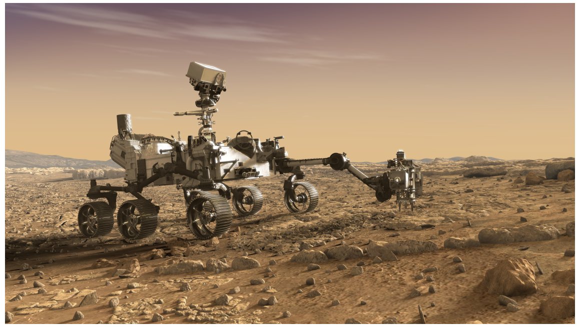
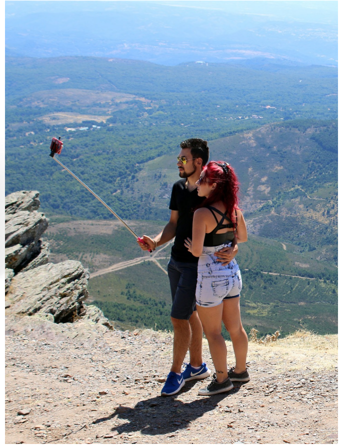
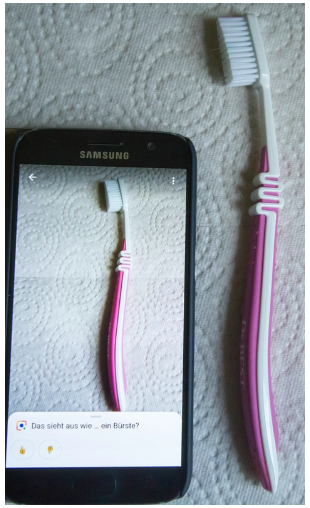
CSC496: Deep learning in computer vision
Prof. Bei Xiao

Lecture 1: Introduction

Welcome to CSC 496

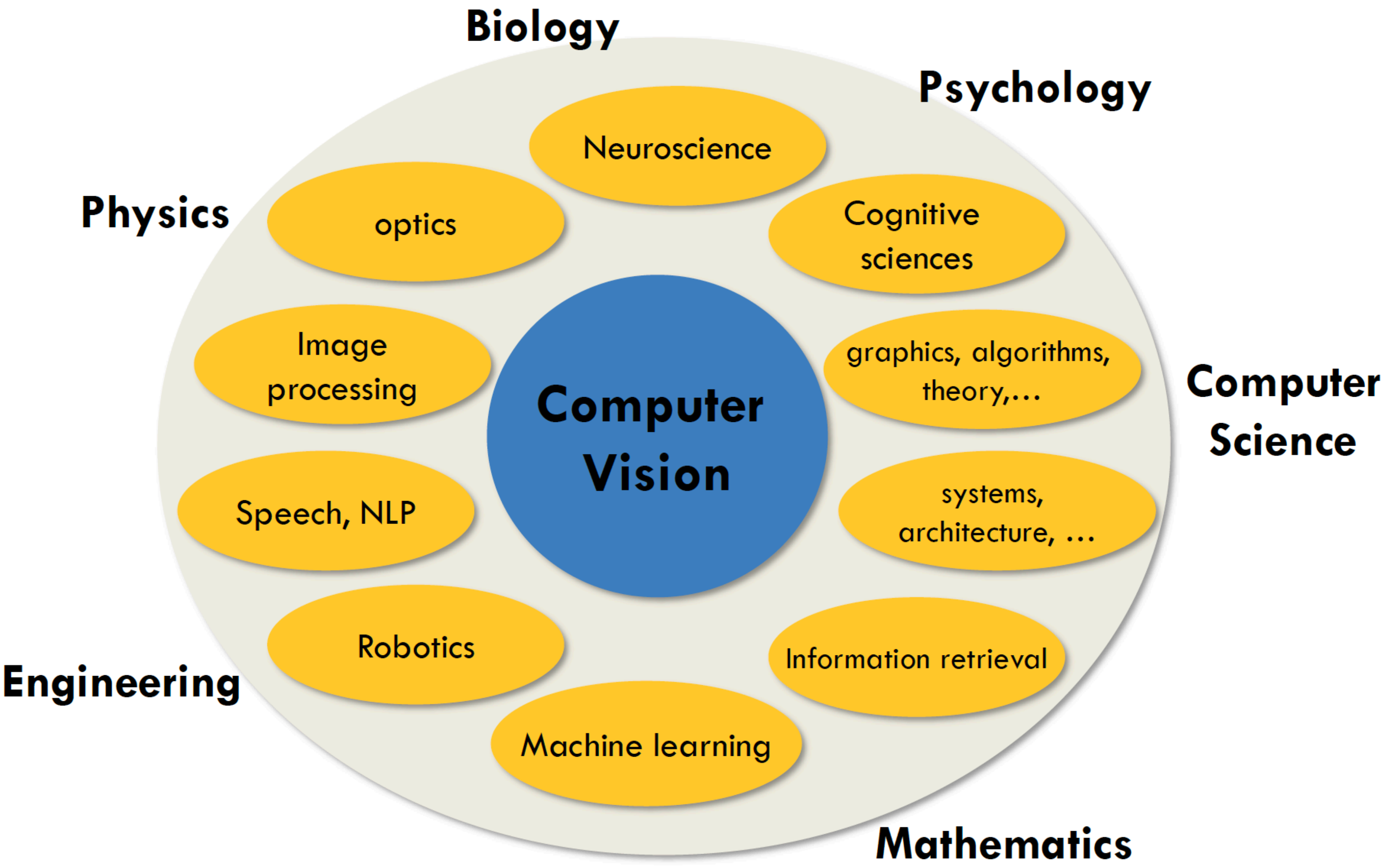






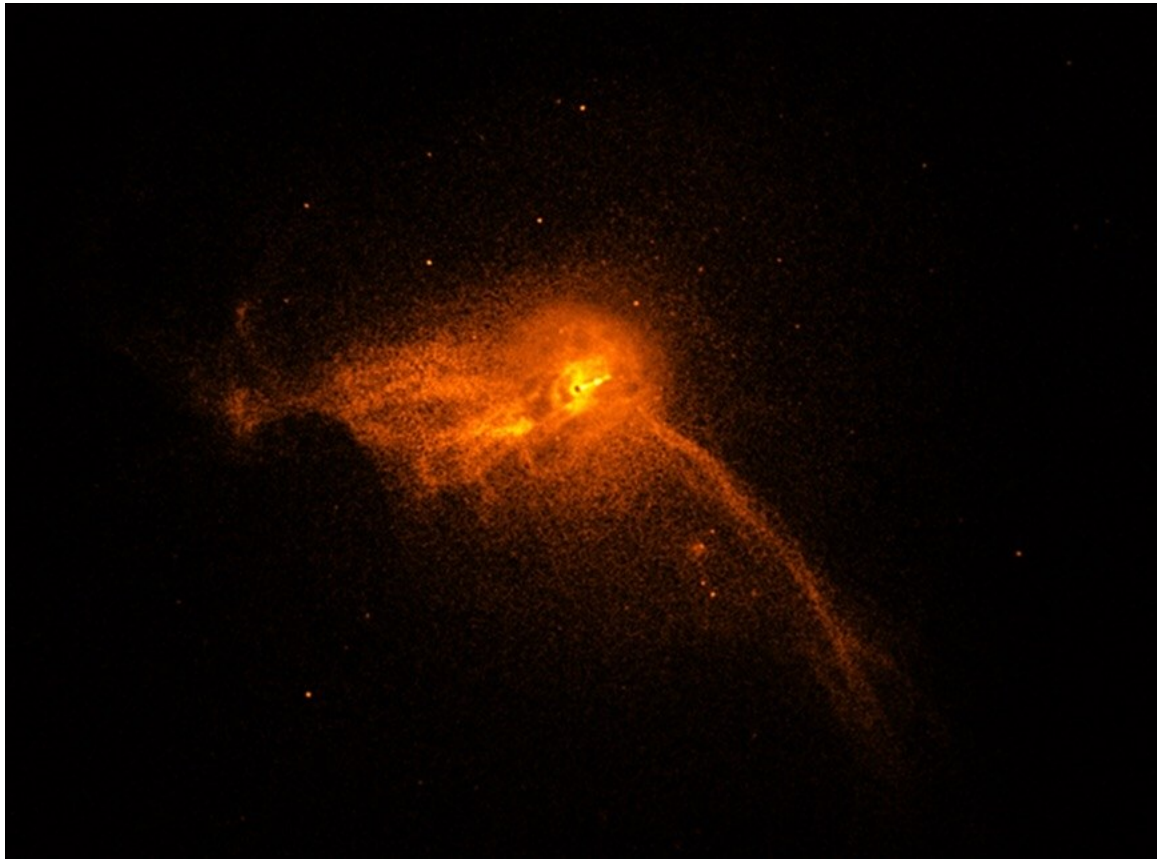
Overview

- What is computer vision?
- History of computer vision
- What is deep learning?
- Instructor/student Introduction
- About this course



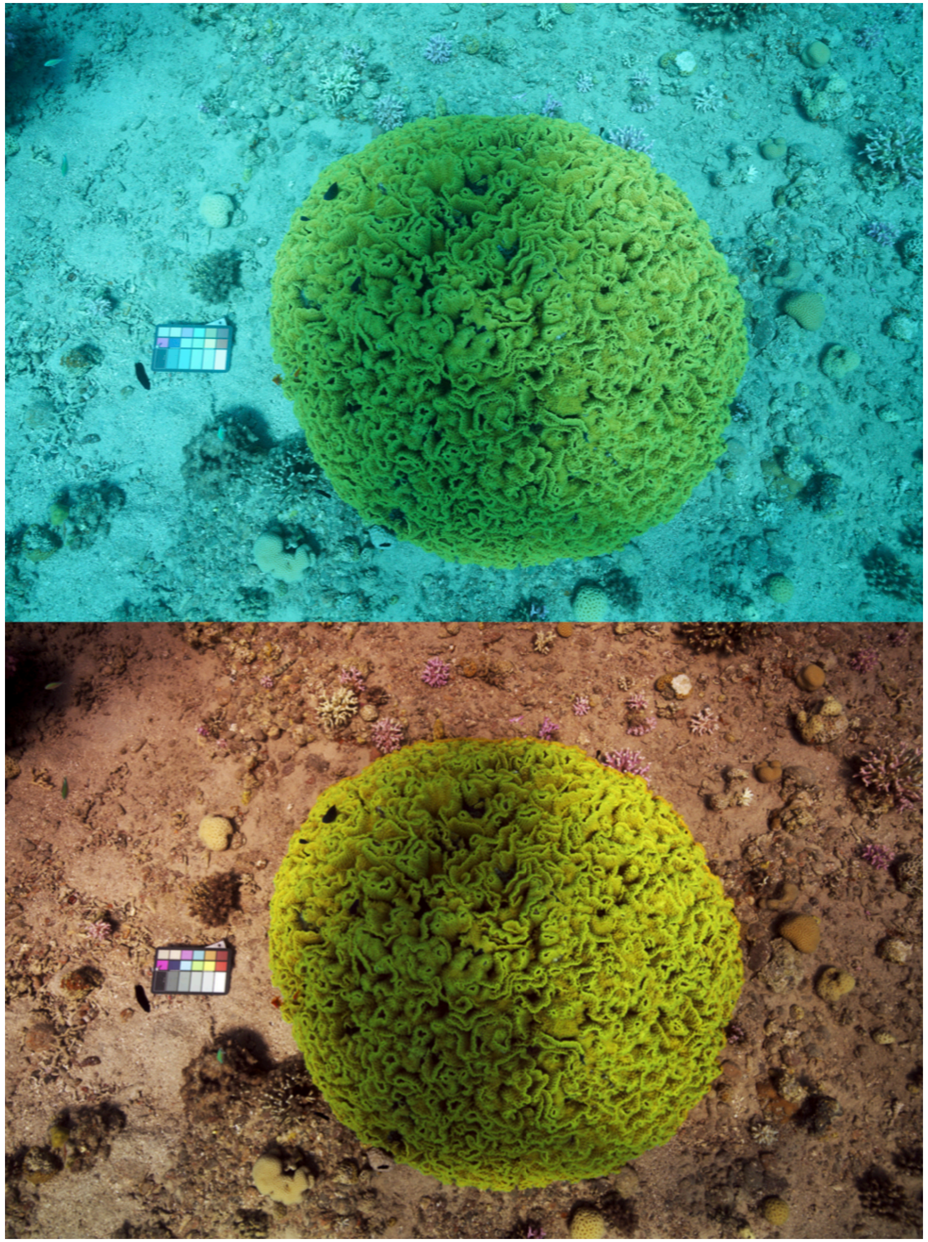
Slide from Stanford cs231n

Astrophysics: Imaging the black hole



https://www.jpl.nasa.gov/edu/news/2019/4/19/how-scientists-captured-the-first-image-of-a-black-hole/

Fun application: remove water from underwater imaging



Akkaynak et al. 2019 CVPR

Medical imaging: classifying cancer

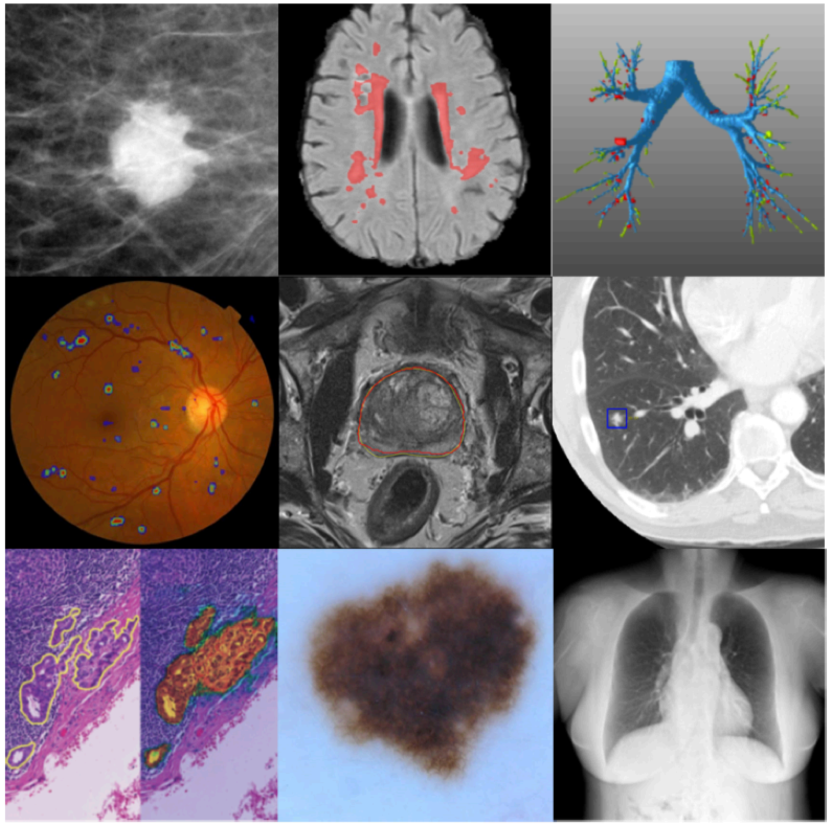
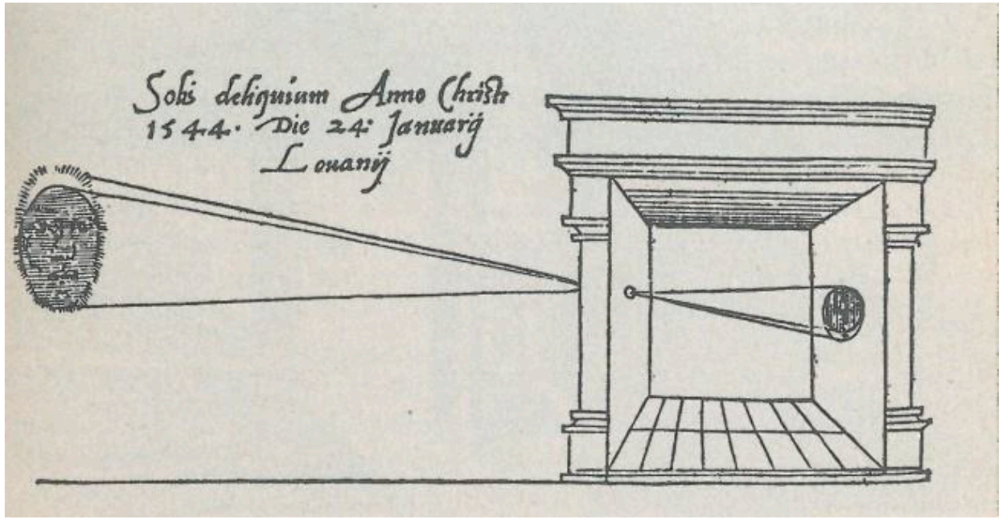


Figure 3: Collage of some medical imaging applications in which deep learning has achieved state-of-the-art results. From top-left to bottom-right: mammographic mass classification (Kooi et al., 2016), segmentation of lesions in the brain (top ranking in BRATS, ISLES and MRBrains challenges, image from Ghafoorian et al. (2016b), leak detection in airway tree segmentation (Charbonnier et al., 2017), diabetic retinopathy classification (Kaggle Diabetic Retinopathy challenge 2015, image from van Grinsven et al. (2016), prostate segmentation (top rank in PROMISE12 challenge), nodule classification (top ranking in LUNA16 challenge), breast cancer metastases detection in lymph nodes (top ranking and human expert performance in CAMELYON16), human expert performance in skin lesion classification (Esteva et al., 2017), and state-of-the-art bone suppression in x-rays, image from Yang et al. (2016c).

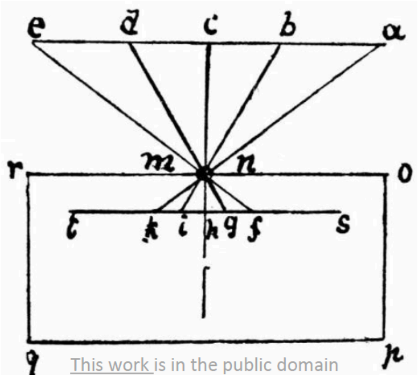
Landmarks in computer vision

Camera Obscura (pinhole camera)

Gemma Frisius, 1545



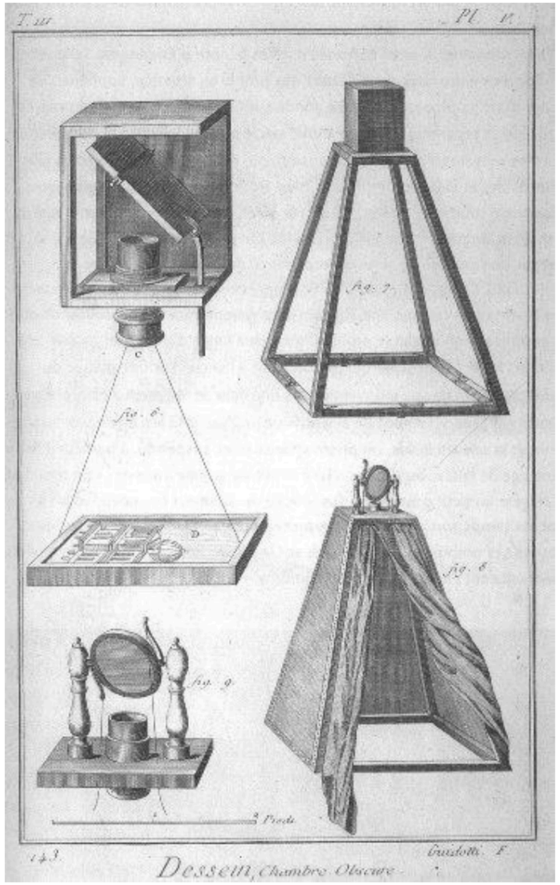
This work is in the public domain



This work is in the public domain

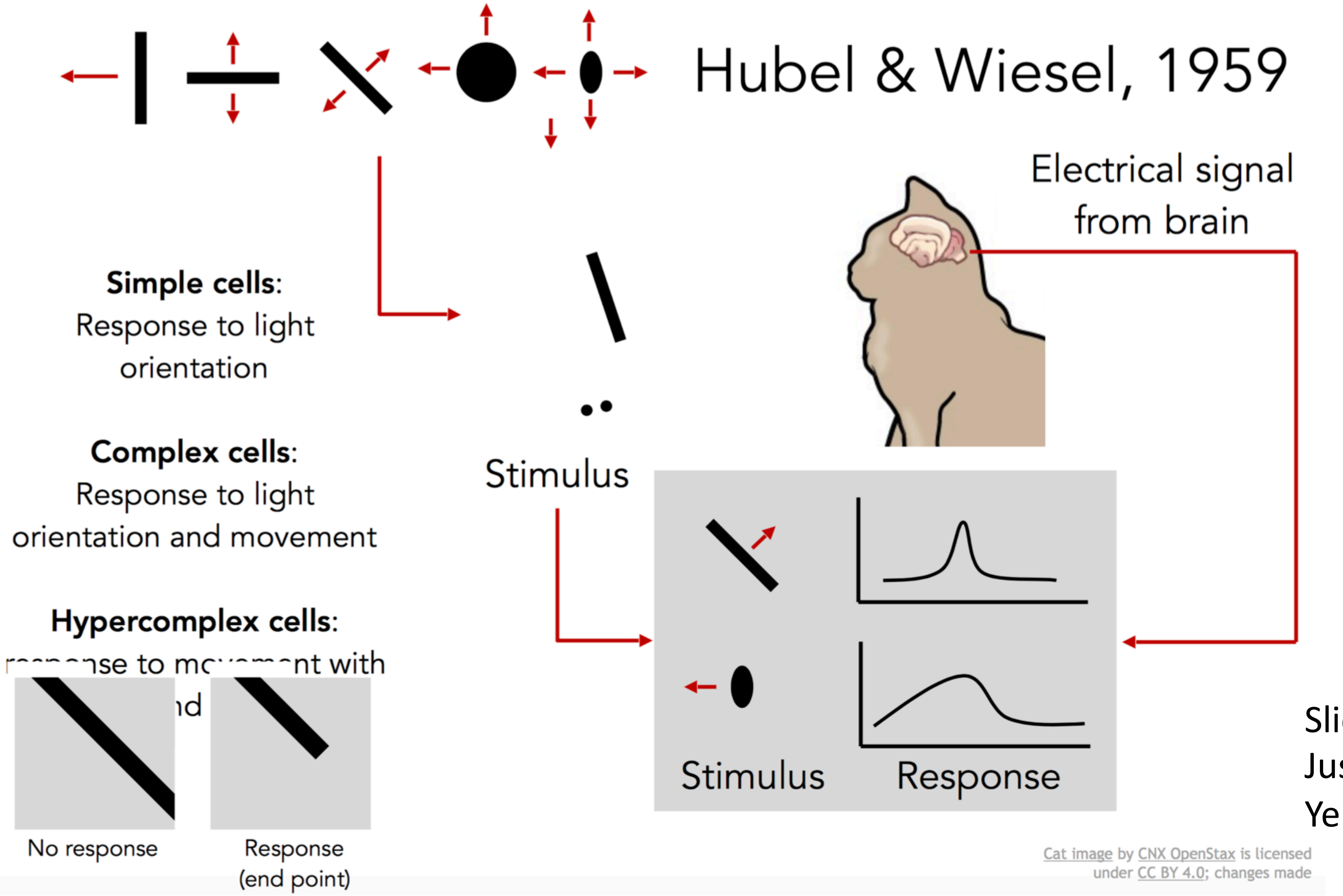
Leonardo da Vinci, 16th Century AD

Encyclopedie, 18th Century



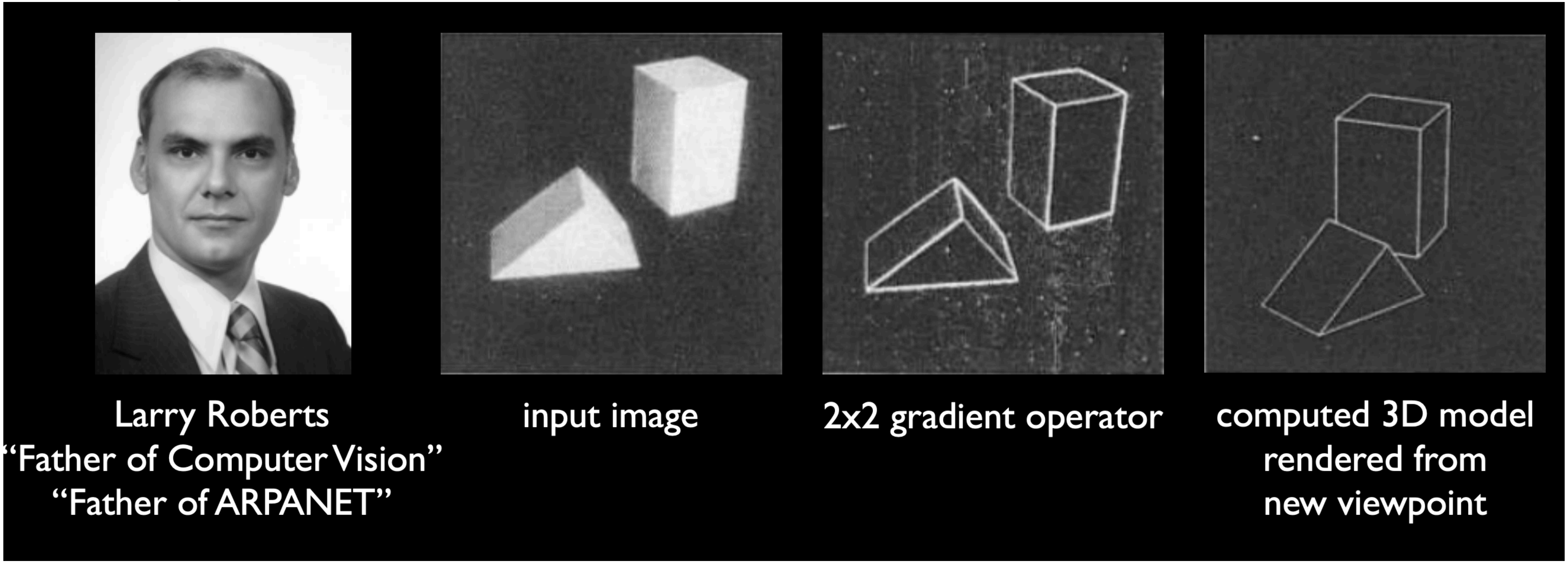
This work is in the public domain

Visual Neuroscience, Nobel Prize 1962



Slide from Fei-Fei Li, Justin Johnson, Serena Yeung, cs231 Stanford

Block world
Larry Roberts, 1963



Slide from Steve Setiz

Block world ++
Culmination: 1970 Copy-demo at MIT

- vision system recovers the structure of a blocks scene, robot plans and builds an exact copy from another set of blocks

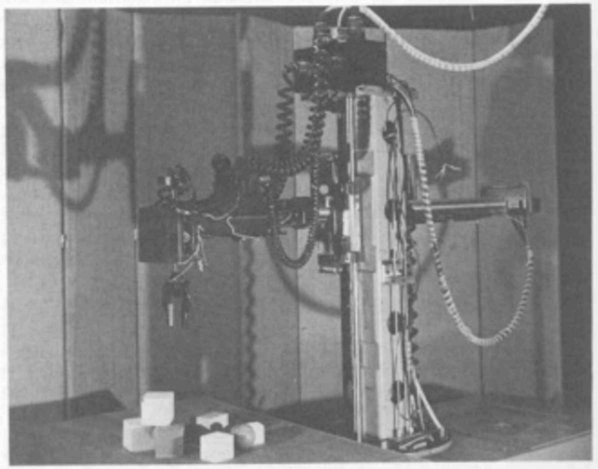


Figure 15-3. The mechanical manipulator used in the "copy-demo," one of the first projects in which visual information was used to plan the motions of an industrial robot. (Photo by Steve Slesinger.)

Winston, Horn, Freuder

Ambitious AI system

- vision + planning + manipulation
- weak link: low-level edge finding not good enough for task
  - led to more attention on low-level vision

Slide from Steve Setiz

Summer project: 1966

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

<u>THE SUMMER VISION PROJECT</u>

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".

Shape from shading: 1970

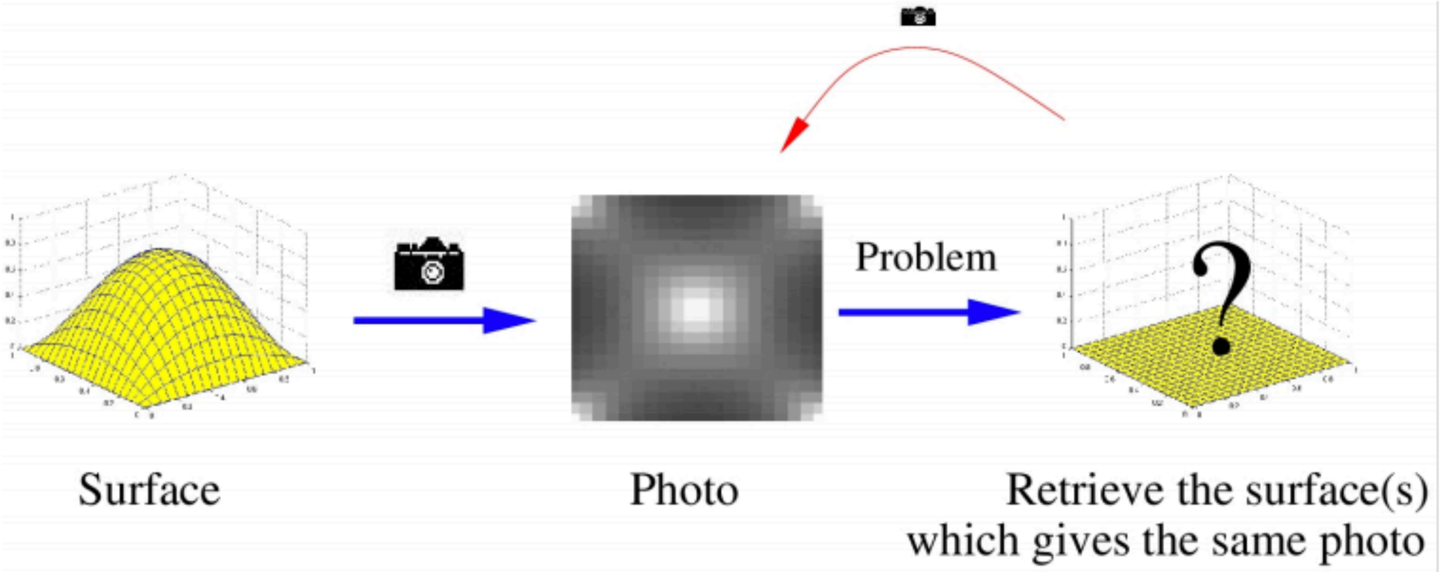


FIGURE 1. The "Shape-from-Shading" problem.

Shape from shading: 1970

1970: Shape-from-shading [Horn, MIT Ph.D., (AITR 232)]

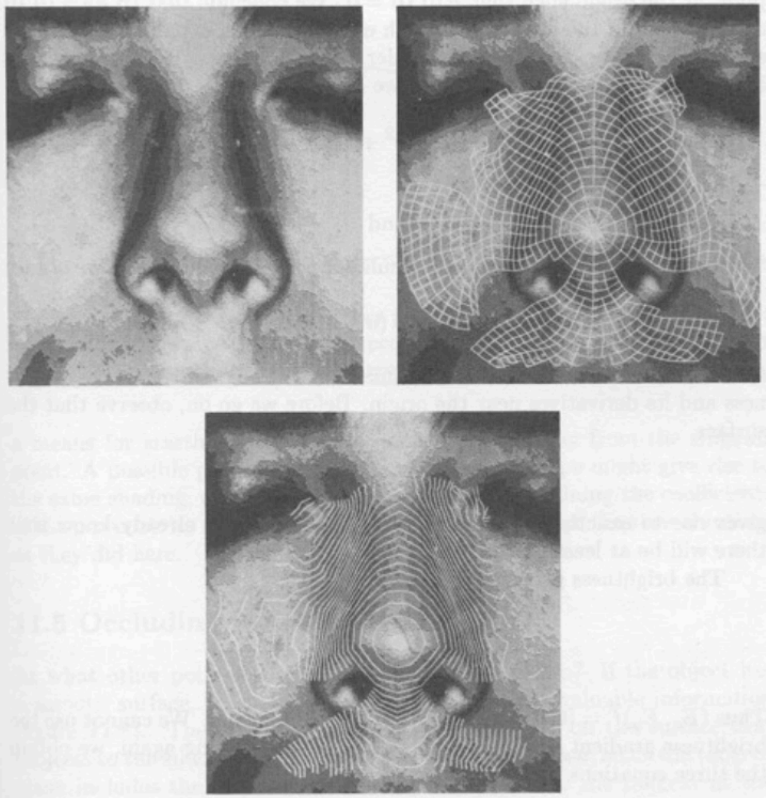


Figure 11-7. The shape-from-shading method is applied here to the recovery of the shape of a nose. The first picture shows the (crudely quantized) gray-level image available to the program. The second picture shows the base characteristics superimposed, while the third shows a contour map computed from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

Also known as *photoclinometry*

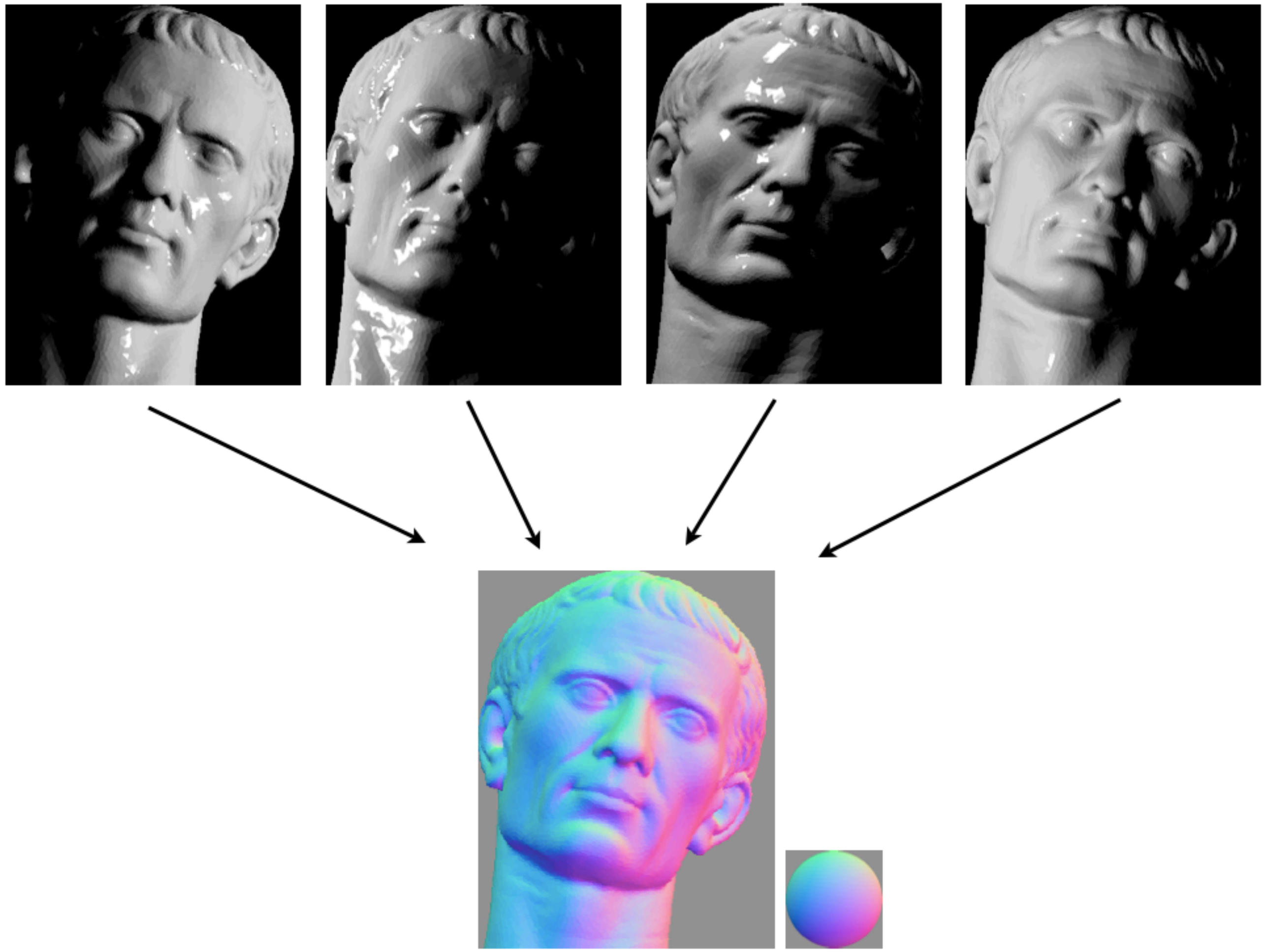
- measure planetary geometry
- Origins from mid 1960's
  - T. Rindfleisch, Photometric Method for Lunar Topography, Photometric Engineering, 1966.
- SfS vs. photoclinometry discussion in Horn 1989

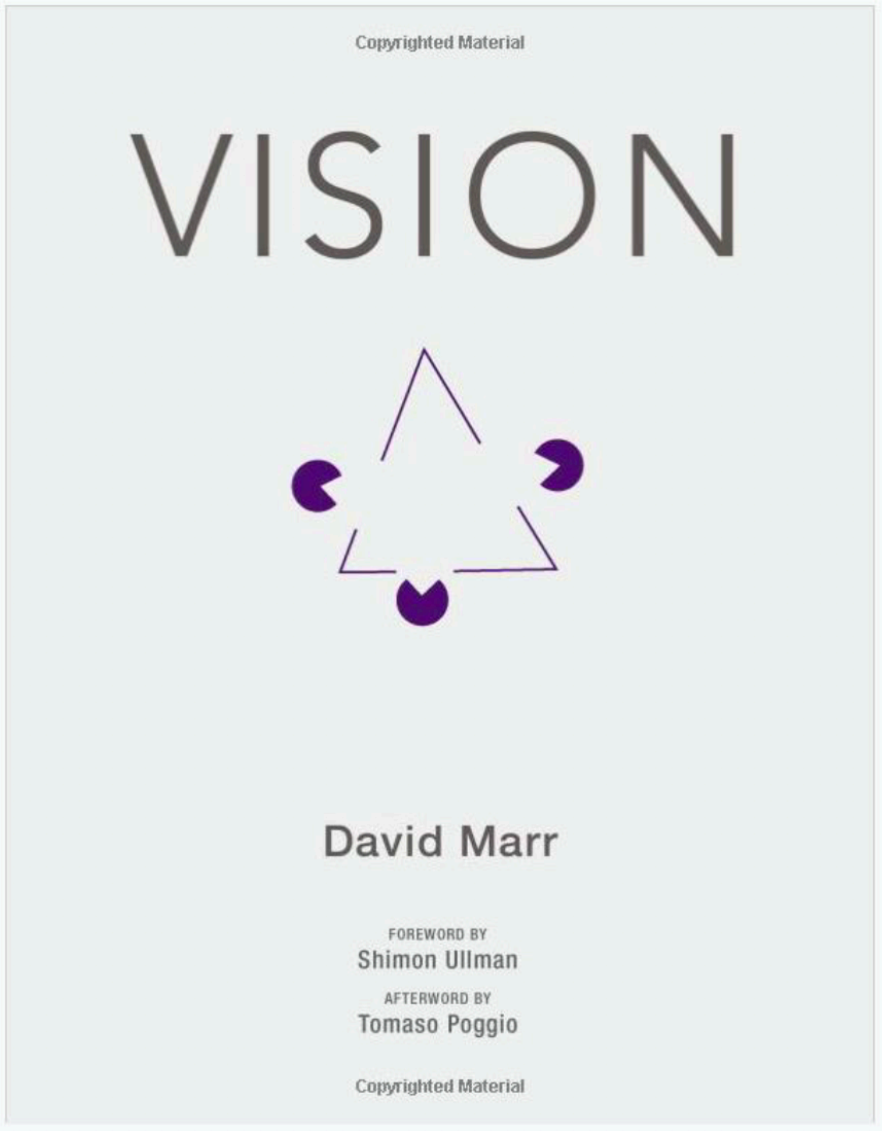
Ernst Mach [1865] formulated image irradiance equation

- concluded shape inference was impossible because of instabilities

Slide from Steve Setiz

Photometric stereo

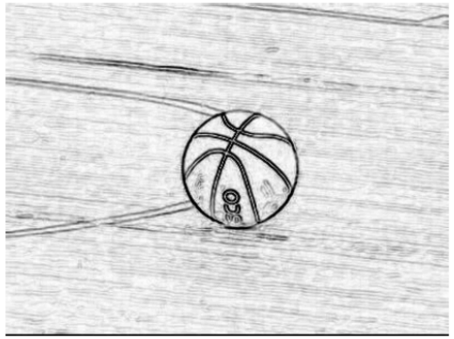




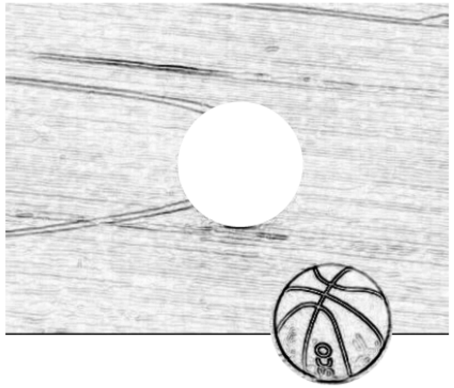
David Marr, 1970s

Input image

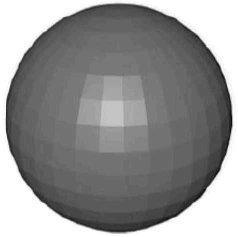
Edge image



2 ½-D sketch



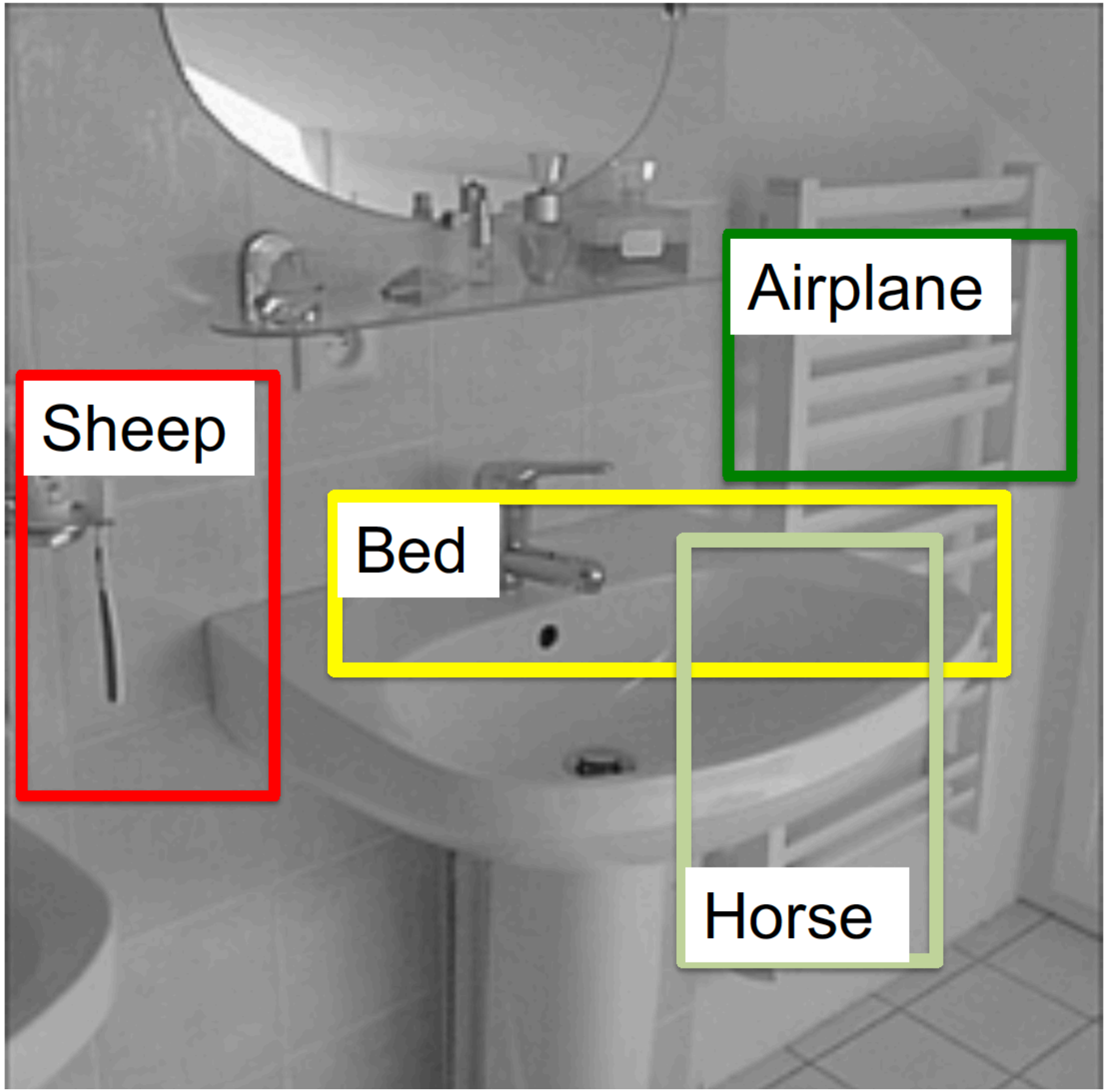
3-D model



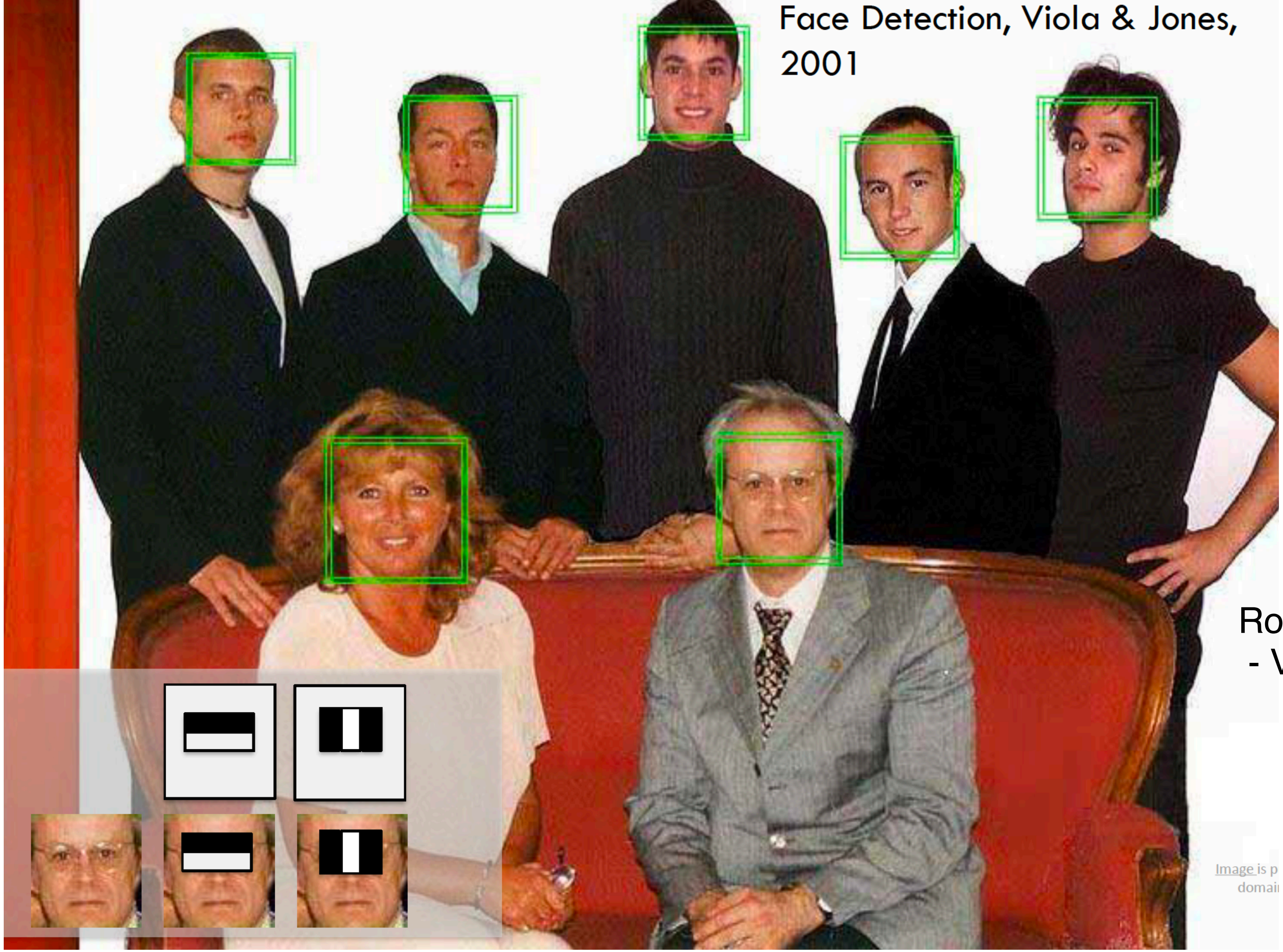
Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface and volumetric primitives

Stages of Visual Representation, David Marr, 1970s

30 years ago...



But 15 years ago…

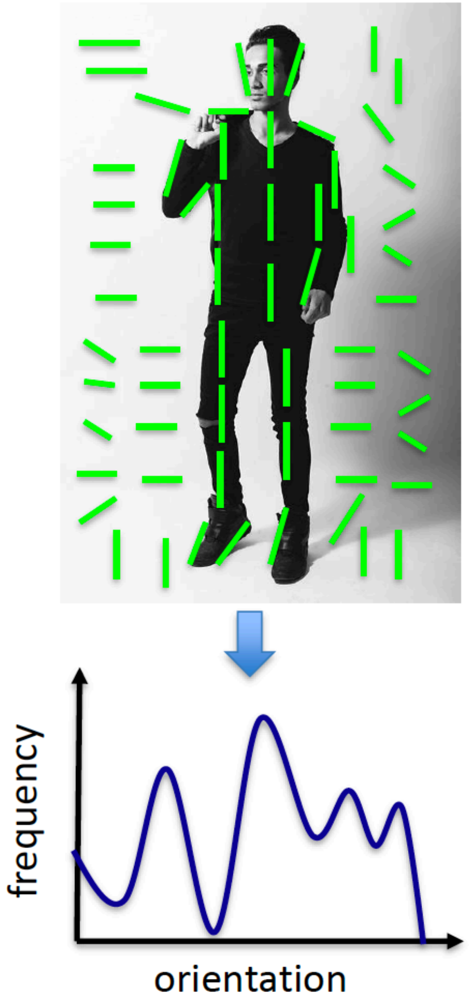


Robust Real-time Object Detection
- Viola * Jones, 2001

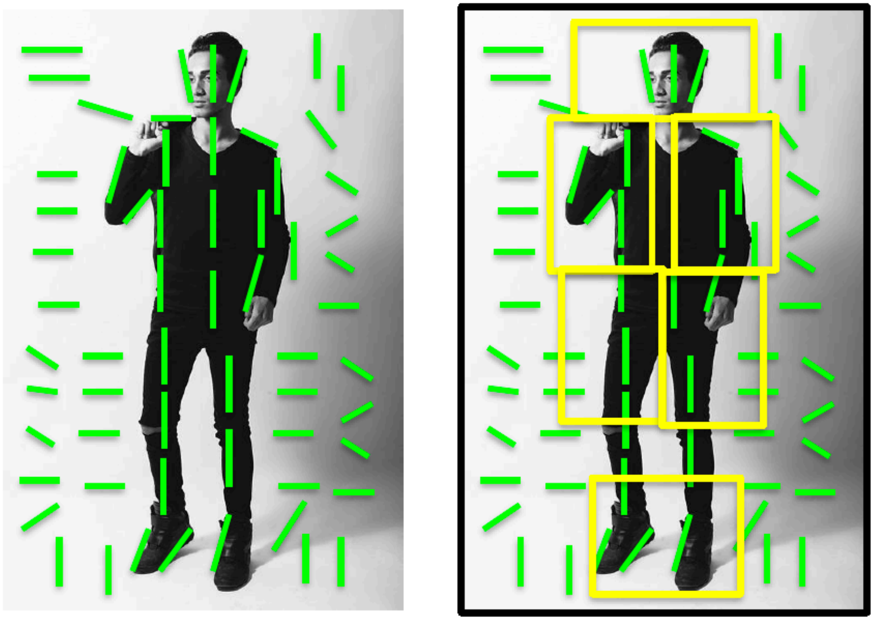
Image is public domain

Image is public domain

“SIFT” & Object Recognition, David Lowe, 1999



Histogram of Gradients (HoG)
Dalal & Triggs, 2005



Deformable Part Model
Felzenswalb, McAllester, Ramanan, 2009

PASCAL Visual Object Challenge (20 object categories)

[Everingham et al. 2006-2012]

Image is CC0 1.0 public domain

Image is CC0 1.0 public domain

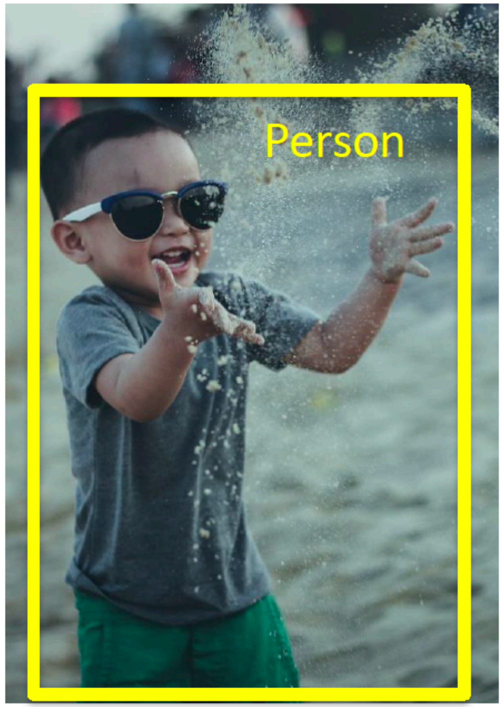
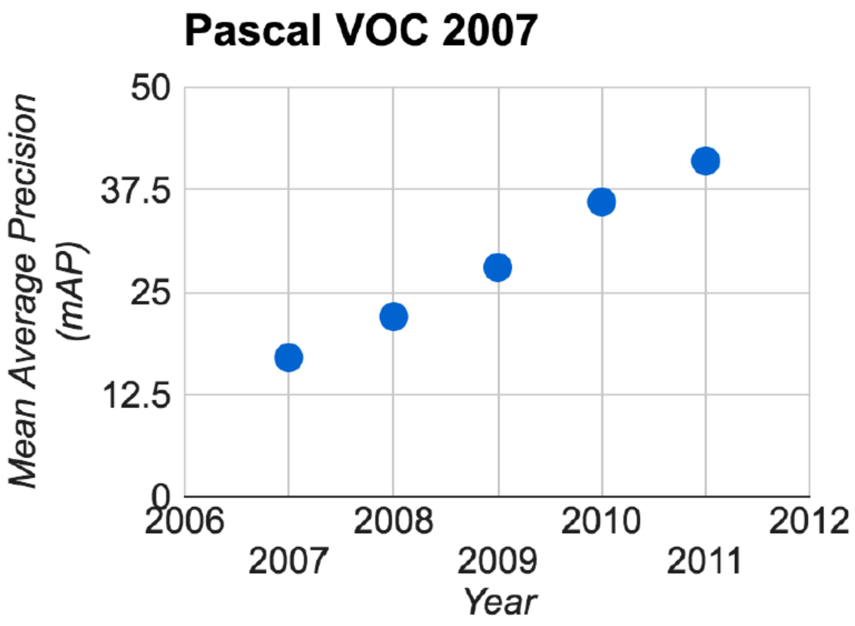


Image is CC0 1.0 public domain



IMAGENET

www.image-net.org

22K categories and **15M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

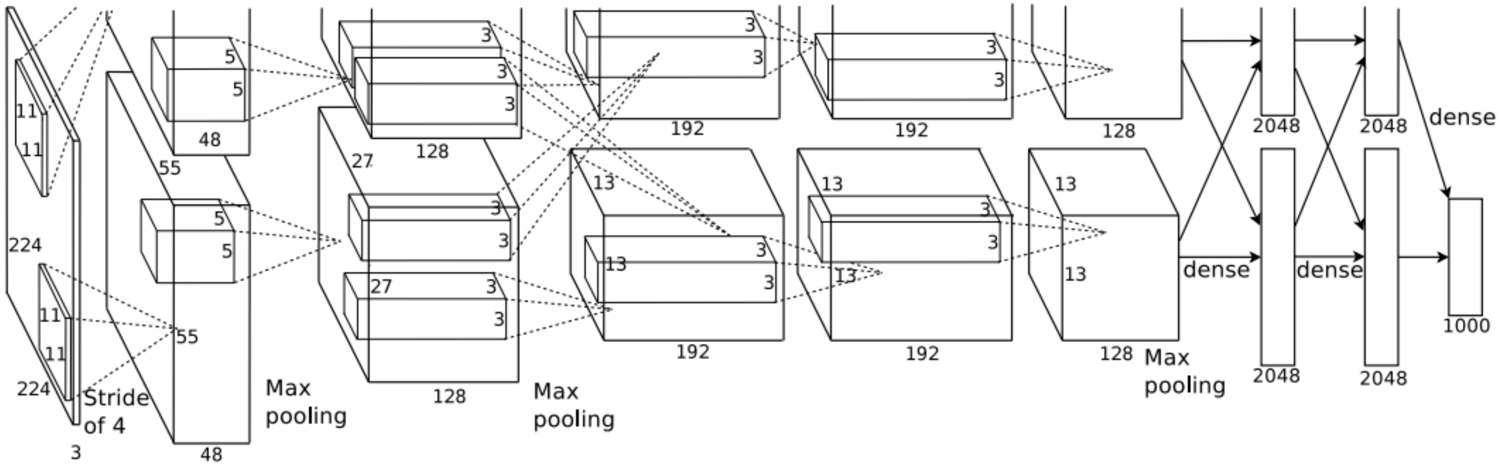
1,431,167 images

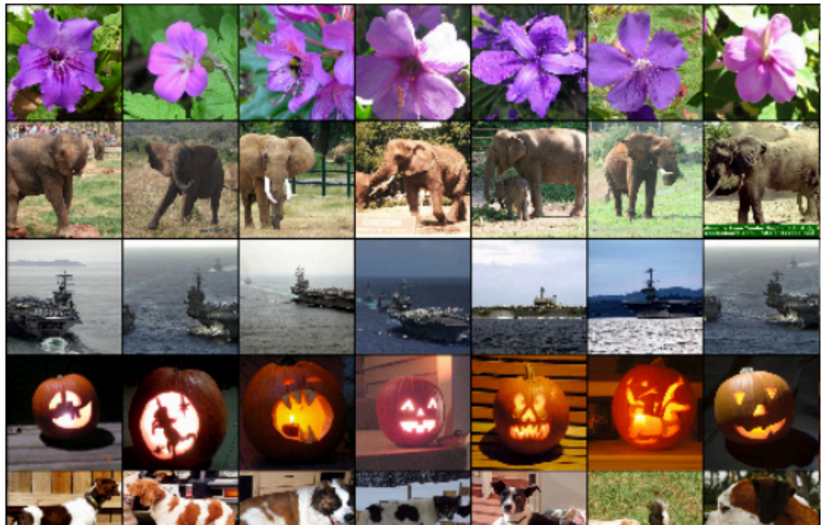
Output:	Output:
Scale	Scale
T-shirt	T-shirt
<u>Steel drum</u>	Giant panda
Drumstick	Drumstick
Mud turtle	Mud turtle

Russakovsky et al. IJCV 2015

https://arxiv.org/pdf/1409.0575.pdf

AlexNet 2012





Trained with ImageNet, Winner of the 2012 IMAGENET CHALLENGE

85% accuracy top-5 accuracy

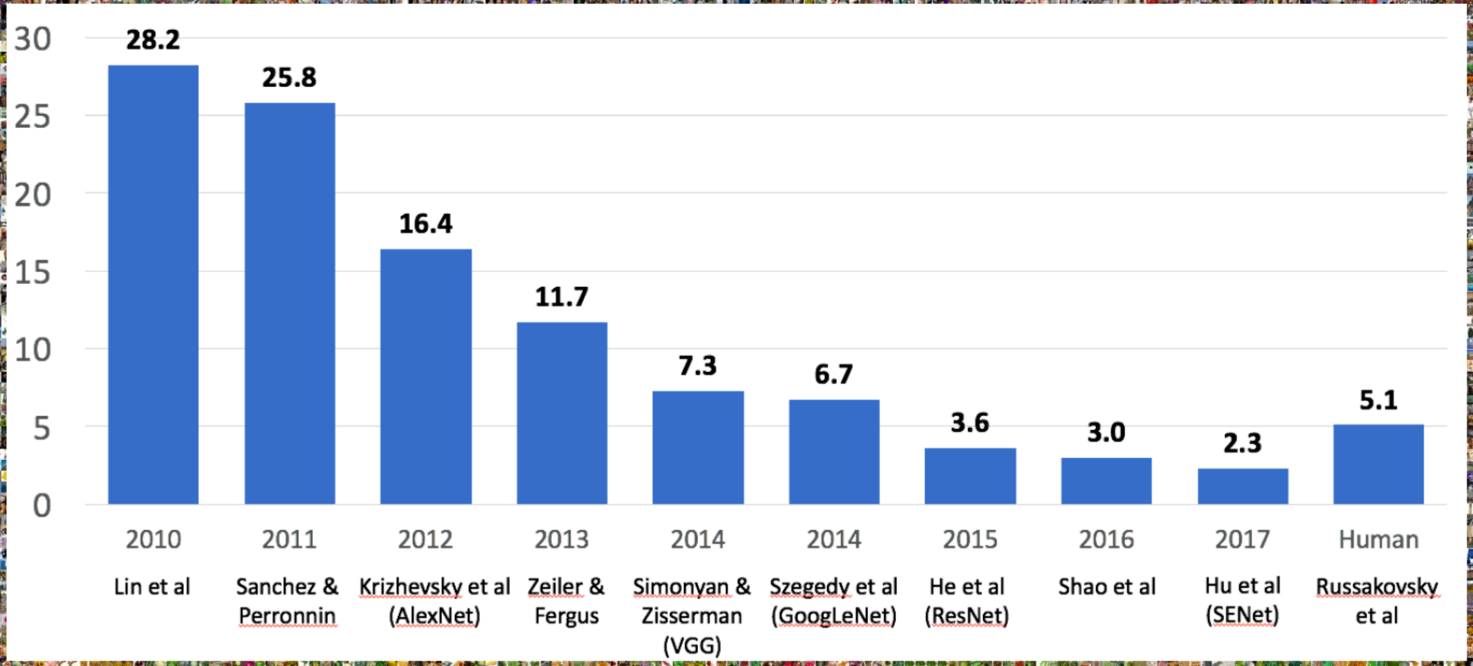
Trained on multiple GPUs

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images



Russakovsky et al. IJCV 2015

The rise of deep learning

Motherboard

'Deep Voice' Software Can Clone Anyone's Voice With Just 3.7 Seconds of Audio

Using snippets of voices, Baidu's 'Deep Voice' can generate new speech, accents, and tones.

By Samantha Cole

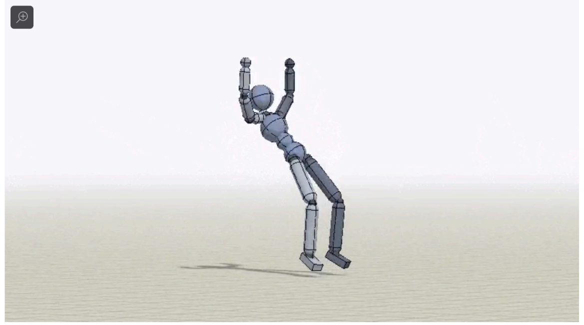
Mar 7 2018, 1:00pm Share Tweet

IMAGE: SHUTTERSTOCK

After Millions of Trials, These Simulated Humans Learned to Do Perfect Backflips and Cartwheels

George Dvorsky
4/11/18 11:55am • Filed to: AI

29.7K 40 13



Simulated humanoid pulling off a perfect backflip, which it did after a month of simulated training.

How an A.I. 'Cat-and-Mouse Game' Generates Believable Fake Photos

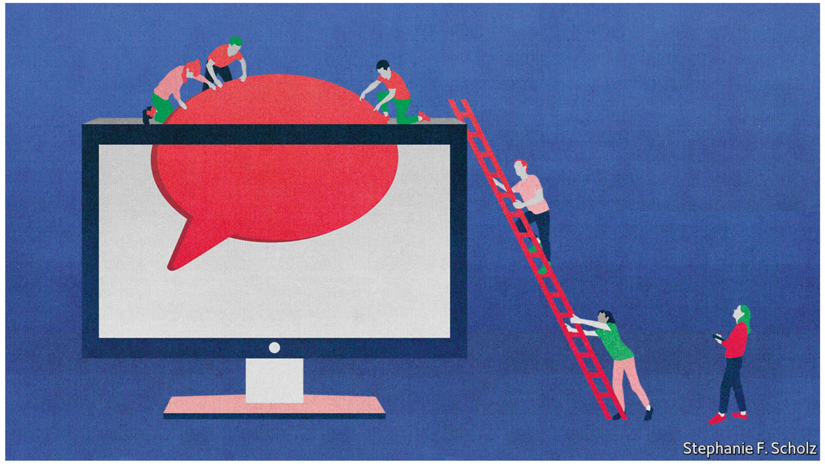
By CADE METZ and KEITH COLLINS JAN. 2, 2018

To create the final image in this set, the system generated 10 million revisions over 18 days.

AI in society

For artificial intelligence to thrive, it must explain itself

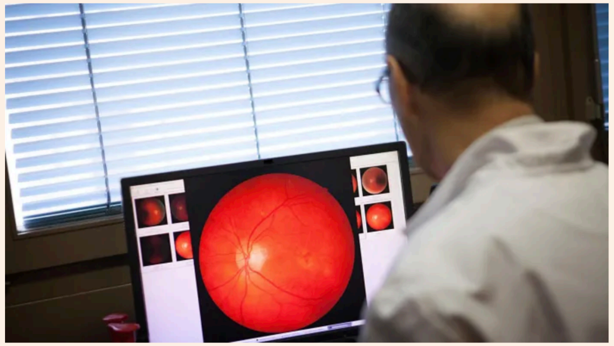
If it cannot, who will trust it?



Stephanie F. Scholz

Google's AI beats doctors at spotting eye disease in scans

Clinical trials planned after early tests of DeepMind software



© Alamy

Aliya Ram in London AUGUST 13, 2018

What is deep learning?

Artificial Intelligence

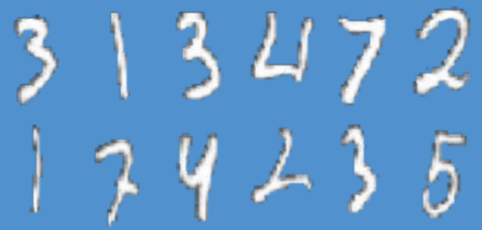
Any technique that enables computers to mimic human behavior

Machine Learning

Ability to learn without explicitly being programmed

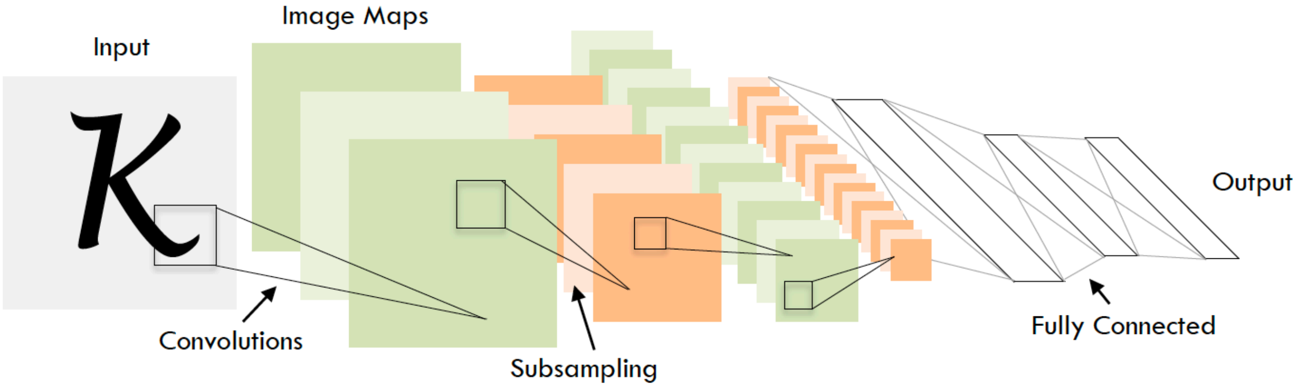
Deep Learning

Extract patterns from data using neural networks



1998

LeCun et al.



# of transistors

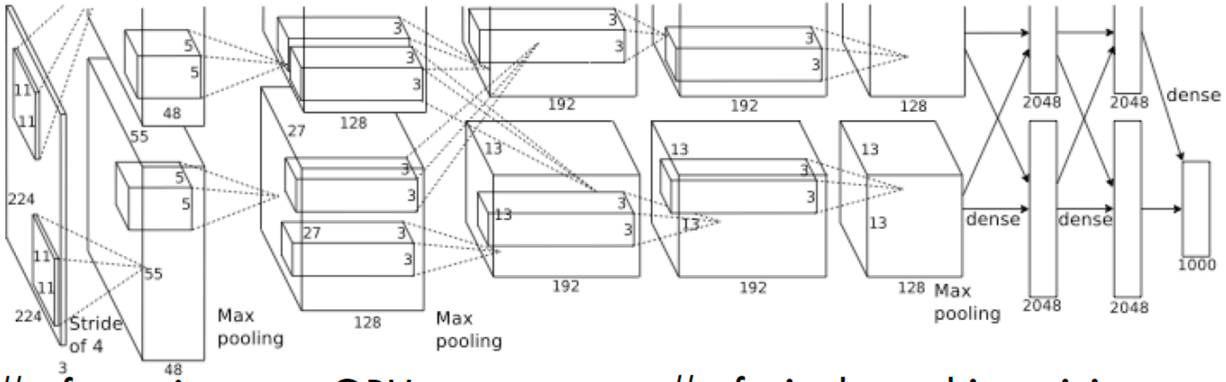
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



# of transistors

10^9

GPUs

# of pixels used in training

10^{14} IMAGENET

Why now?

1952	Stochastic Gradient Descent
1958	Perceptron • Learnable Weights
⋮	
1986	Backpropagation • Multi-Layer Perceptron
1995	Deep Convolutional NN • Digit Recognition
⋮	

Neural Networks date back decades, so why the resurgence?

1. Big Data

- Larger Datasets
- Easier Collection & Storage

IMAGENET

WIKIPEDIA
The Free Encyclopedia

2. Hardware

- Graphics Processing Units (GPUs)
- Massively Parallelizable

3. Software

- Improved Techniques
- New Models
- Toolboxes

TensorFlow

Ingredients for deep learning

Algorithms

Data

Computation (hardware, architecture)

Skeptical view of current status of AI revolution (assigned reading)

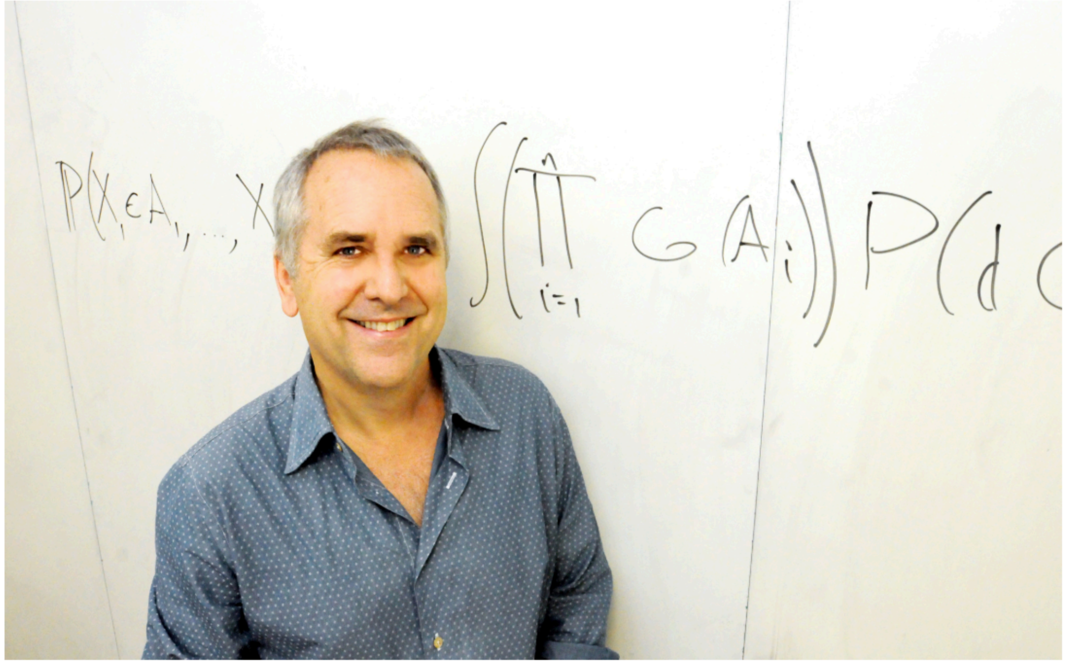


Photo credit: Peg Skorpinski

Artificial Intelligence — The Revolution Hasn't Happened Yet

Discussion:

What is the difference between "AI" and "Machine learning"?

According to the author, what aspects of the "AI" are most successful so far?

What are some of the skepticism the author raised for the current "AI" revolution?

What algorithm is considered the "core" of AI revolution?

Who am I?

- Bei Xiao, Assistant Professor of Computer Science, Vision Reseacher (bxiao@american.edu)

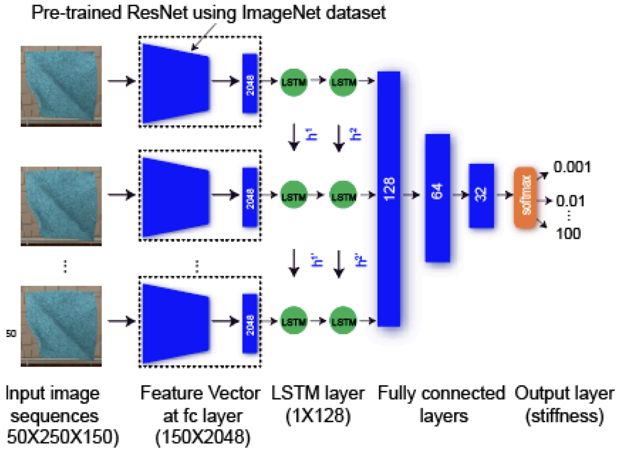
- Research interests:
  - Human vision
  - Computer vision
  - Computer Graphics

  - Specifically, I am interested in
    - Human and machine estimation materials of objects
    - Tactile visual integration
    - Perception in Virtual Environment (VR/AR)
    - Using g human intuition to guide machine algorithm

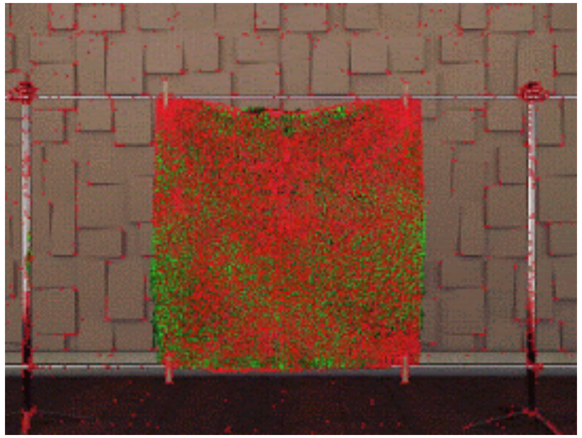
Current research in Xiao Lab

Interactive Material Perception in VR/AR

Deep learning cloth properties

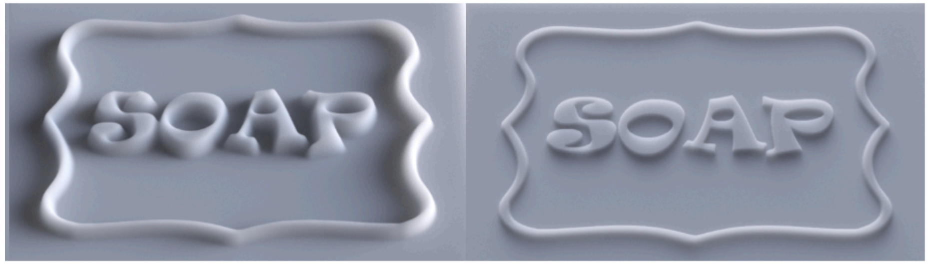


Motion and perception of deformable objects

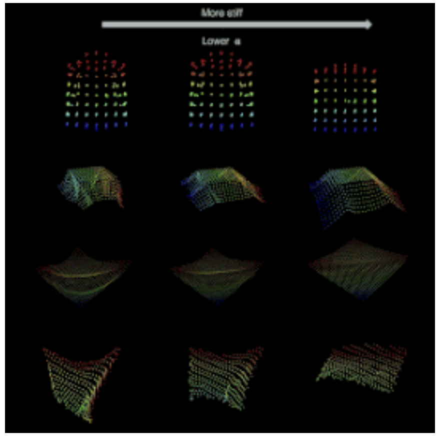


Tactile and vision integration

Perception of translucent materials



Simulate deformable objects



Pre-reqs

Some math: especially probability, college, calculus and linear algebra

Proficient with at least one high-level programming languages

Some basic backgrounds in image processing/signal processing

This course is still about computer vision with an overview of on deep learning

Part 1 (August 26-September 23rd): **Overview of Computer Visio**n: Image formation, Camera, Light and color, Signal processing, Statistical modeling of images, Basic math tutorials, Numpy tutorials.

Part 2: (September 30-October 17th): **Basic machine learning**. Linear classification. Loss function. Tensorflow tutorials.

Part 3 (October 21st- December 9): **Deep learning architectures and applications in CV.** Neural networks, multilayer perceptron, backpropagation, CNNs, RNNs, GANs, representational learning, GPU computing and HPC.

Pre-course survey: diverse backgrounds!

- Good math: 75% have taken probability and statistics. 62% has taken calculus, and 75% have taken linear algebra.
- Mixed-python programming: 12.5% have never programmed in Python. 38% has done some Python after 280.
- Numpy: 37.5% are beginner level.
- Lack of computer vision: 50% have never done image convolution. 25% have never done image processing.

- Mixed machine learning: 50% has taken machine learning.

How do I respond to this?

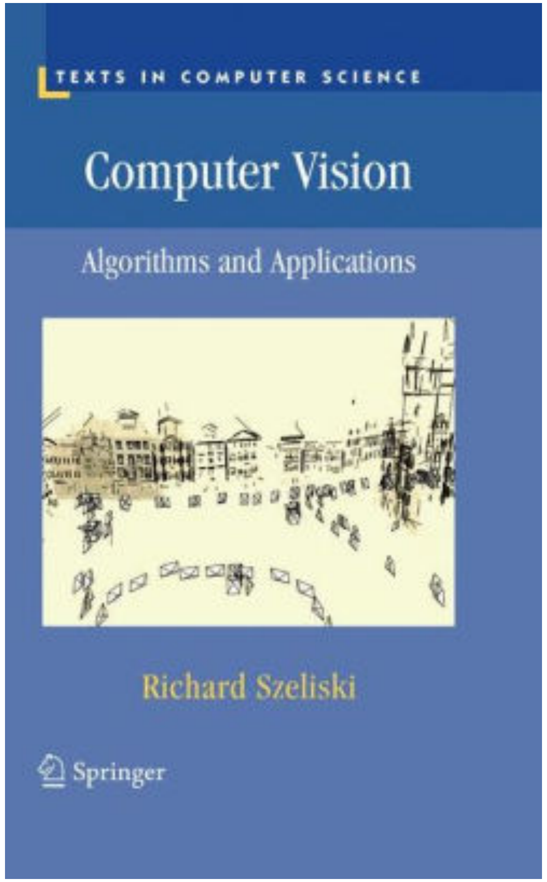
- A review of basic computer vision, with a focus on an signal processing, multiscale pyramids, image formation, statistical modeling of images.
- A tutorial and homework on basic image processing with Numpy.
- A tutorial of basic machine learning methods and algorithmic foundation of neural networks.
- **Expectation:** if you have taken Computer Vision (CSC476), You might have heard some of the content we will cover in the first 3 weeks. You can help your follow classmates by answering their questions or discuss the topics.

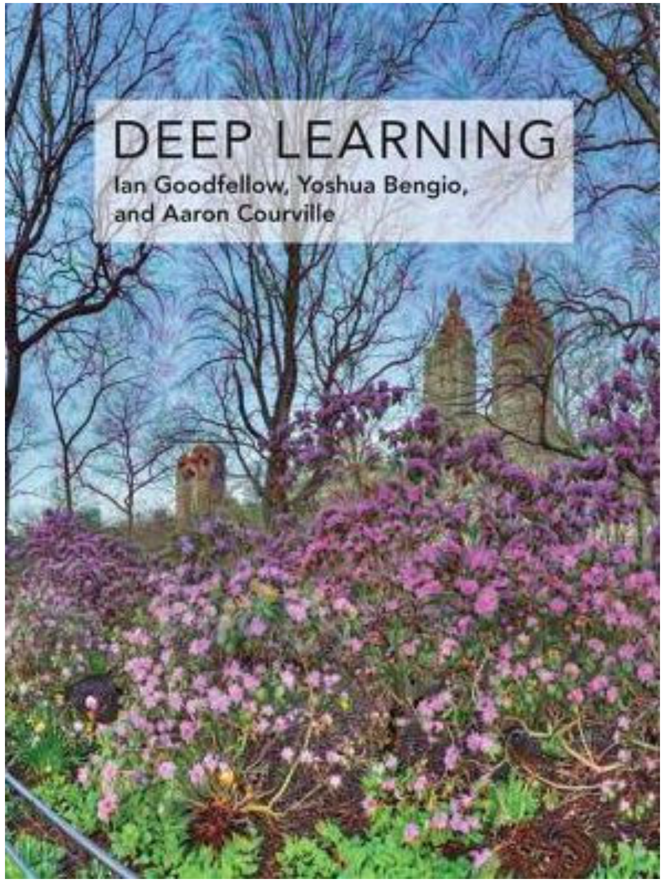
How do you get started?

- Do the tutorials on Python and Numpy in the first two weeks
- http://cs231n.github.io/python-numpy-tutorial/
- https://sites.engineering.ucsb.edu/~shell/che210d/numpy.pdf

- Read Computer Vision (Chapter 3) as much as you can.

Textbooks (both have online versions)





Course communication

- Updated slides/demo code/reading/ will be posted on course GitHub page:
- https://github.com/fruittree/CSC496_2019F
- Homework assignment and submission, announcements, grading will ONLY be happen in blackboard.
- Restrict your questions of homework in office hours. If you can't make to office hour, schedule with me with a 24 hour notice.

What is office hour used for?

Office hour is for?

1. Ask specific questions about contents covered in lectures or any course related content.
2. High-level discussion of projects and algorithms.
3. Ask well-formed questions about topic in computer vision, deep learning, and software/hardware tools.
4. Express concerns, ask for advice in how to learn better, and give feedback on the course.

Office hour is NOT for:

1. Go through lecture slides if you missed the class.
2. Debugging homework solutions.
3. Installing softwares (you can still ask specific questions)

Grading

- 65% homework projects (5-6 projects)
- 10% mid-term (in-class) closed book exam
- 15% Final project
- 5% in-class quiz
- 5% attendances. Missing one class will result in 2% reduction of the total grade (reasons accepted, sickness, pre-registered sports, and religious holidays).

Late Policy

- We **do not accept late submissions for this course**. Homework is typically due 11:59pm. The submission deadline has a 24-hour soft cut-off; after midnight, submissions are penalized 5% per hour late round up to the next hour. So, you turned it in 2:59am. You will get 3 hours penalty with a 15% point reduction. There is no negotiation about this.

- Homework CANNOT be submitted via Email. Only homework submitted to Blackboard will be graded. All the technical issues will have to be resolved earlier than the deadline (e.g. if data is too big you are allowed to submit a link of dropbox to download). It is your responsibility to make sure all of your code compile and run.

Collaboration Policy

- No sharing of code/solutions among students (unless team work). However discussing general strategies is fine.
- No copying from the Internet (GitHub, etc) unless specific by instructor. Everything you submitted should be your work.
- No “detailed” walk through of homework project by a tutor, another student, etc. You can come to office hour to discuss with me. Only high-level discussion is allowed.
- No posting homework to public websites before one hour after the deadline.
- Turning in something non-perfect/incomplete is better than violate the honor code.

Softwares

On your laptop:

Main tools,: Python 3, tensorflow with keras.

Pip install Jupyter, OpenCV, Tensorflow, matplotlib, Numpy.

If you use Anaconda, you still have to install Tensorflow.

Next week, we will learn how to set up virtual box to avoid compatibility issues.

GPU computing resources

AU HPC cluster GPU node (available in October). Tutorials offers in October by CTRL.
One GPU computer available in Room 203.
AWS educate.

Office hours (DMTI 204)

- Monday: 4pm-5pm (Away October 28th). The office hour on that day will be moved to the previous Friday.
- Wed: 3:30pm-5pm (Traveling September 18th). On that day the office hour will be moved to the morning.
- If you can't make the office hours, you can schedule with me another time to meet up. I need at least 48 hours notice to respond to your email for the scheduling.

Take-home exercise

- Check blackboard for tutorials
- Install Numpy, Scippy, skimage, matplotlib.
- You can install Anaconda 3
- Numpy tutorial

- https://stackabuse.com/numpy-tutorial-a-simple-example-based-guide/
- https://www.youtube.com/watch?v=xECXZ3tyONo

Next class:

- Simple visual system
- Image formation
- Pin hole camera
- Python/Numpy primer