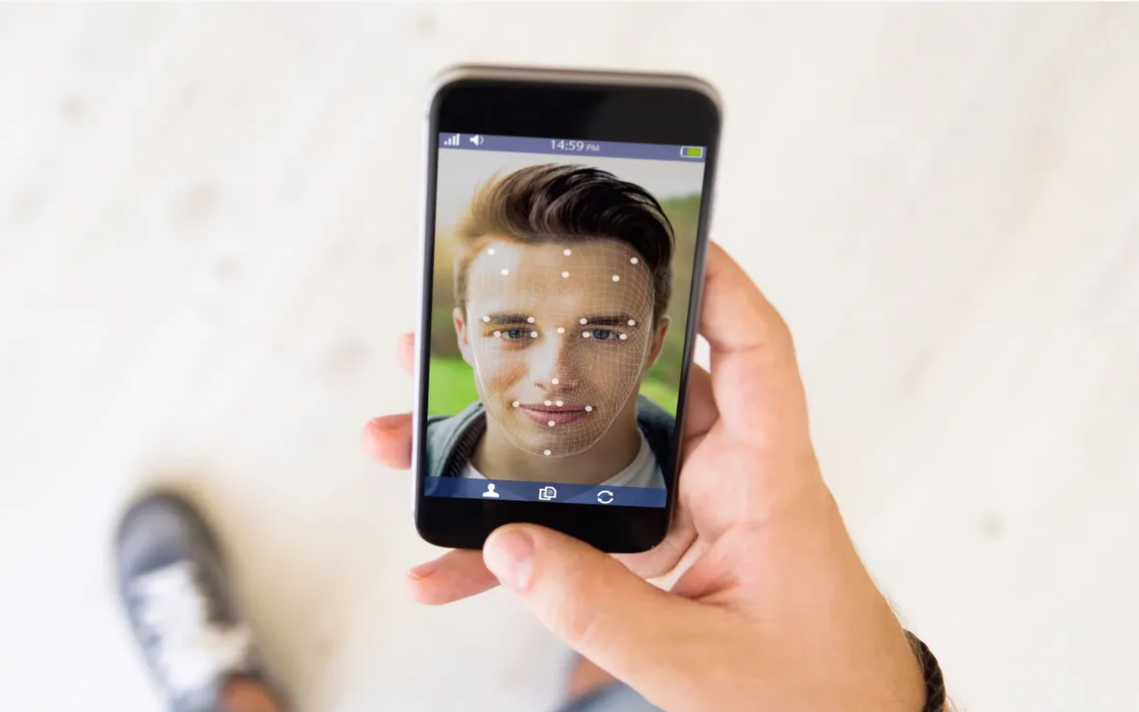




Find your photos faster

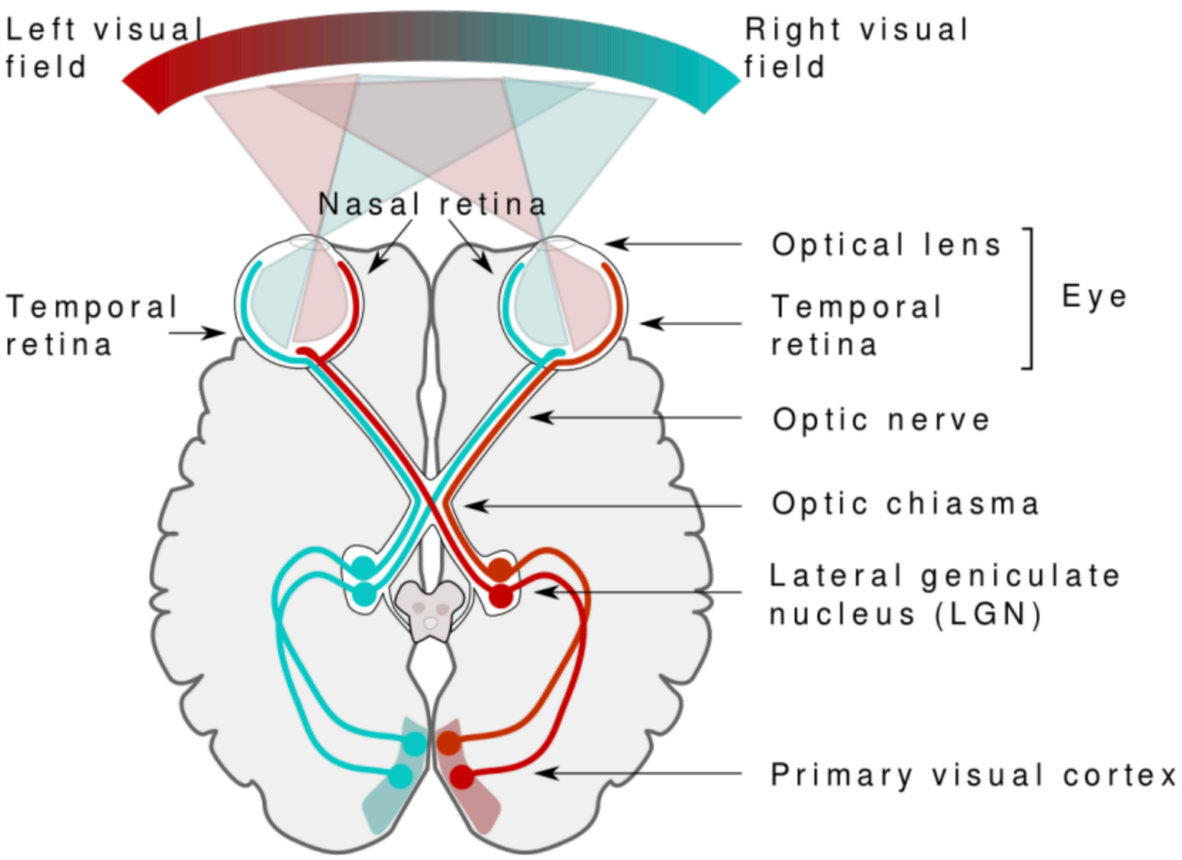
Your photos are organized and searchable by the places and things in them – no tagging required. Just search "dog" to find all the photos of your pup.





Computer Vision vs.
Biological Vision

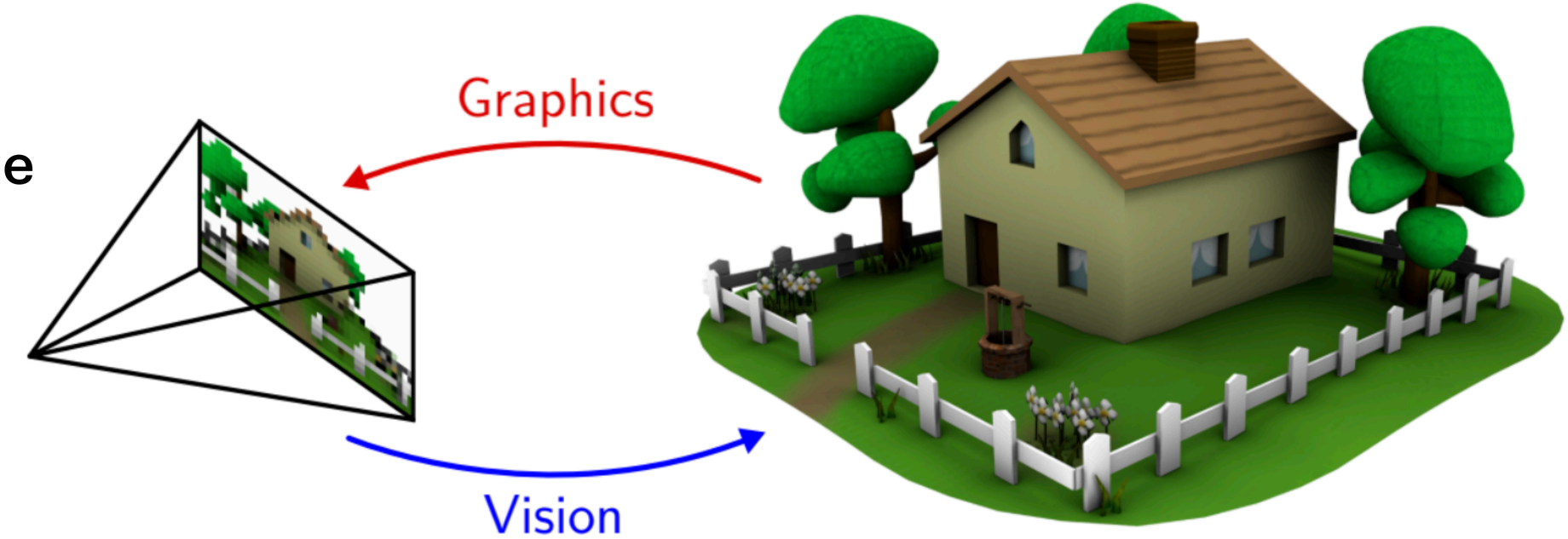
Over 50% of the processing in the human brain is dedicated to visual information.



Computer Vision vs.
Computer Graphics

3D world

2D image

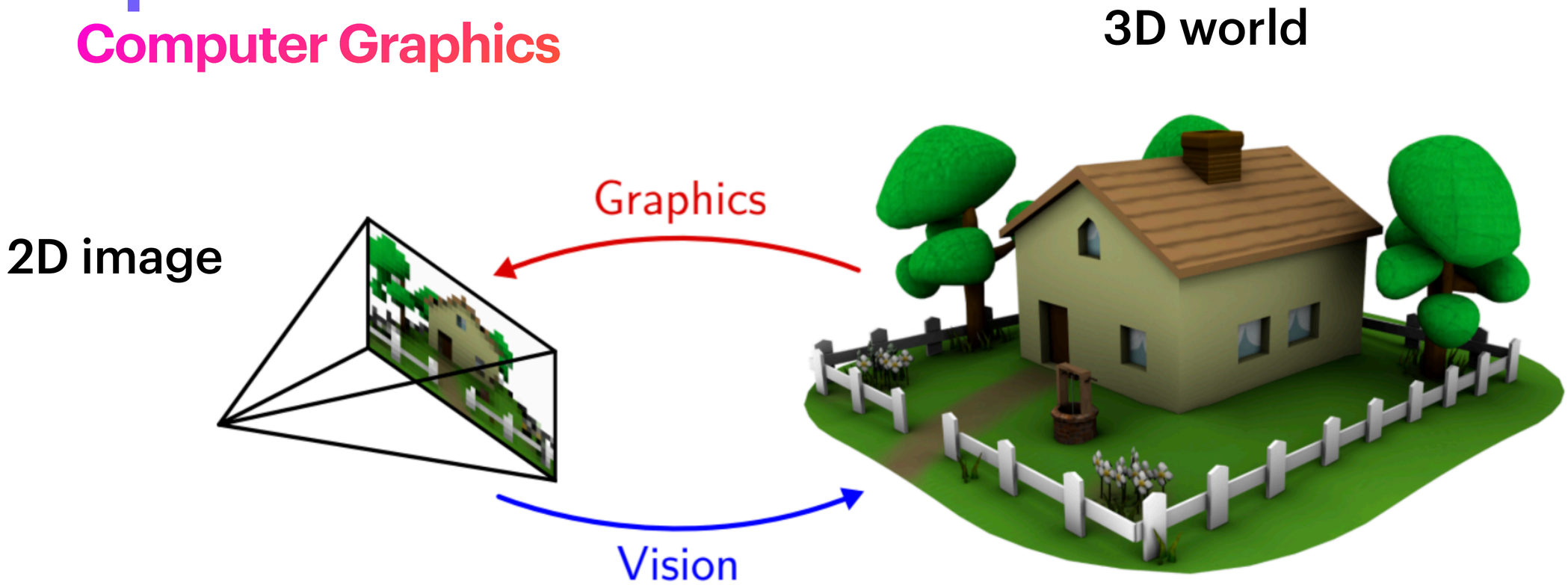


Pixel Matrix

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

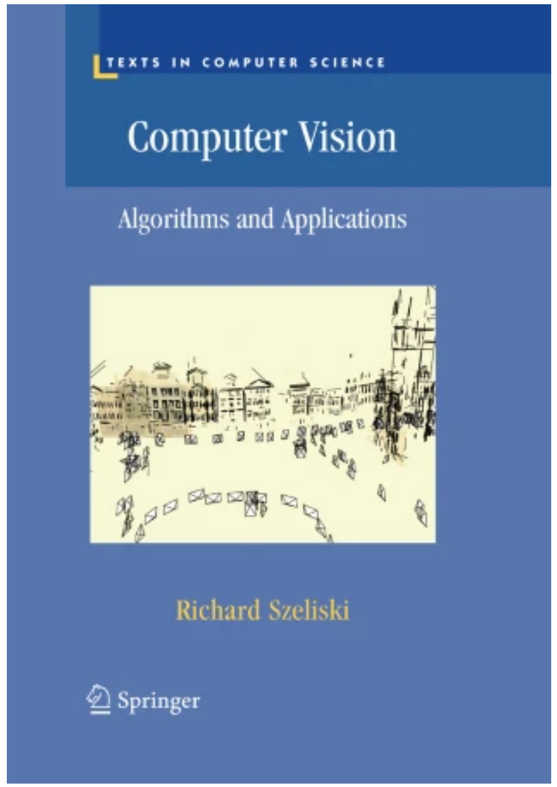
Computer Vision vs.
Computer Graphics

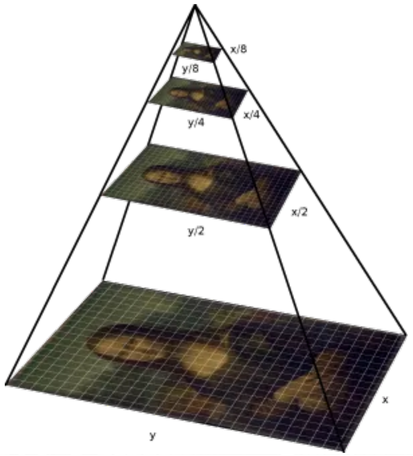


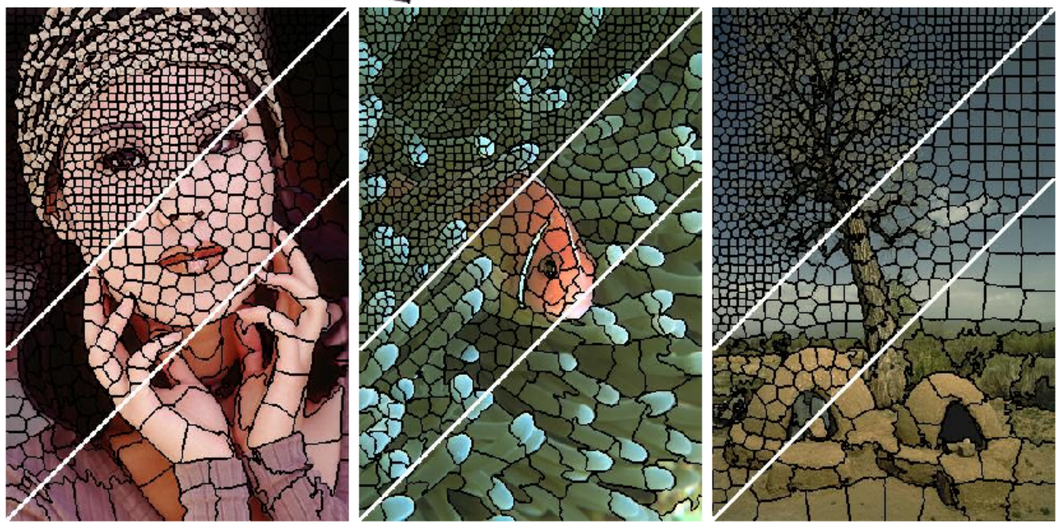
Computer Vision is an ill-posed inverse problem:

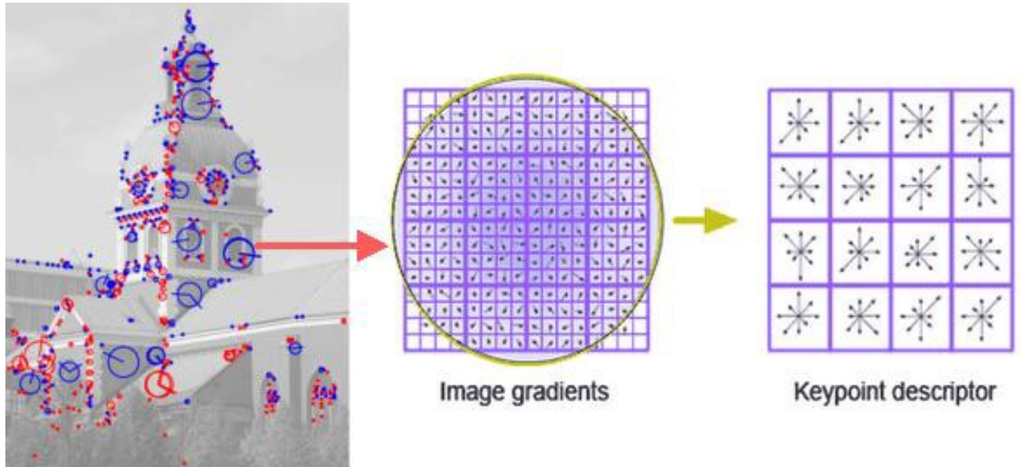
- Many 3D scenes yield the same 2D image
- Additional constraints (knowledge about world) required

Computer Vision vs.
Image Processing





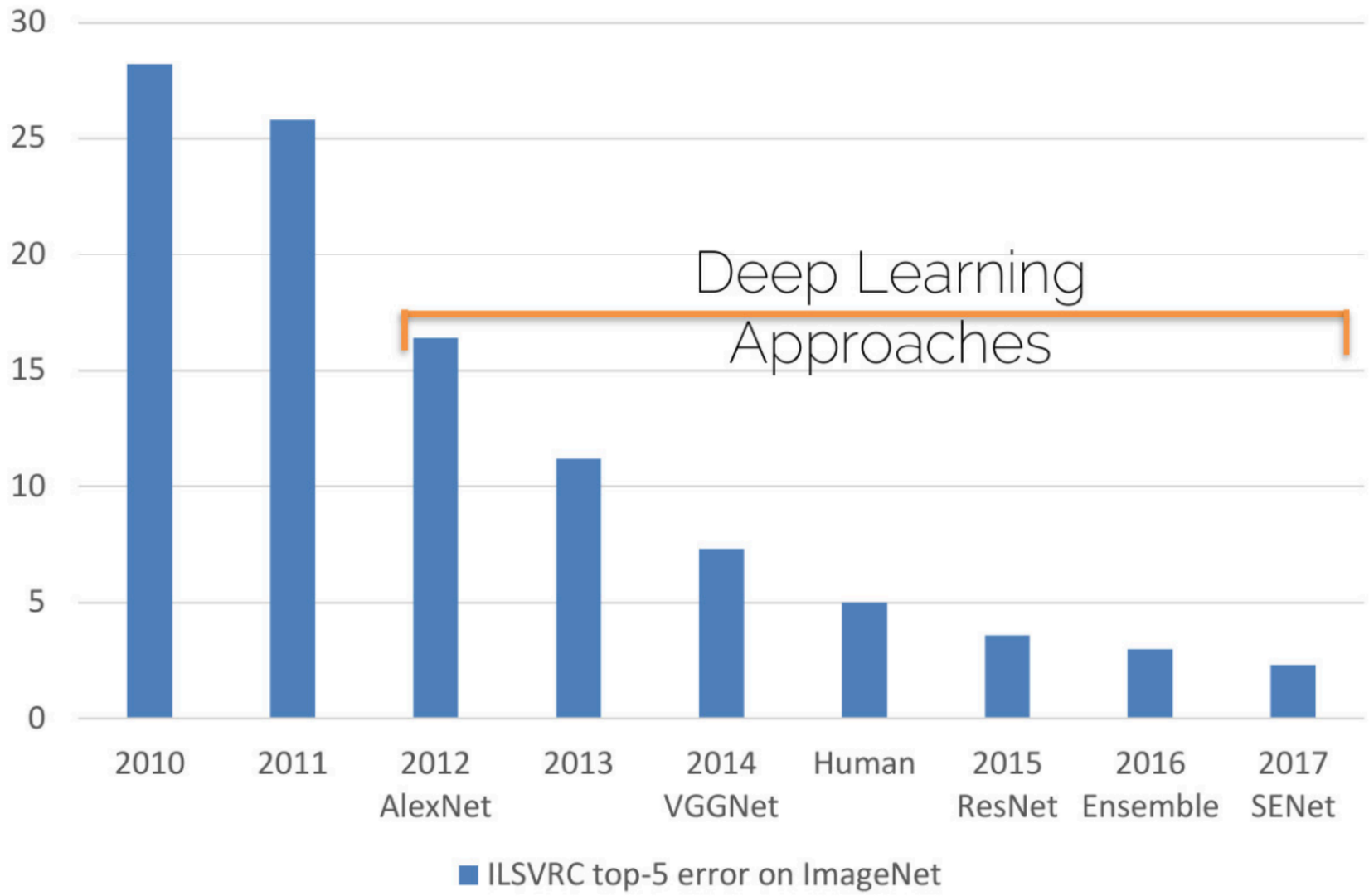




Computer Vision vs.
Machine Learning

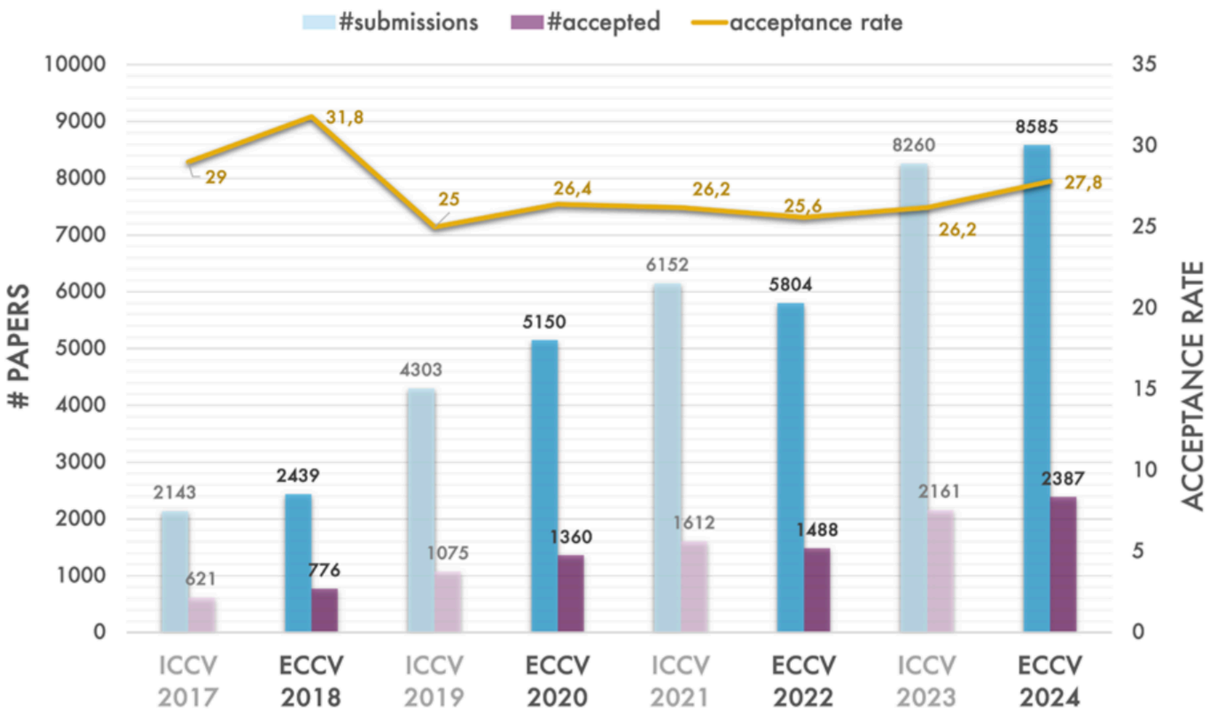
IMAGENET

The Deep Learning Revolution



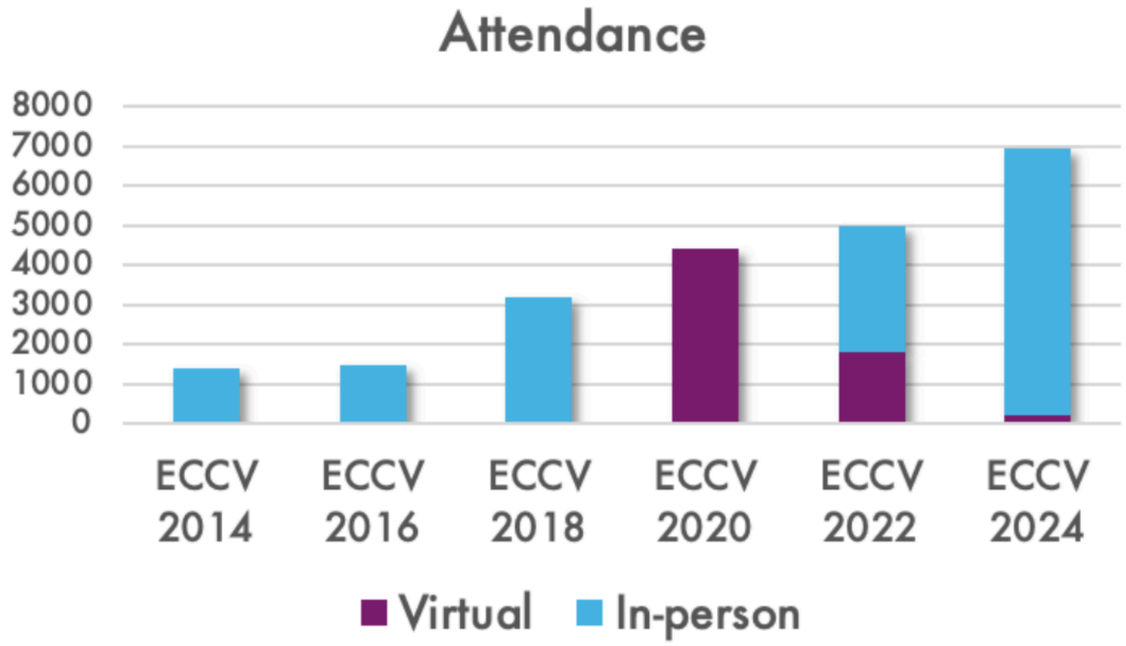
EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO
2024



EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO
2024



amazon | science
AMD
ByteDance
captions
Google
Hyperbolic
intel.
Lambda
美团
Meta
Microsoft
Qualcomm
SONY
THE AI INSTITUTE
Adobe
Alibaba Cloud × TONGYI
ANT RESEARCH
Baidu 百度
comet
Gatik
HYUNDAI MOTOR GROUP
Jane Street
LATITUDE
scale
CHEN INSTITUTE TIANQIAO & CHRISSY
TOYOTA RESEARCH INSTITUTE
VOXEL51
WAYVE
Weights & Biases
WYZE
ZOOX

ASML
B GARAGE
CEBS
DeepAuto.ai
helm.ai
Determined AI
iMerit
kitware
Lightning AI
Creators of PyTorch Lightning
NEC
NEC Laboratories America
nexdata
oppo
sapien
SIMULAR
synthesia
AKOOL
ANDURIL
AKOOL
ANDURIL
Disney Research
GE Aerospace
Disney Research
GE Aerospace
KUAISHOU
LG AI Research
KUAISHOU
LG AI Research
PROPHESEE METAVISION TECHNOLOGIES
PROPHESEE METAVISION TECHNOLOGIES
Tencent
TENYKS
WAYMO

Why is Computer Vision Hard?

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

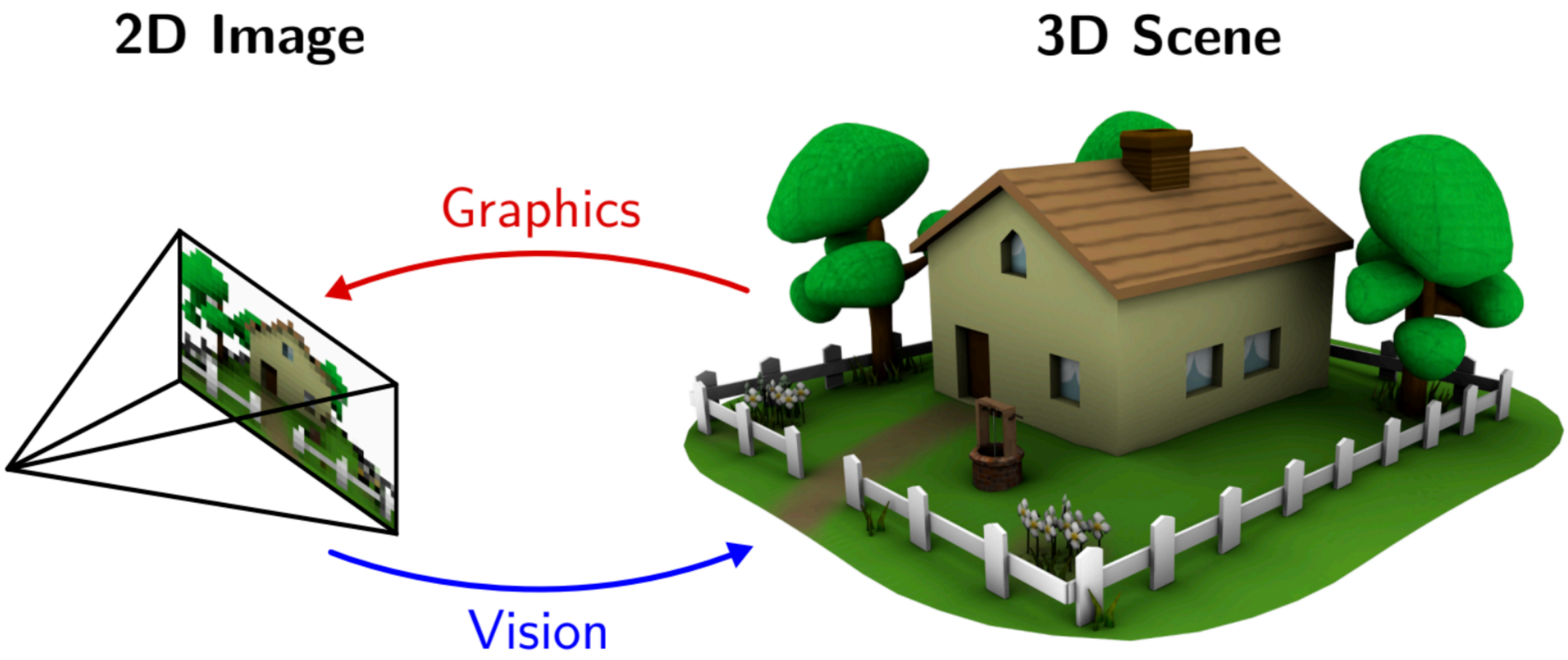
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

The Summer Vision Project. MIT AI Memos, 1966.

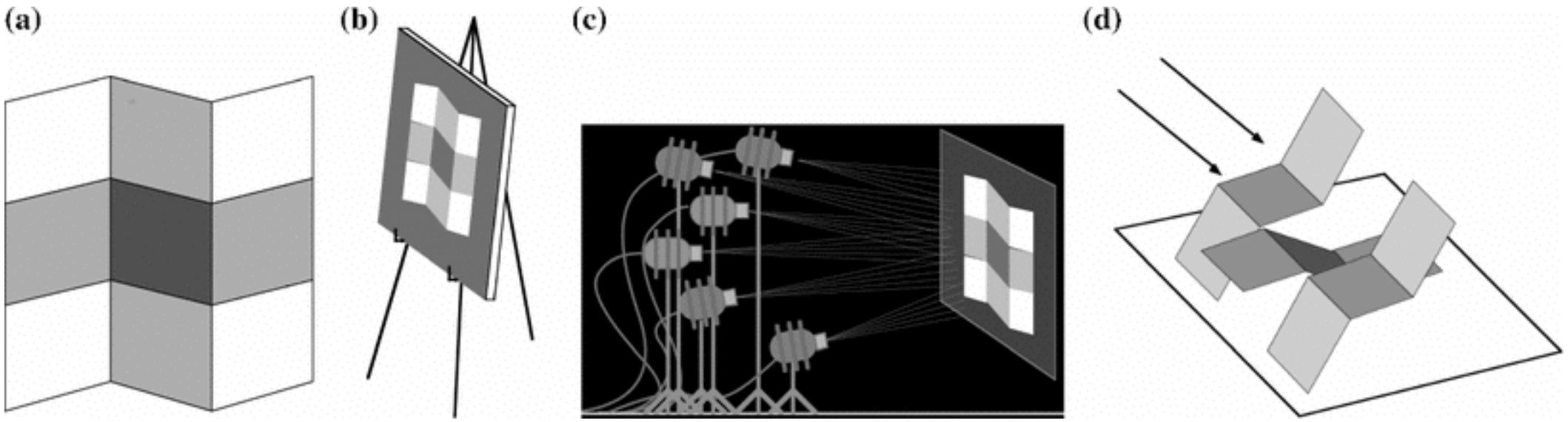
What we see

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	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 computer sees

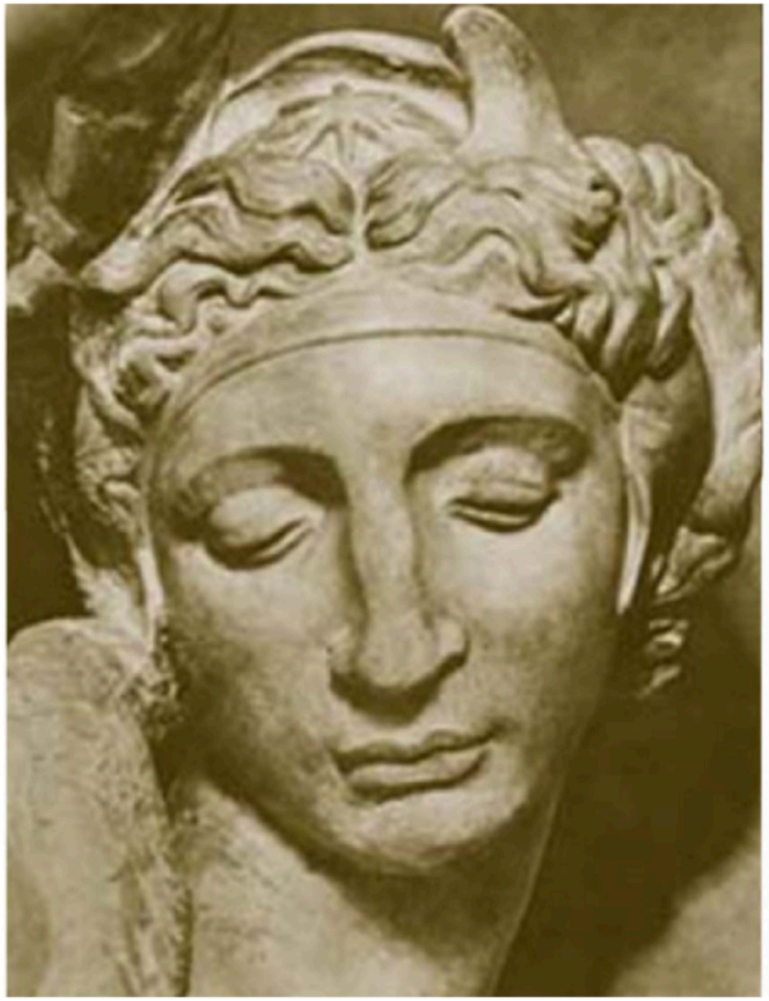


Images are 2D Projections of the 3D World.

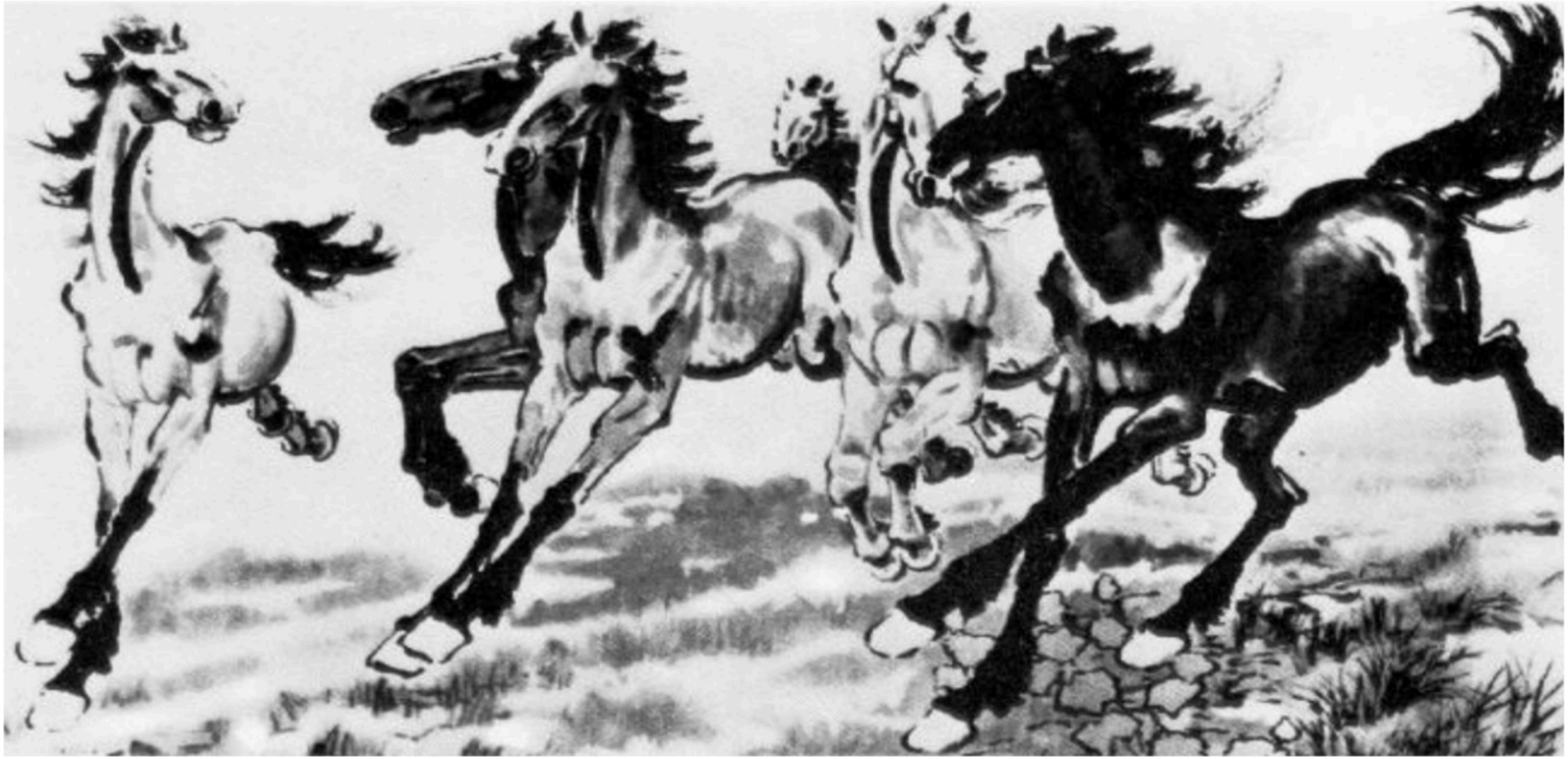


Adelson and Pentland's workshop metaphor: To explain an image (a) in terms of reflectance, lighting and shape, (b) a painter, (c) a light designer and (d) a sculptor will design three different, but plausible, solutions.

Ames Room Illusion



Viewpoint Variation



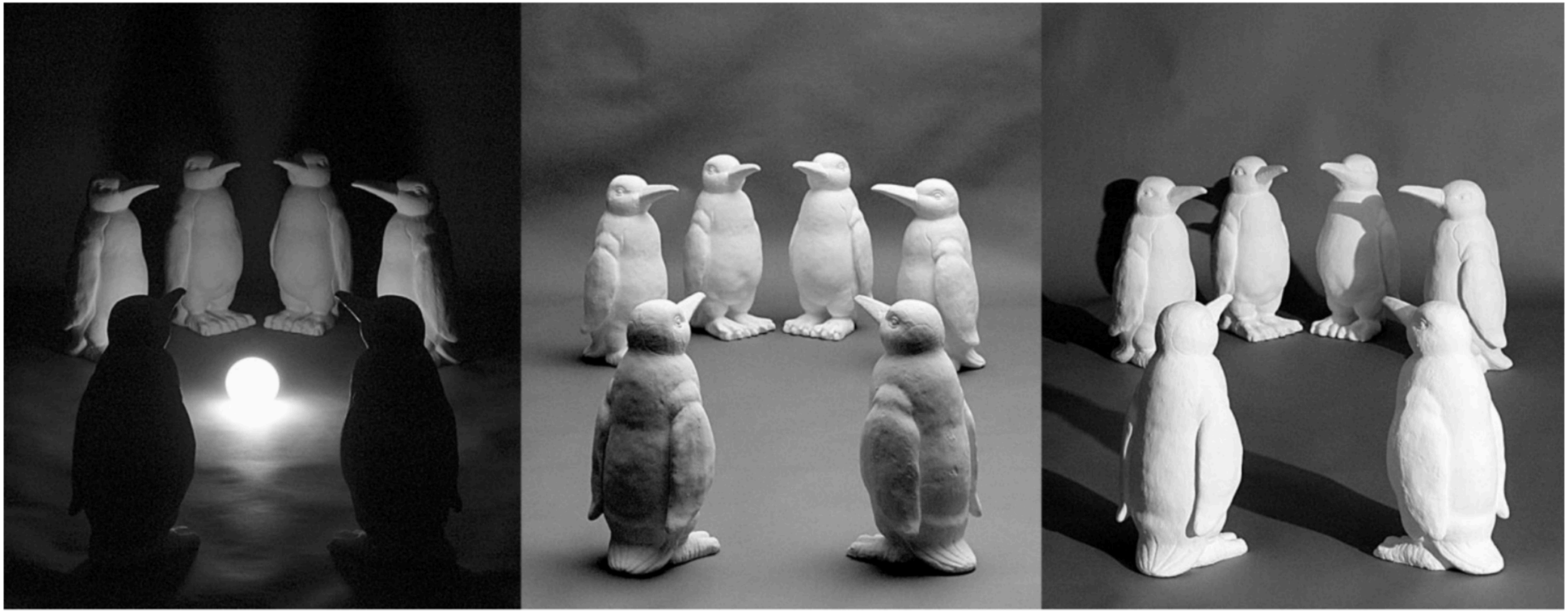
Deformation

Occlusion

Illumination

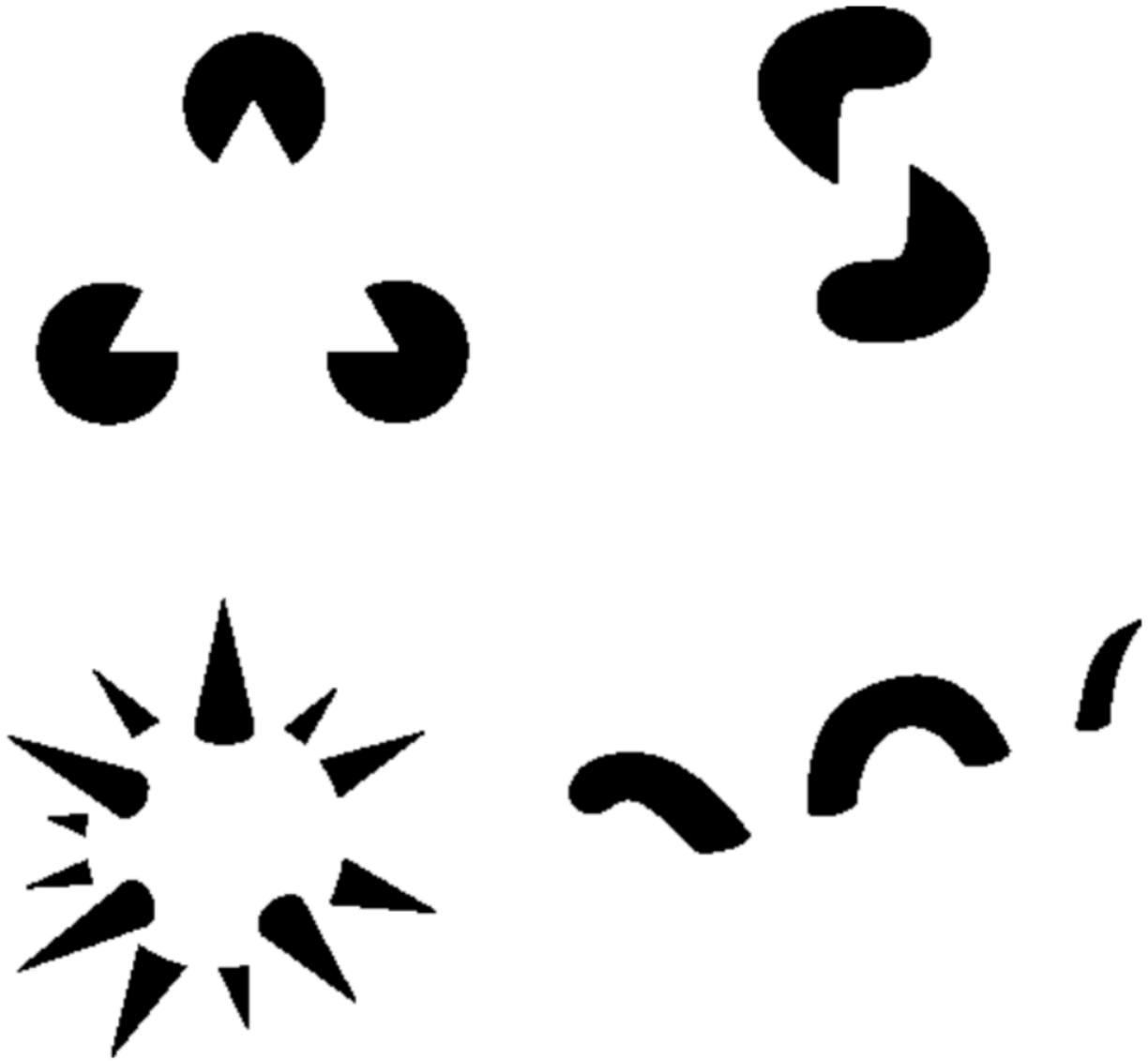
Illumination

Illumination



Illumination

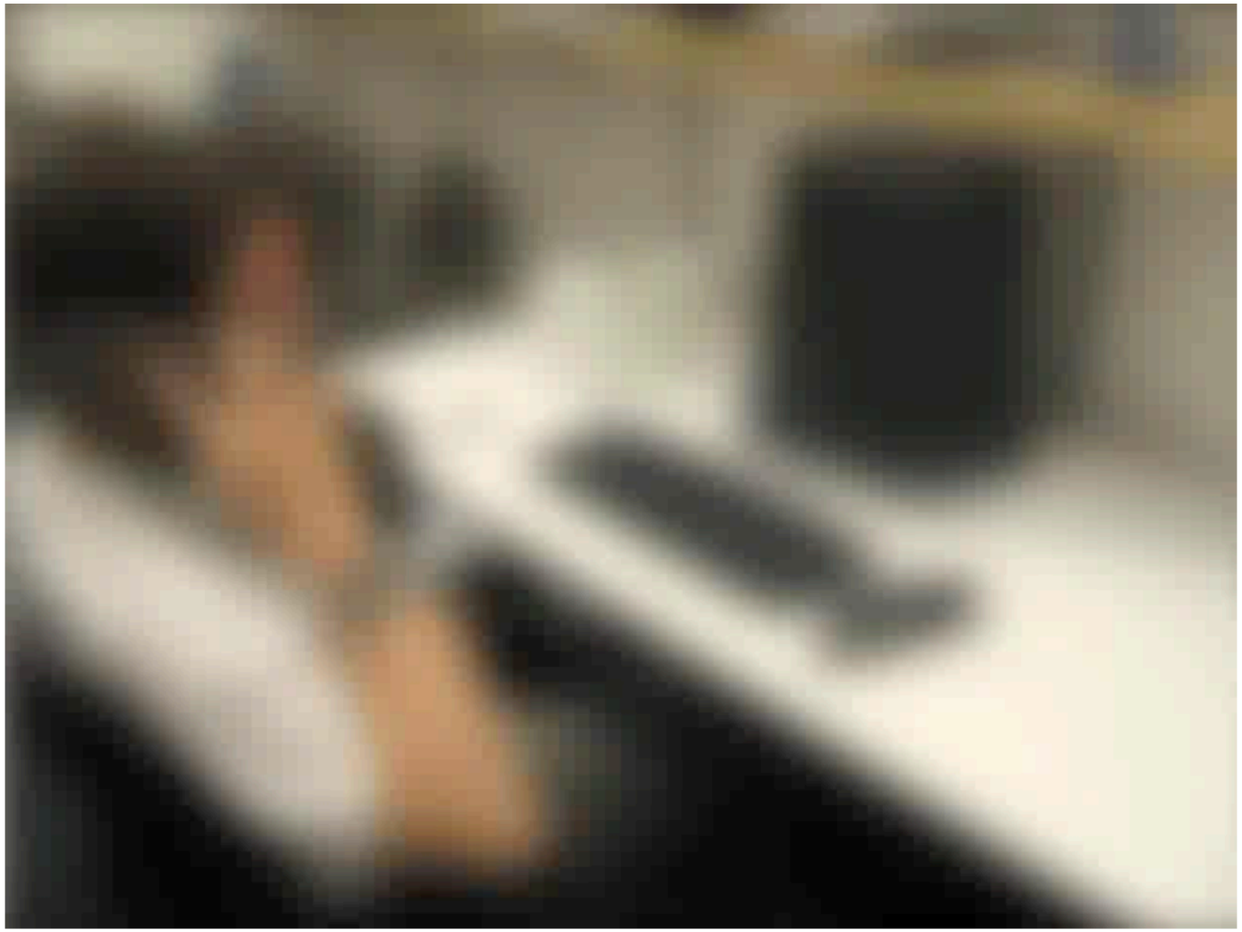
Motion



Perception vs. Measurement

Perception vs. Measurement

Local Ambiguities



Local Ambiguities

Local Ambiguities

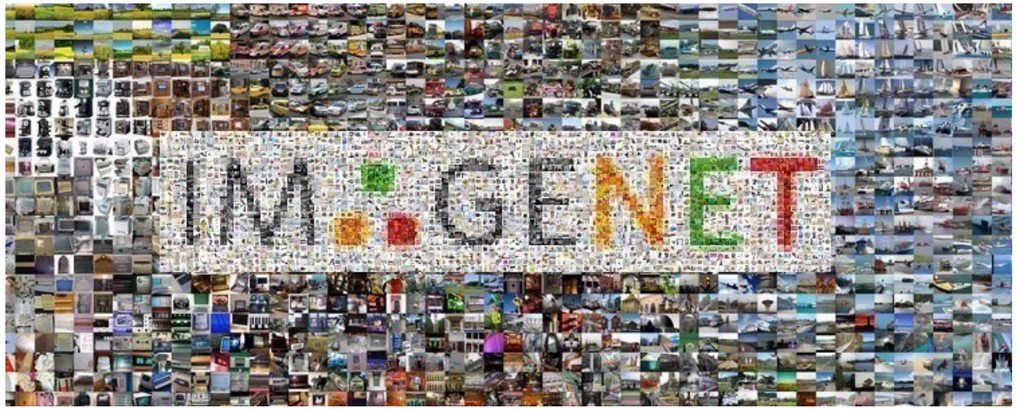
Intra-Class Variations

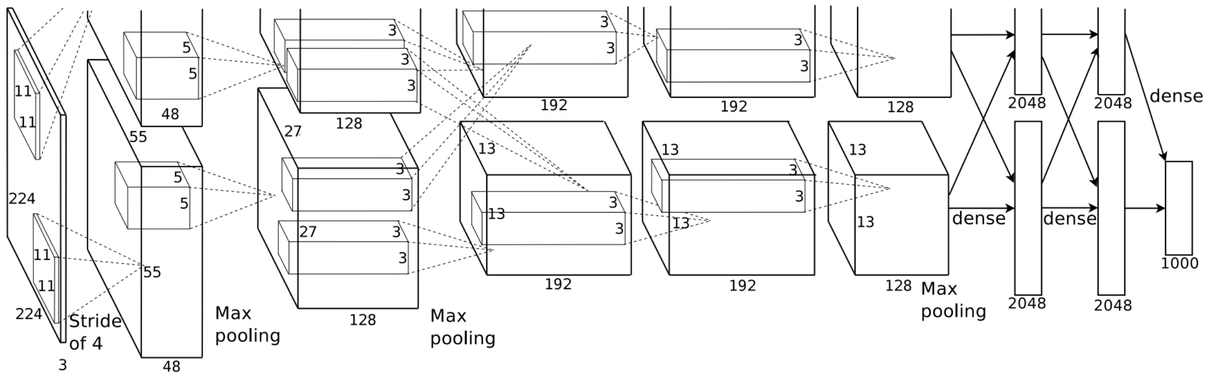
Number of Object Categories

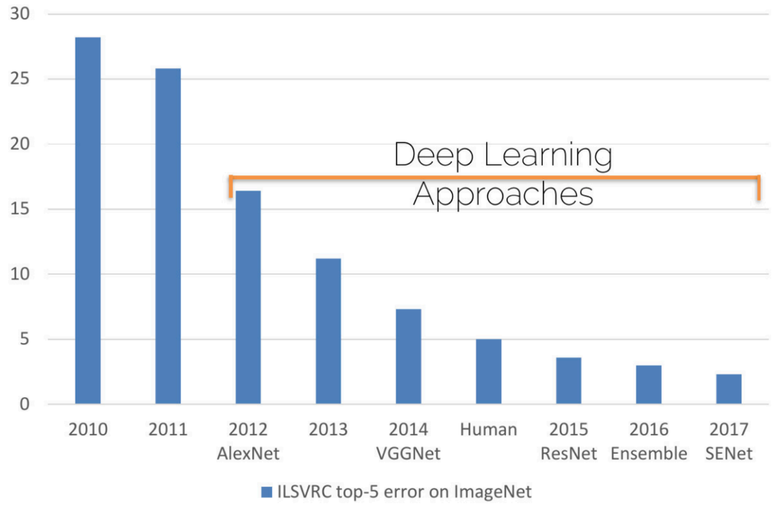
A Brief History of Computer Vision with Deep Learning

ImageNet and AlexNet
2009-20012

- **ImageNet**
  - Recognition benchmark (ILSVRC)
  - 10 million images, 1000 categories annotated.
- **AlexNet**
  - First neural network to win ILSVRC via **GPU + deep models + data**
  - Sparked the deep learning revolution



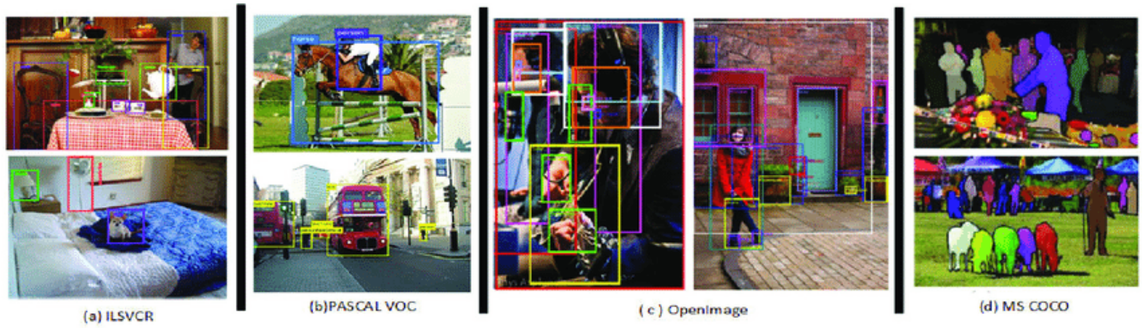


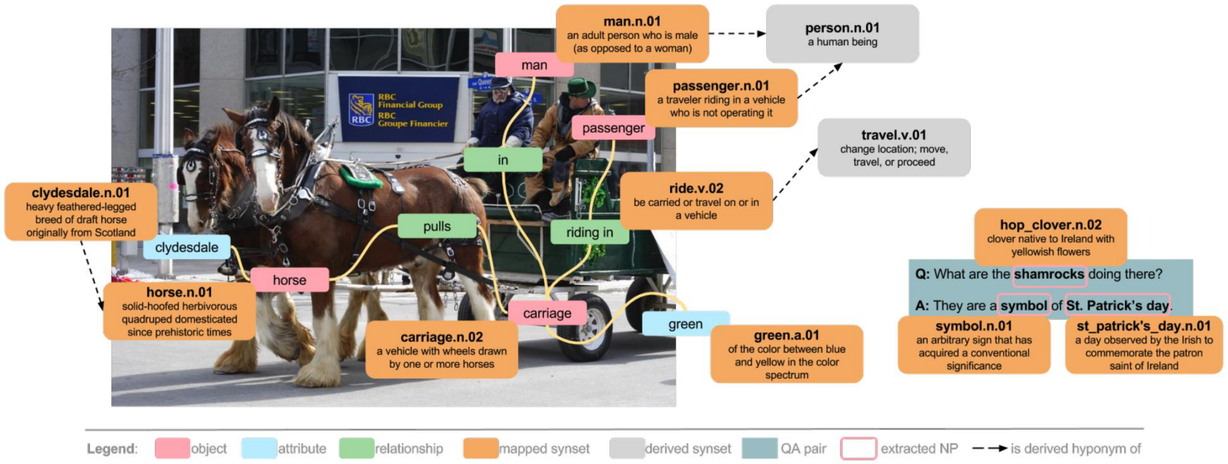


Golden Age of Datasets

2002-Now

- KITTI, Cityscapes: Self-driving
- PASCAL, MS COCO: Recognition
- ShapeNet, ScanNet: 3D DL
- Visual Genome: Vision/Language





Golden Age of Datasets
2002-Now

Synthetic Data:

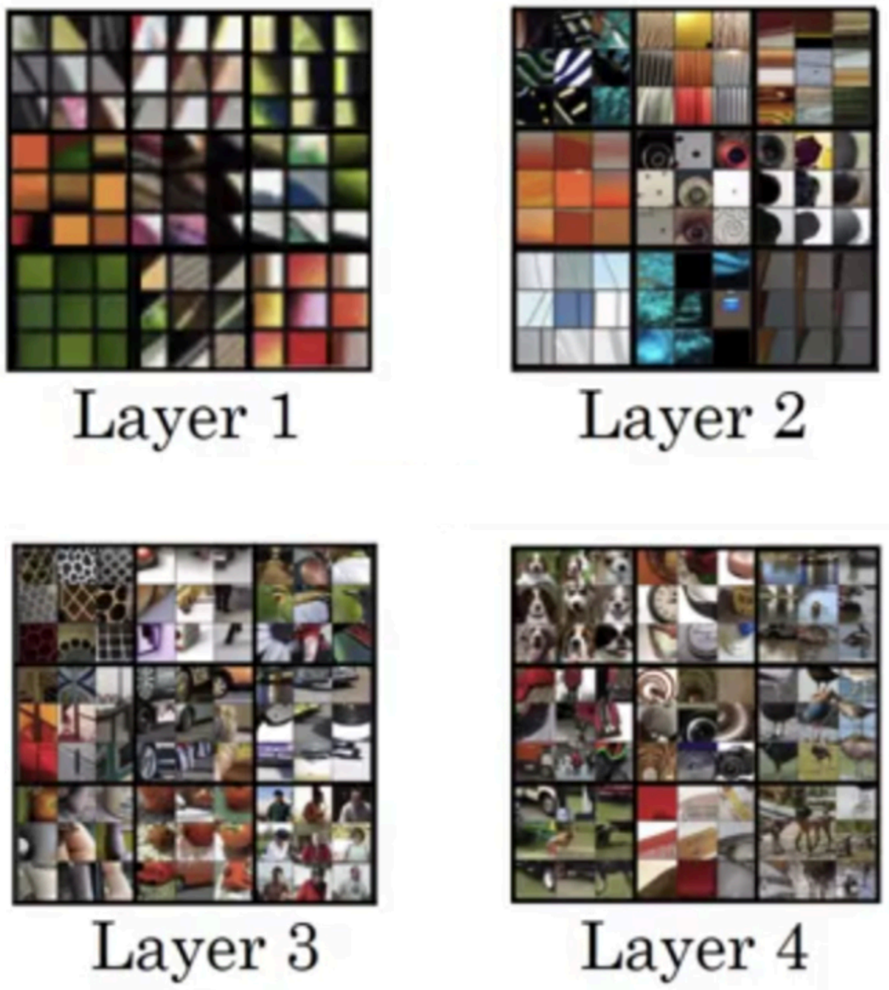
- Annotating real data is expensive
- Led to surge of synthetic datasets
- Creating 3D assets is also costly
- But even very simple 3D datasets proved tremendously useful for pre-training (e.g., in optical flow)

Visualization

2014

Goal: provide insights into what the network (black box) has learned.

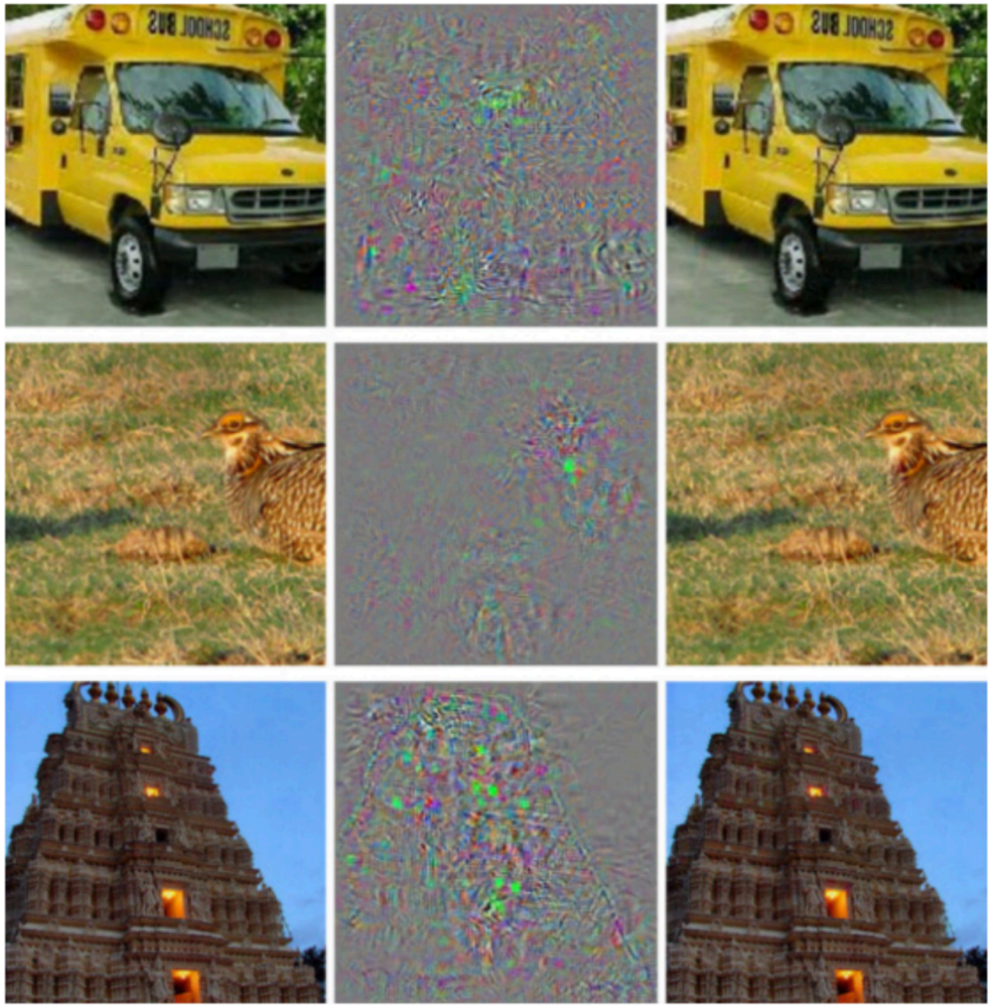
- Visualized image regions that most strongly activate various neurons at different layers of the network
- Found that higher levels capture more abstract semantic information



Adversarial Examples

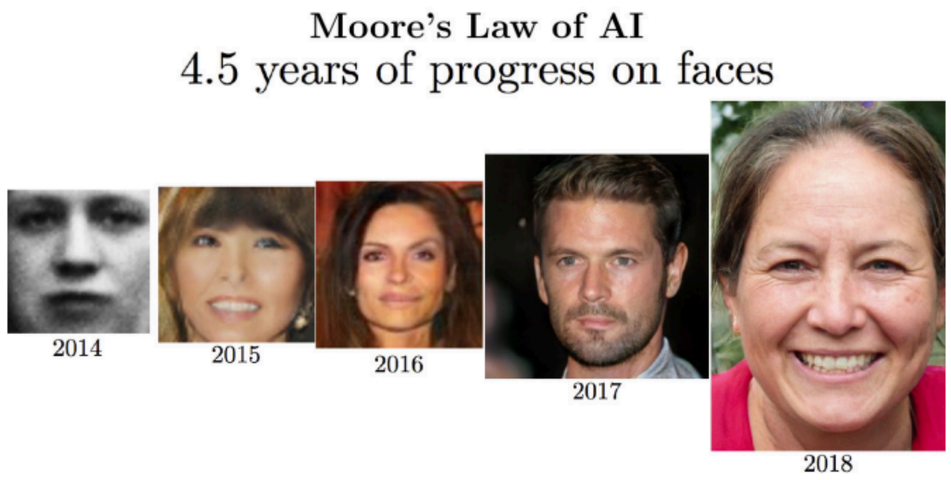
2014

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



Generative Adversarial Networks
2014

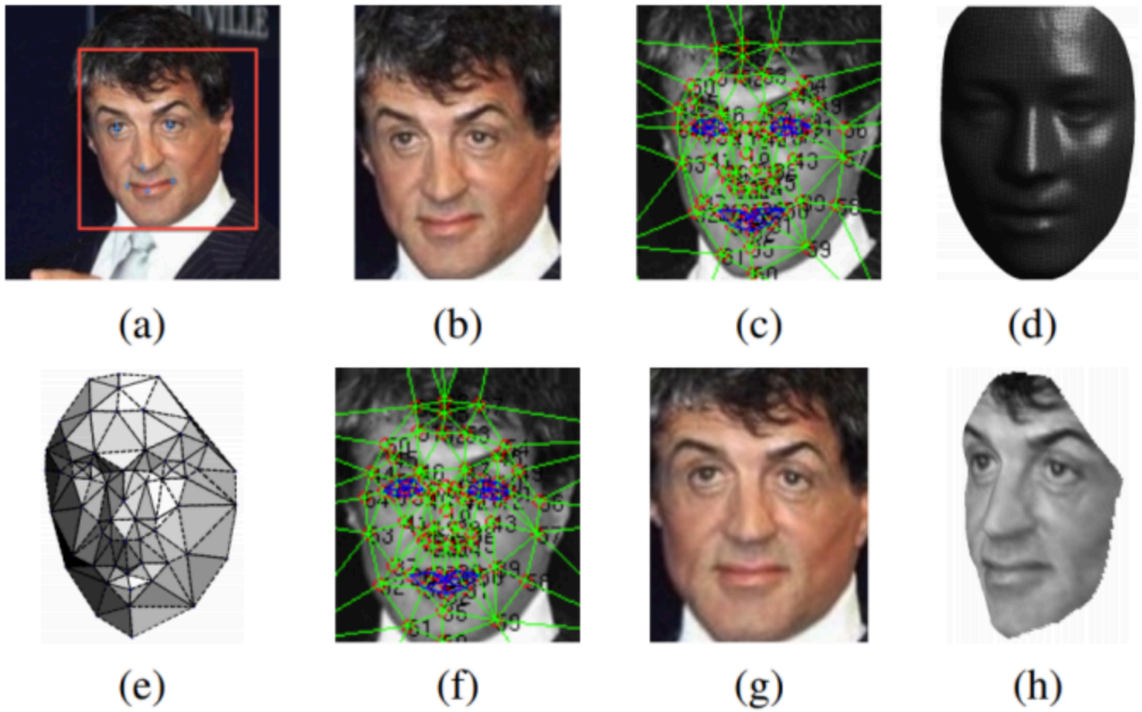
- Deep generative models (VAEs, GANs) produce compelling images
- StyleGAN2 is state-of-the-art
- Results on faces hard to distinguish from real images
- Active research on image translation, domain adaptation, content and scene generation and 3D GANs

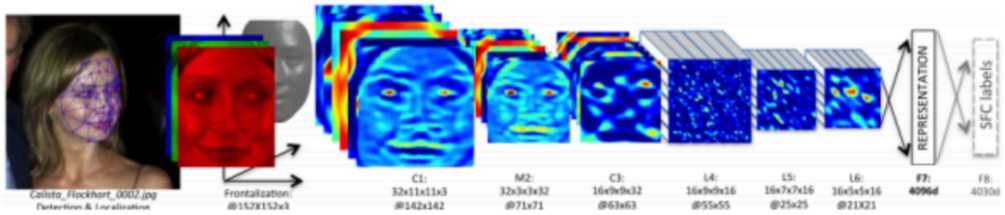


DeepFace

2014

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

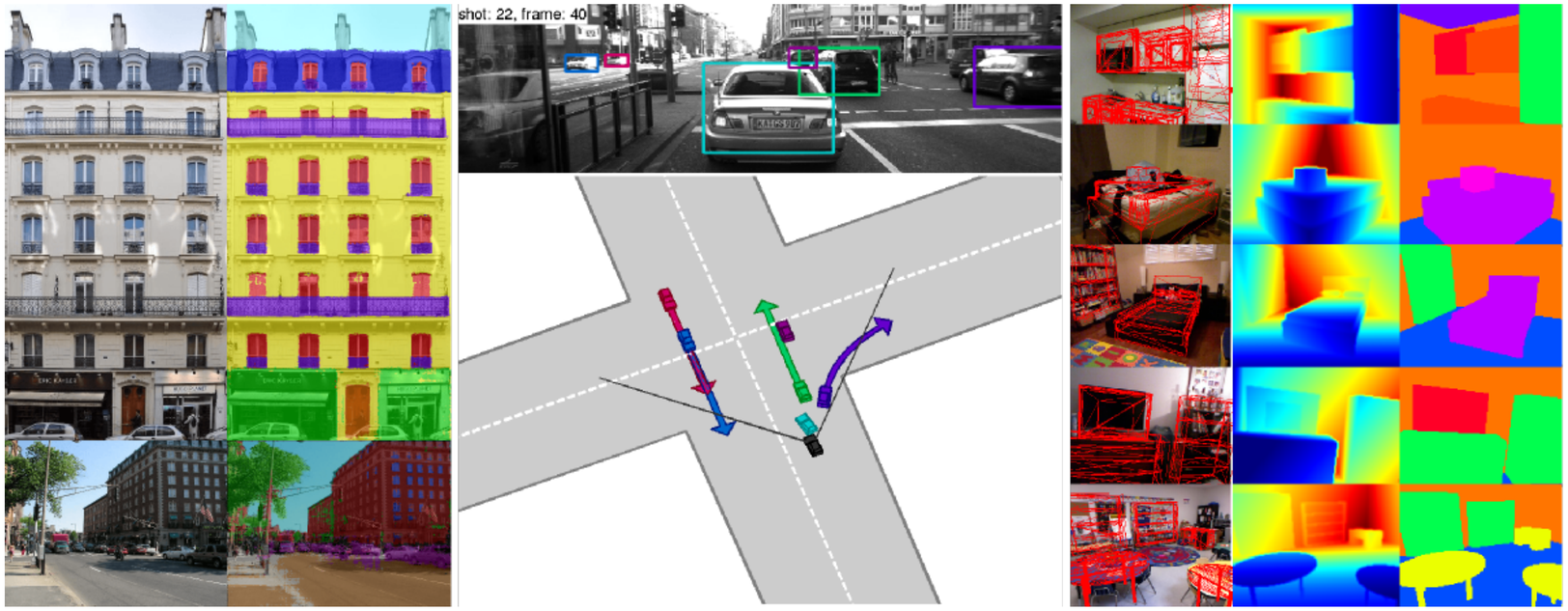




3D Scene Understanding

2014

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



3D Scanning

2014

- 3D scanning techniques allow for creating accurate replicas
- Debevec’s team scans Obama
- Exhibition in Smithsonian

Style Transfer

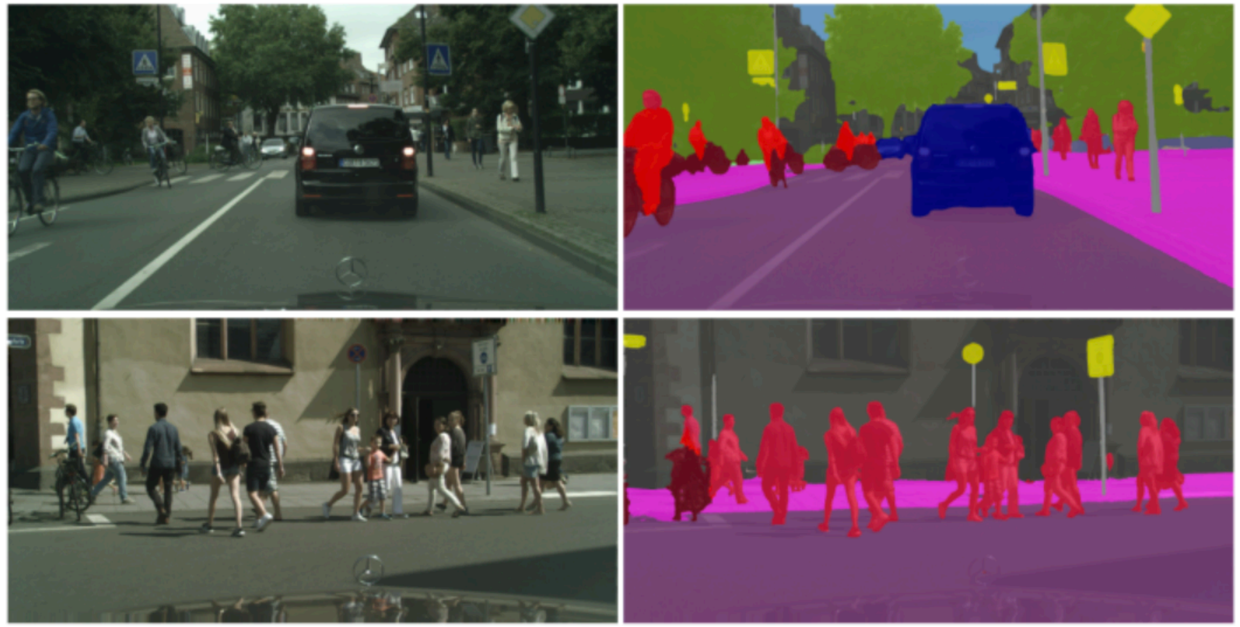
2016

- Manipulate photograph to adopt style of a another image (painting)
- Uses deep network pre-trained on ImageNet for disentangling content from style

`https://deepart.io/`

Semantic Segmentation
2015-2017

- 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



Mask R-CNN

2017

- 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

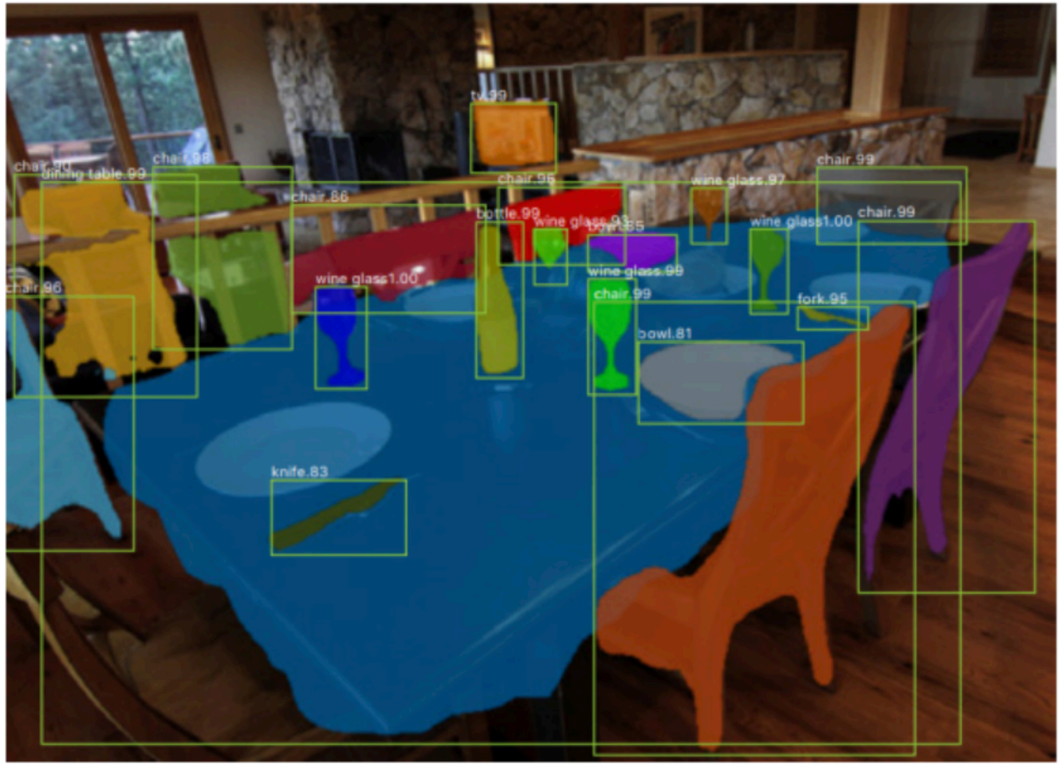
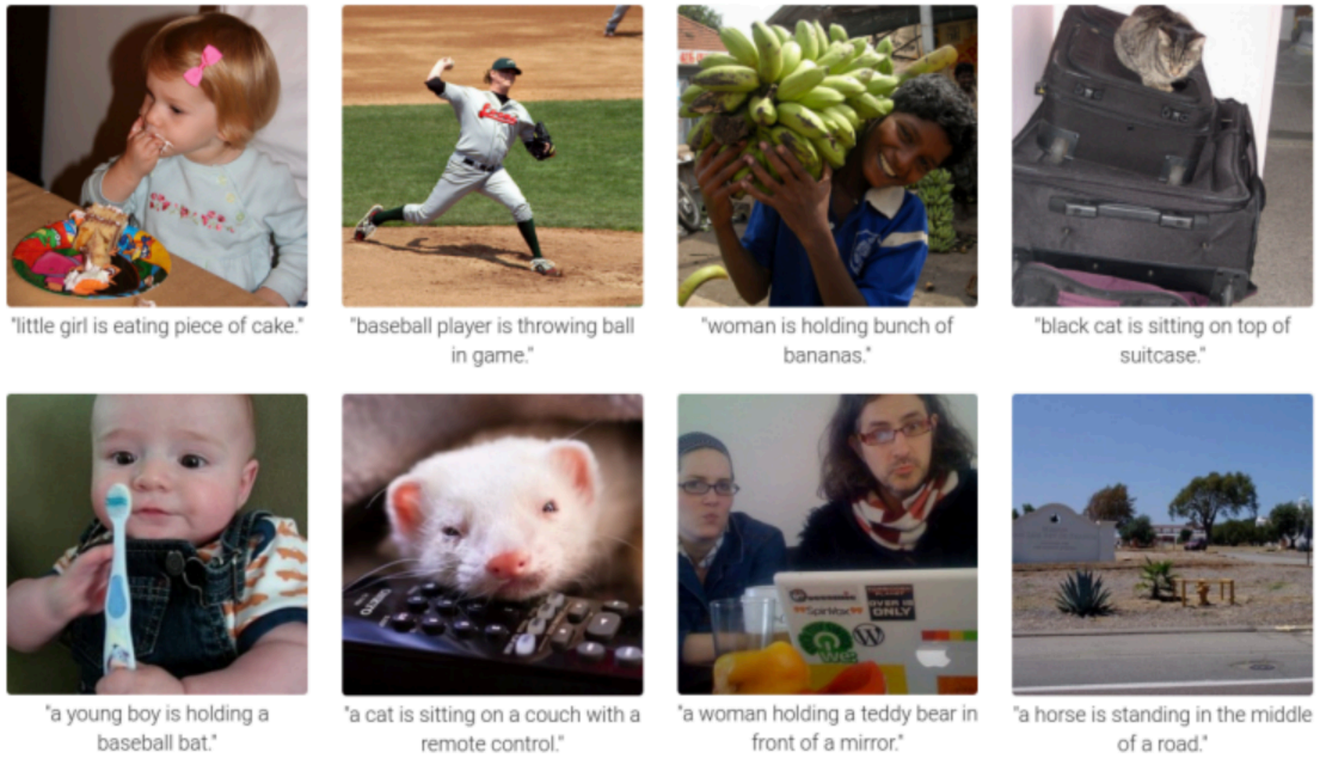


Image Captioning

2017

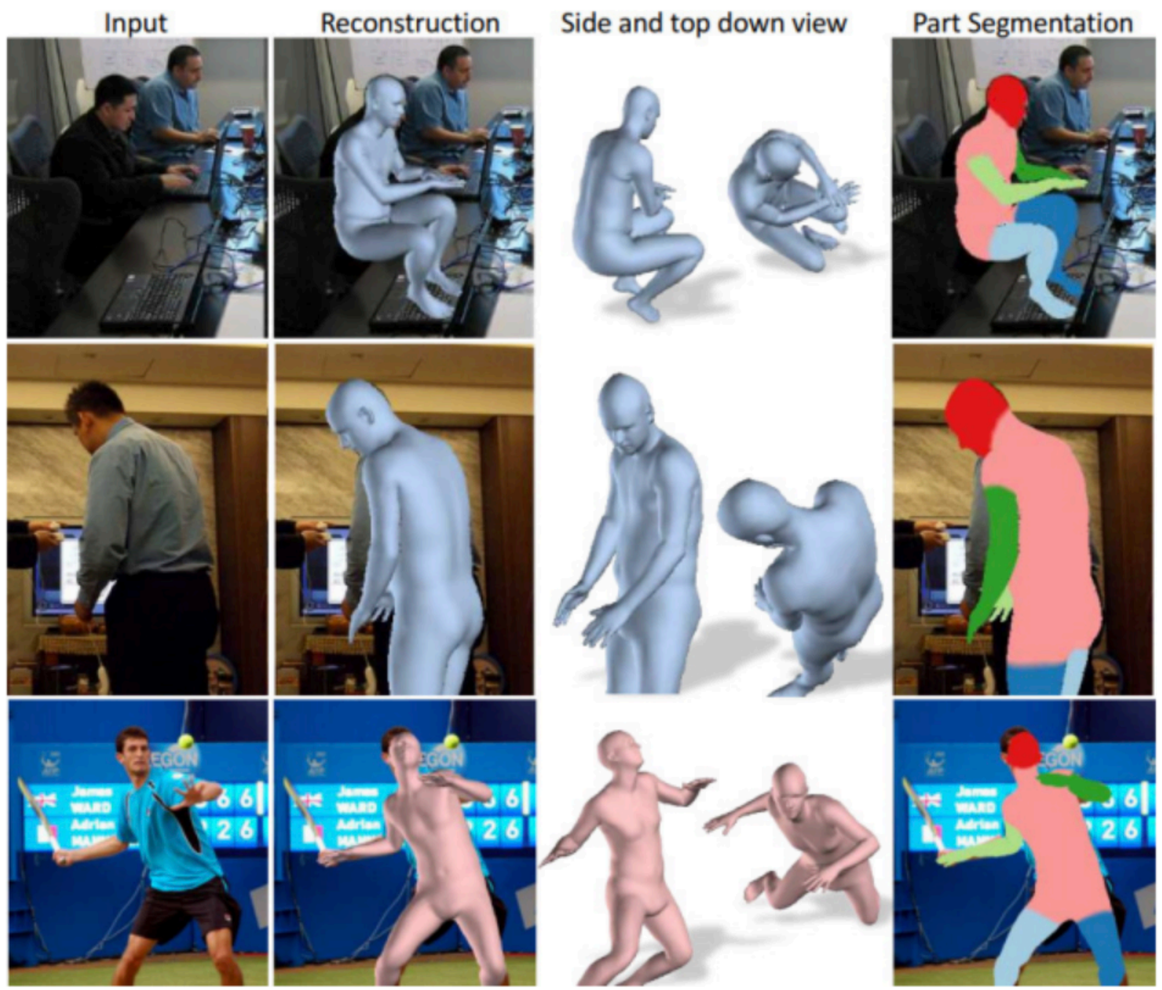
- Growing interest in combining vision with language
- Several new tasks emerged including image captioning and visual question answering
- However, models still lack understanding / commonsense



Human Shape & Pose

2018

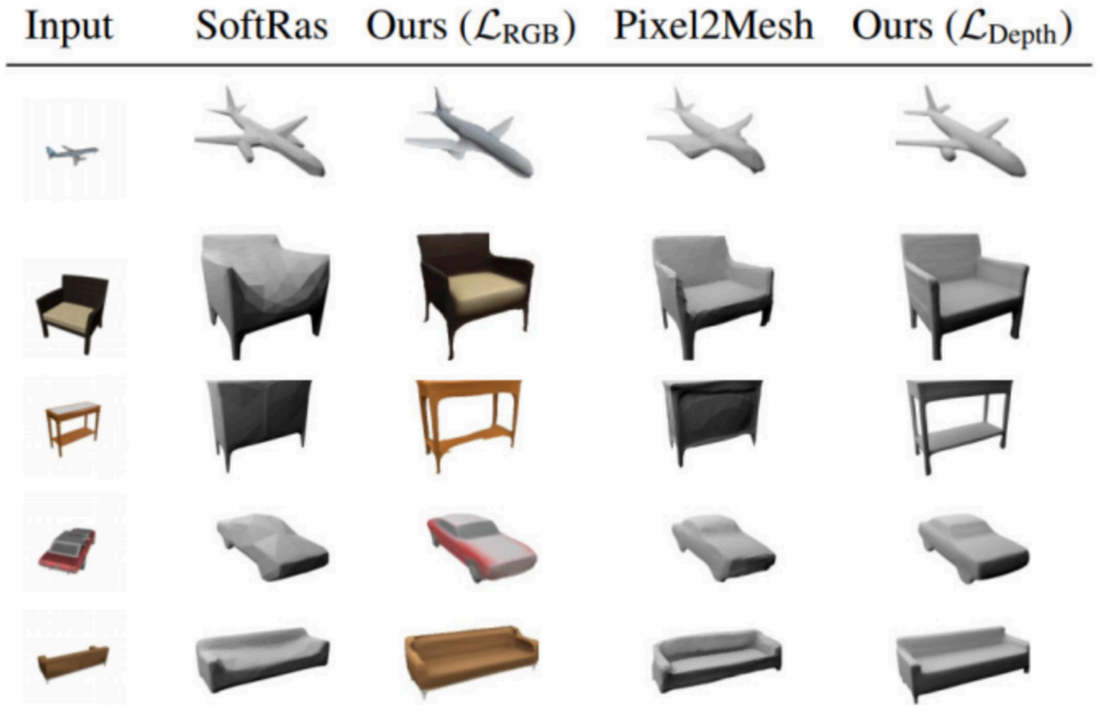
- Human pose/shape models mature
- Rich parametric models (e.g., SMPL, STAR)
- Regression from RGB images only
- Models of pose-dependent deformation and clothing



3D Deep Learning

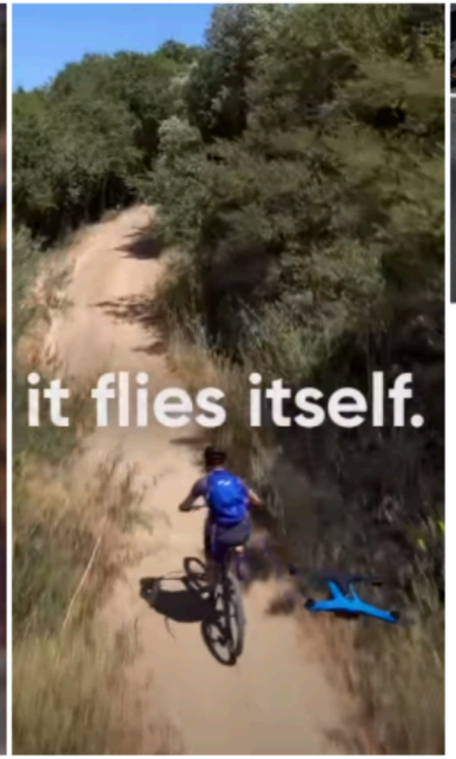
2016-2020

- 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



Applications & Products

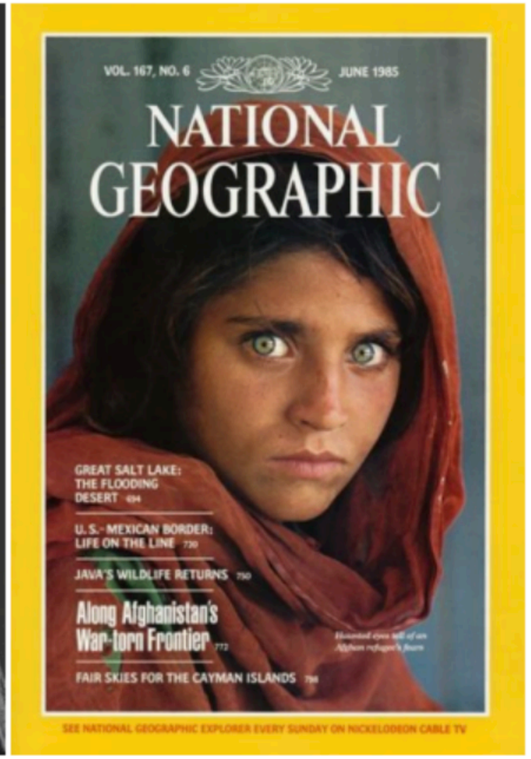
Google Portrait Mode



Skydio 2 Drone

Self-Driving Cars

Microsoft Hololens

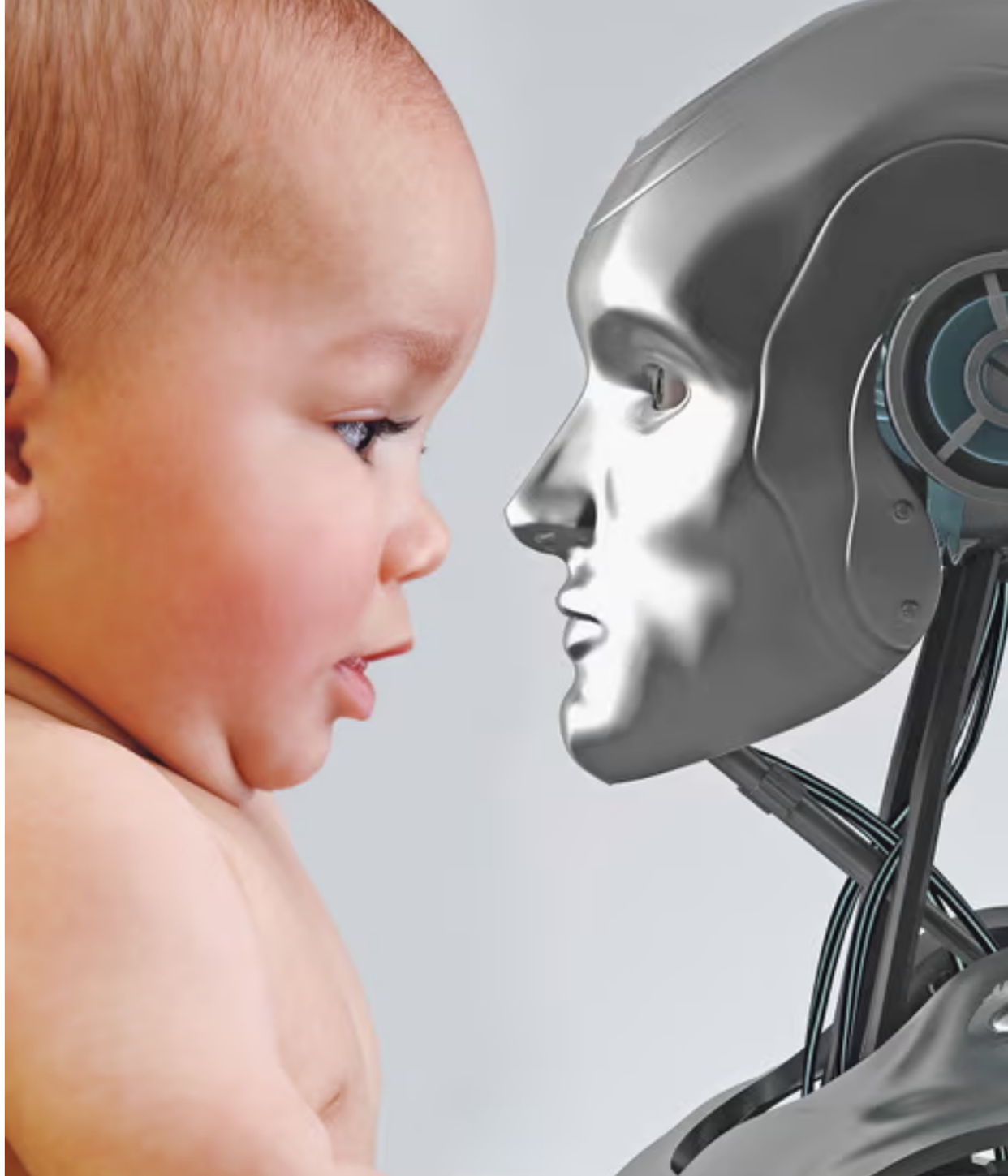


Iris Recognition

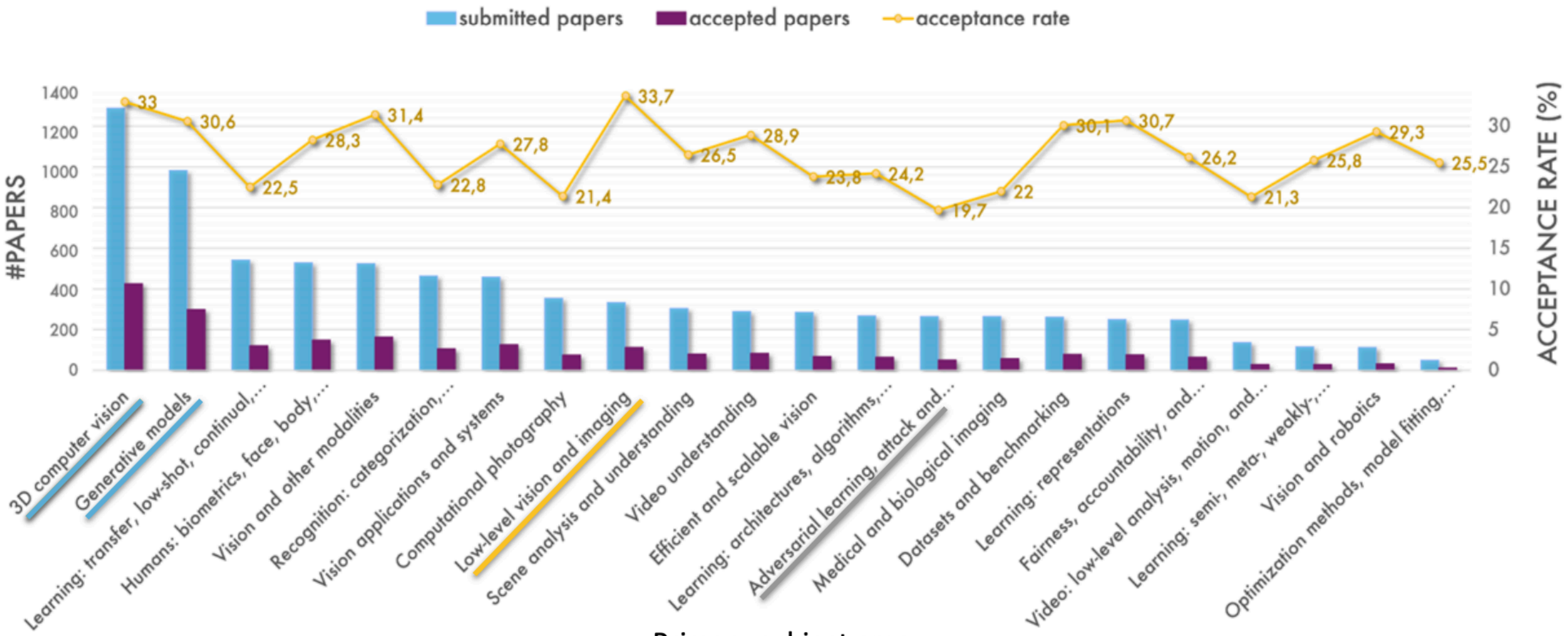
Remaining Challenges

Next

- 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



ECCV2024


submitted papers
accepted papers
acceptance rate
#PAPERS
ACCEPTANCE RATE (%)
1400
1200
1000
800
600
400
200
0
30
25
20
15
10
5
0
33
30,6
22,5
28,3
31,4
22,8
27,8
21,4
33,7
26,5
28,9
23,8
24,2
19,7
22
30,1
30,7
26,2
21,3
25,8
29,3
25,5
3D computer vision
Generative models
Learning: transfer, low-shot, continual,...
Humans: biometrics, face, body,...
Vision and other modalities
Recognition: categorization,...
Vision applications and systems
Computational photography
Low-level vision and imaging
Scene analysis and understanding
Video understanding
Efficient and scalable vision
Learning: architectures, algorithms,...
Adversarial learning, attack and...
Medical and biological imaging
Datasets and benchmarking
Learning: representations
Fairness, accountability, and...
Video: low-level analysis, motion, and...
Learning: semi-, meta-, weakly-,...
Vision and robotics
Optimization methods, model fitting,...
Primary subject area