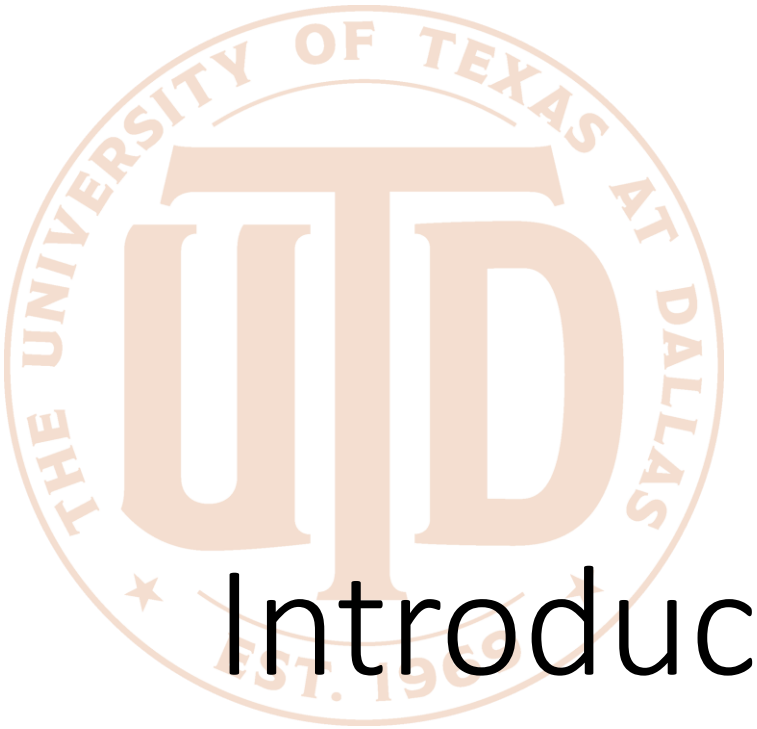


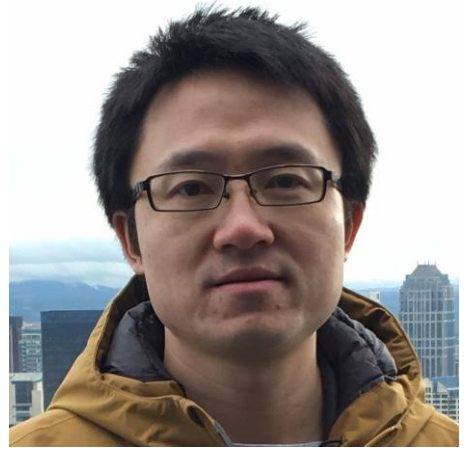
Introduction to Computer Vision

CS 4391 Introduction Computer Vision

Professor Yu Xiang

The University of Texas at Dallas

Who am I?



- Assistant Professor in CS at UTD (joined Fall 2021)
- Intelligent Robotics and Vision Lab at UTD

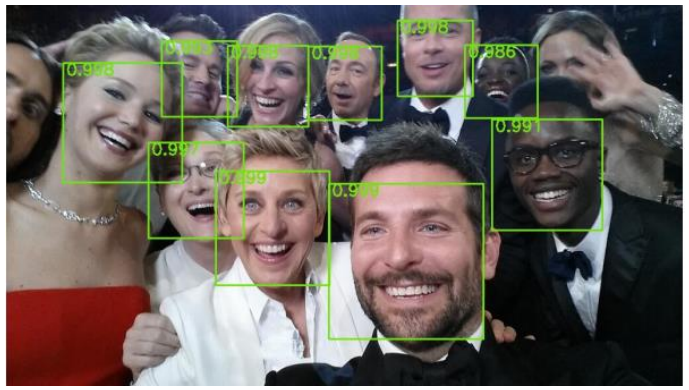
https://labs.utdallas.edu/irvl/

- Senior Research Scientist at NVIDIA (2018 – 2021) Robotics
- Ph.D. University of Michigan at Ann Arbor 2016

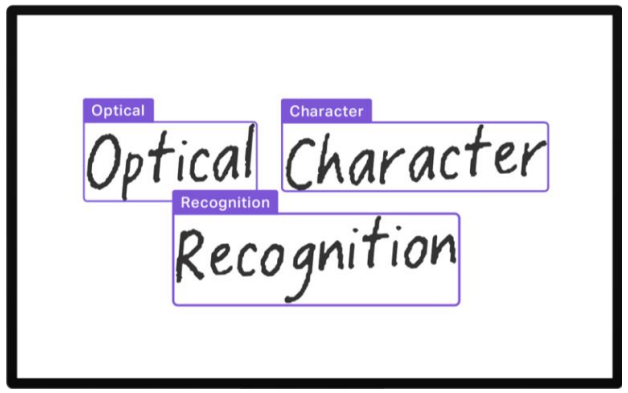
Introduce yourself

- Name
- Major program
- Which year in the program?
- Why are you interested in computer vision?

What is Computer Vision?



Face Detection



Optical Character Recognition (OCR)

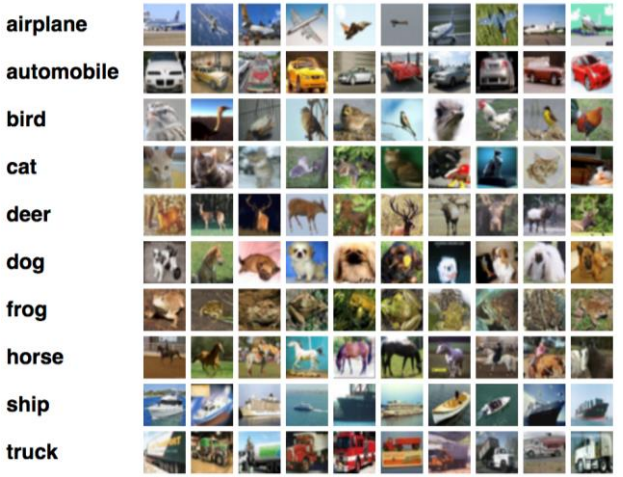
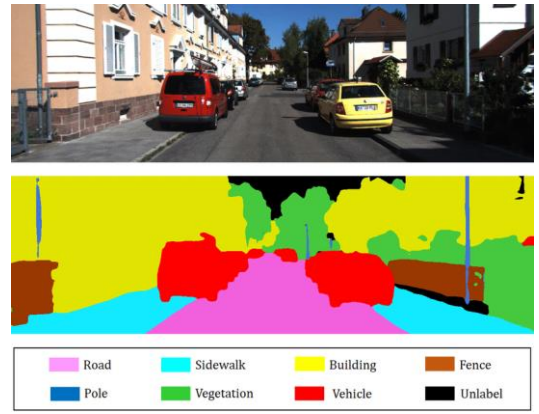


Image Classification

Panorama Stitching

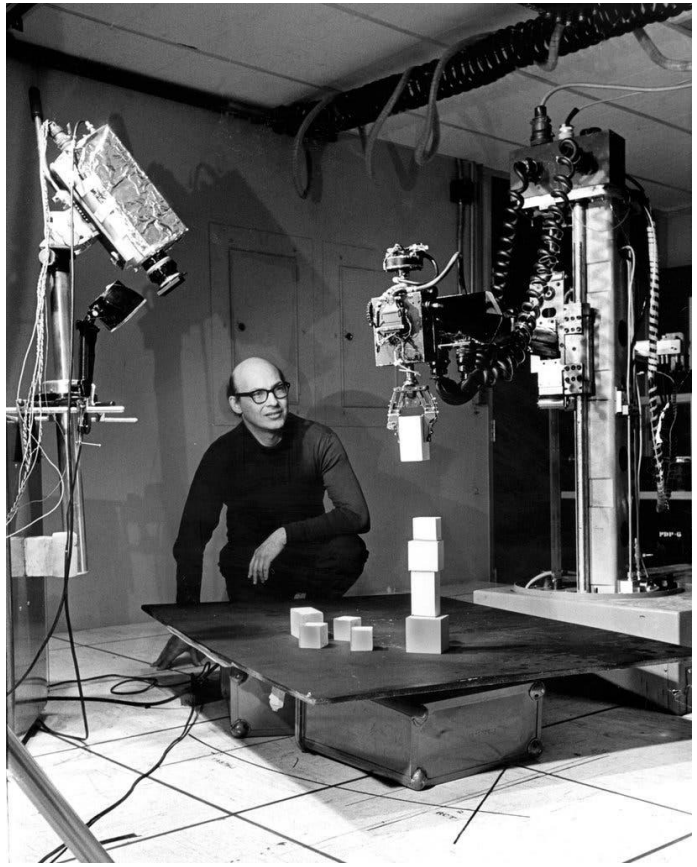
Surveillance



Semantic Segmentation

Computer vision is much more beyond image classification and processing

The Origin of Computer Vision



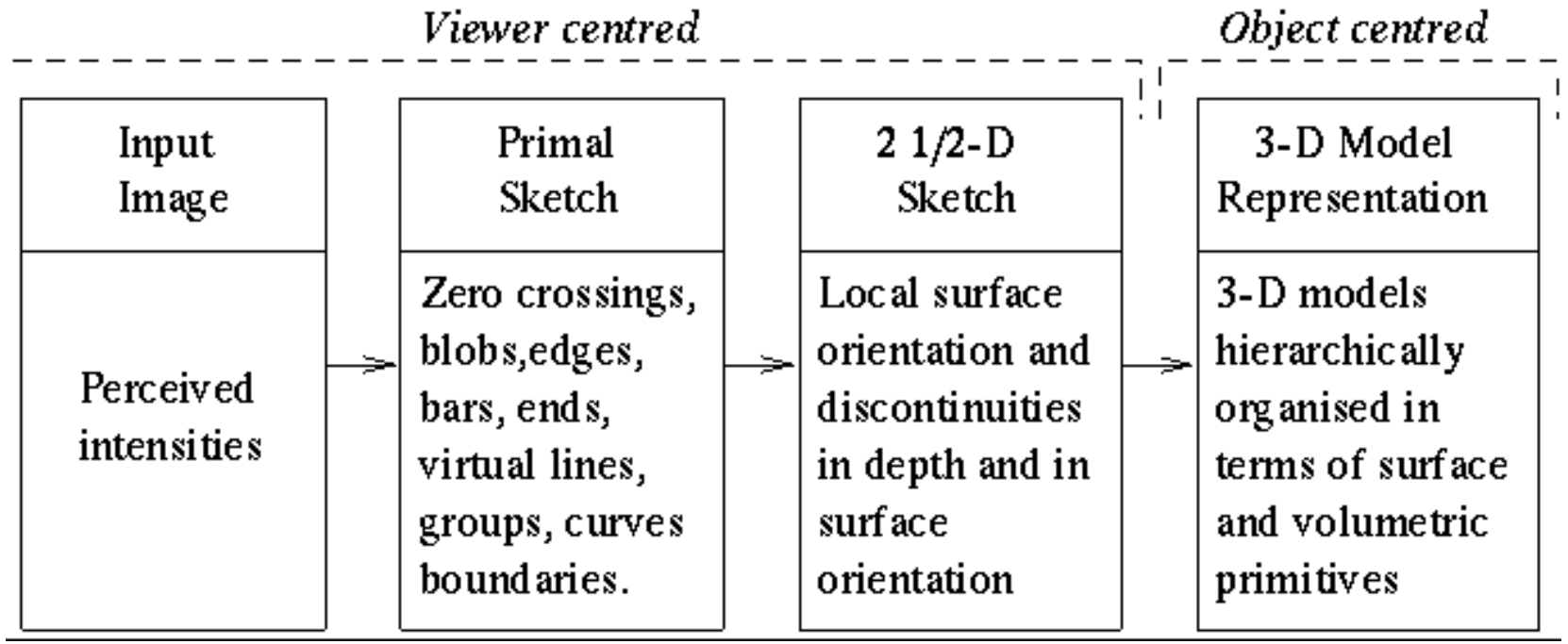
Marvin Minsky in a lab at MIT in 1968

An undergraduate project assigned by Marvin Minsky in 1966

“spend the summer linking a camera to a computer and getting the computer to describe what it saw”

Understand the 3D world from 2D images like humans

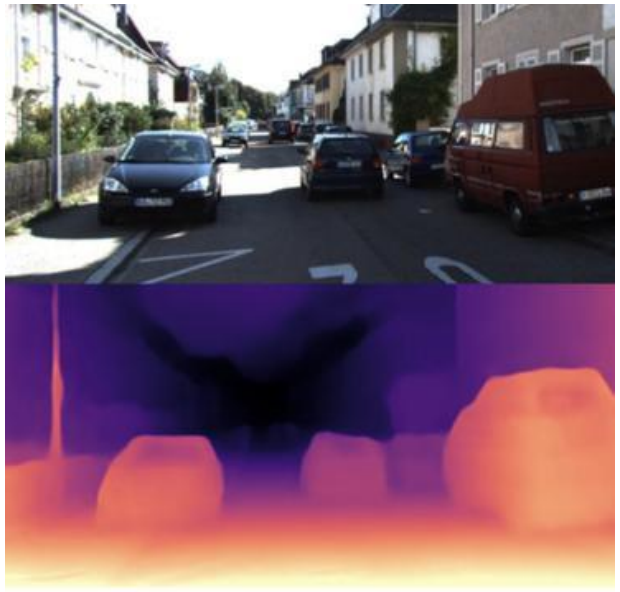
David Marr's Theory of Vision (Neuroscientist)



https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GOMES1/marr.html

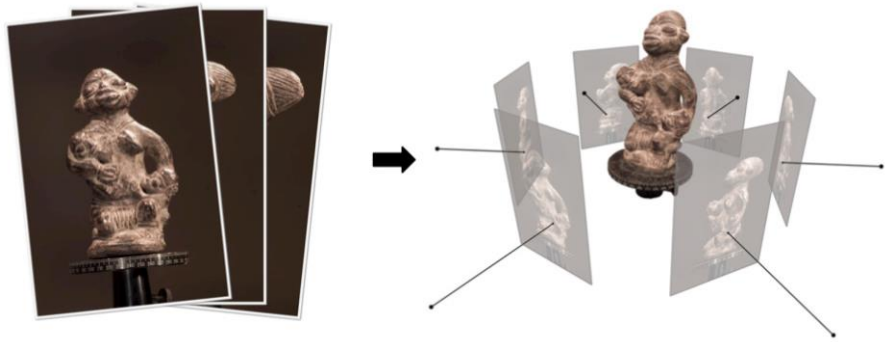
D. Marr. Vision. W. H. Freeman and Co., 1982.

What is Computer Vision?

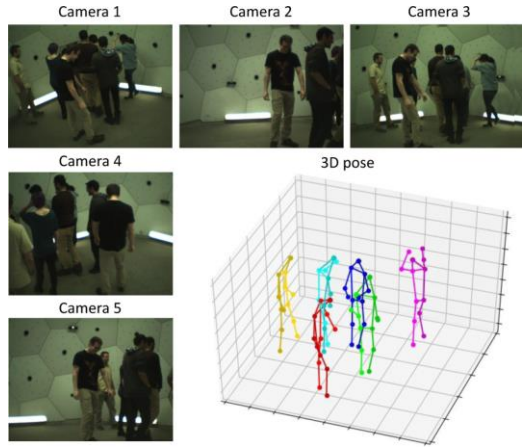


Depth Estimation

Structure from Motion



3D Reconstruction

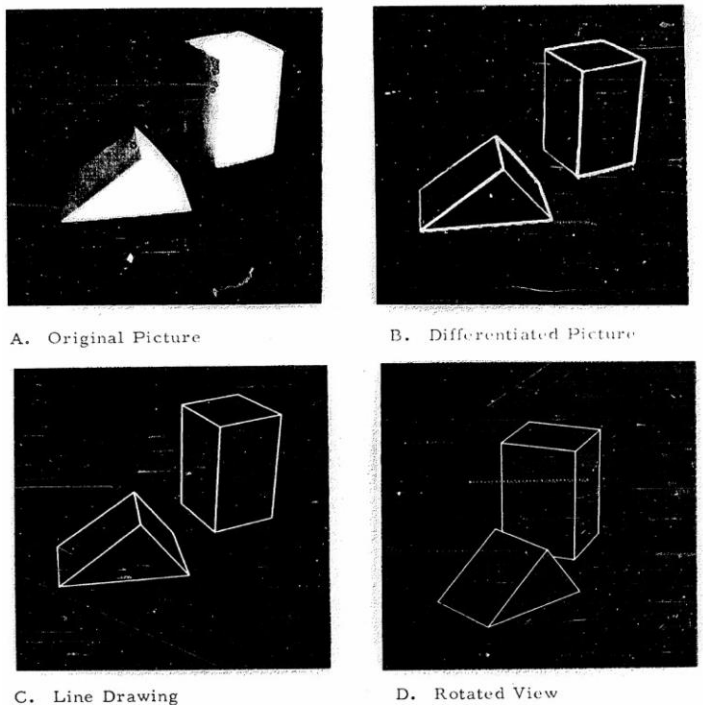


3D Human Pose Estimation
Dong et al. CVPR'19

Understand the 3D world from 2D images

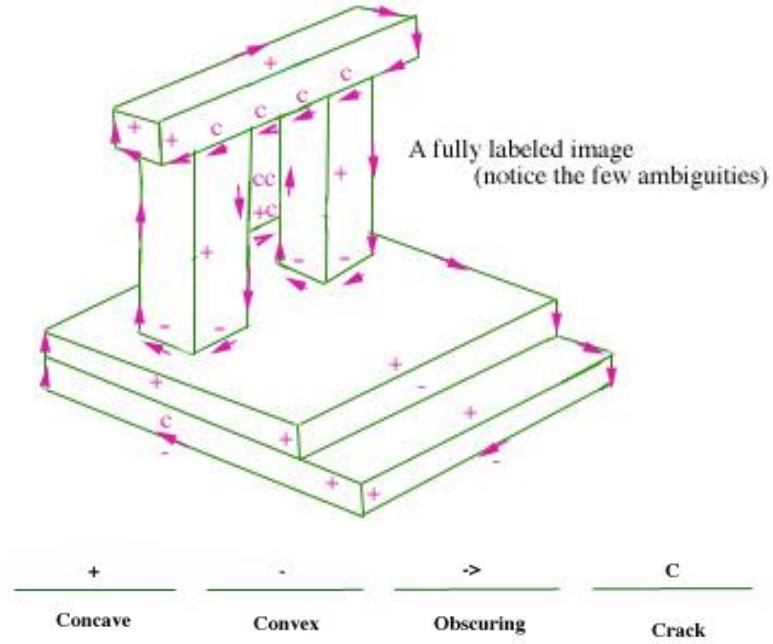
A Brief CV History and My Chosen Milestones

- 1970s
  - Recover 3D structure of the world from images

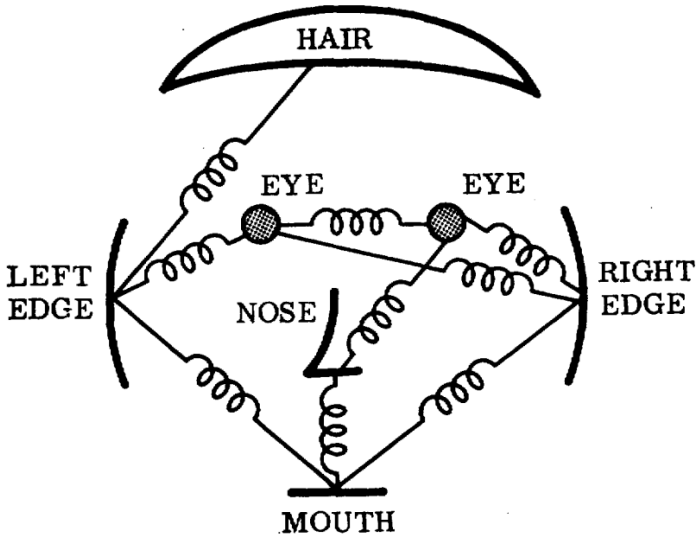


Blocks World

Roberts: Machine perception of three-dimensional solids. PhD Thesis, 1963



Line Labeling

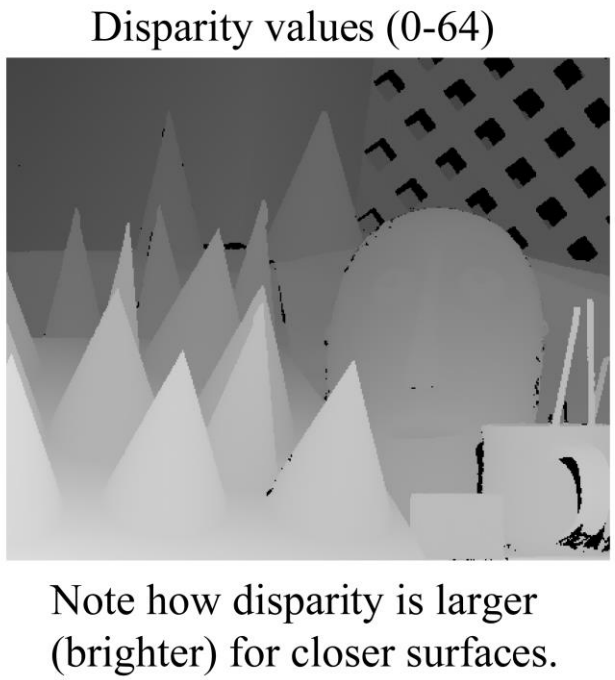


Pictorial Structure

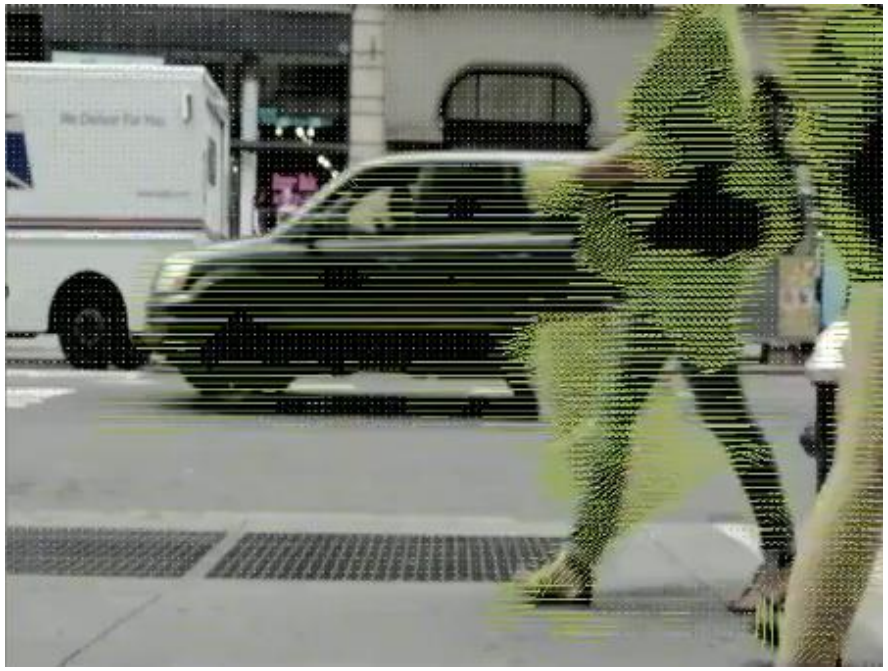
Fischler and Elschlager 1973

A Brief CV History and My Chosen Milestones

- 1980s
  - Stereo correspondence algorithms and optical flow algorithms



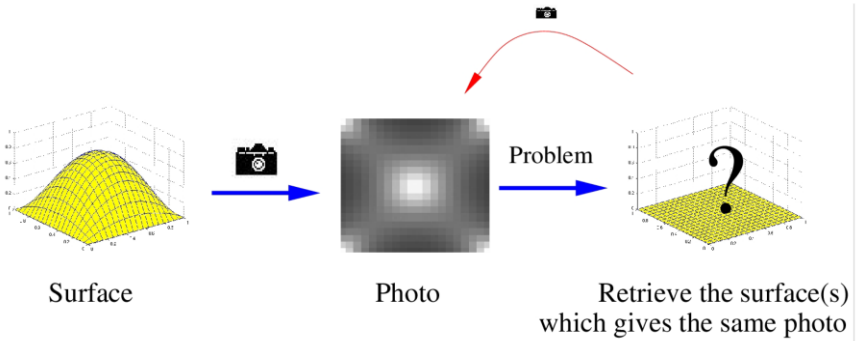
Stereo Correspondence



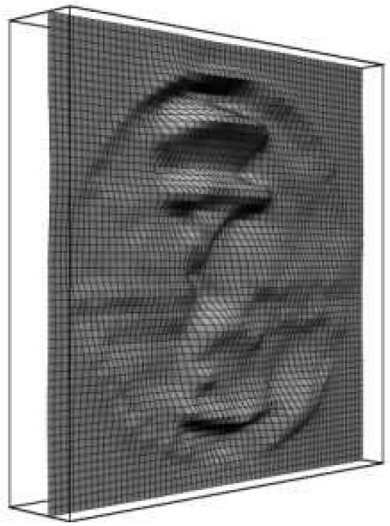
Optical Flow

A Brief CV History and My Chosen Milestones

- 1980s
  - Shape from X techniques (shape from shading, shape from texture, shape from shadows)
  - Edge and contours



Shape from shading



Freeman and Adelson 1991

Canny edge detector. Canny, 1986

A Brief CV History and My Chosen Milestones

- 1980s
  - Markov Random Fields (MRFs)

$$E(x) = \sum_i \underbrace{\Psi_i(x_i)}_{\text{Unary}} + \sum_{i \sim j} \underbrace{\Psi_{i,j}(x_i, x_j)}_{\text{Pairwise}}$$

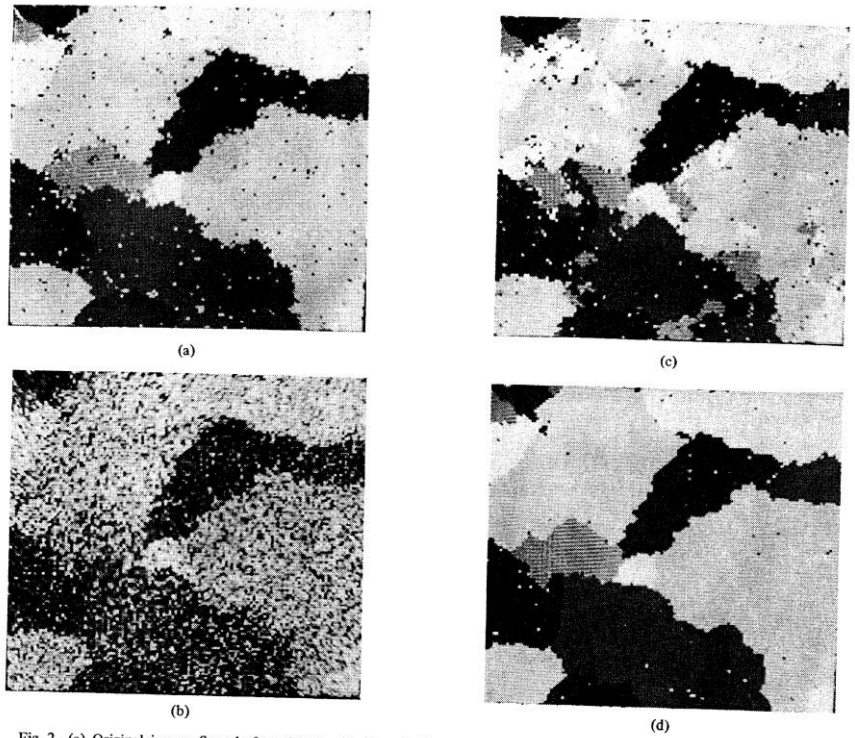
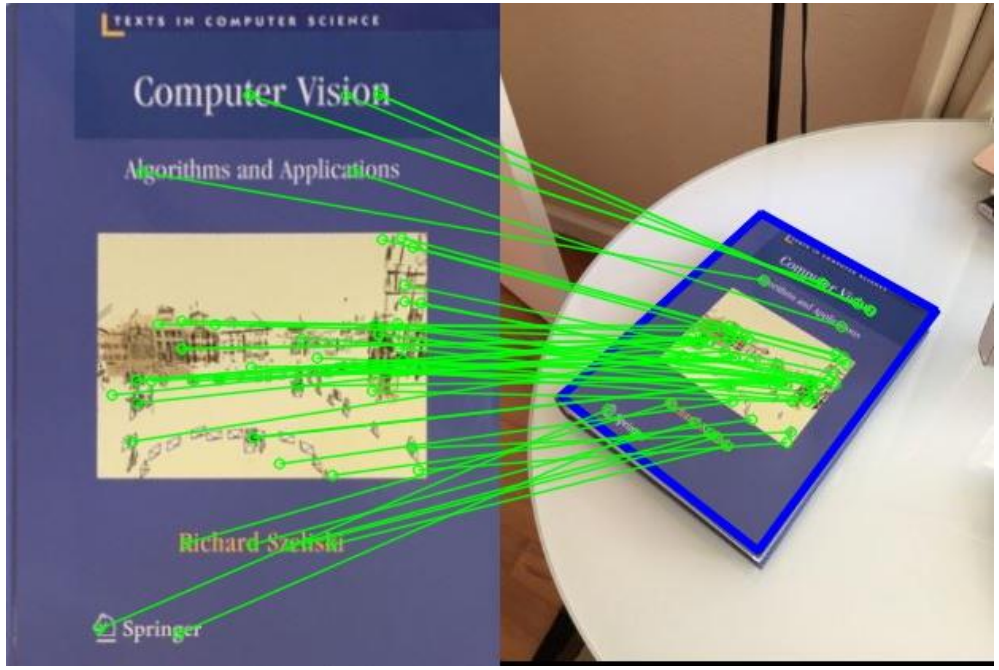


Fig. 2. (a) Original image: Sample from MRF. (b) Degraded image: Additive noise. (c) Restoration: 25 iterations. (d) Restoration: 300 iterations.

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

A Brief CV History and My Chosen Milestones

- 1990s
  - Structure from Motion and Multi-view Reconstruction
  - Scale Invariance Feature Transform (SIFT)



David Lowe: Object recognition from local scale-invariant features. ICCV, 1999.

A Brief CV History and My Chosen Milestones

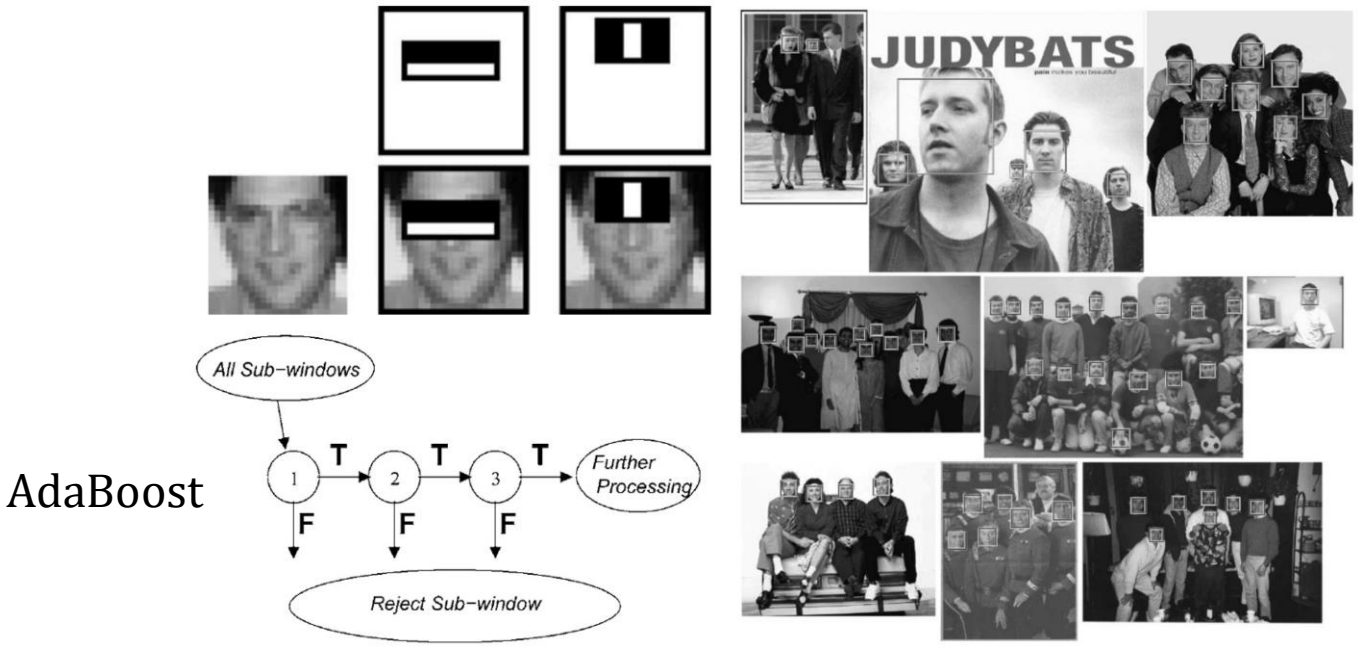
- 1990s
  - Statistical learning techniques started appearing

Eigenfaces

Turk and Pentland: Face recognition using Eigenfaces. CVPR, 1991

A Brief CV History and My Chosen Milestones

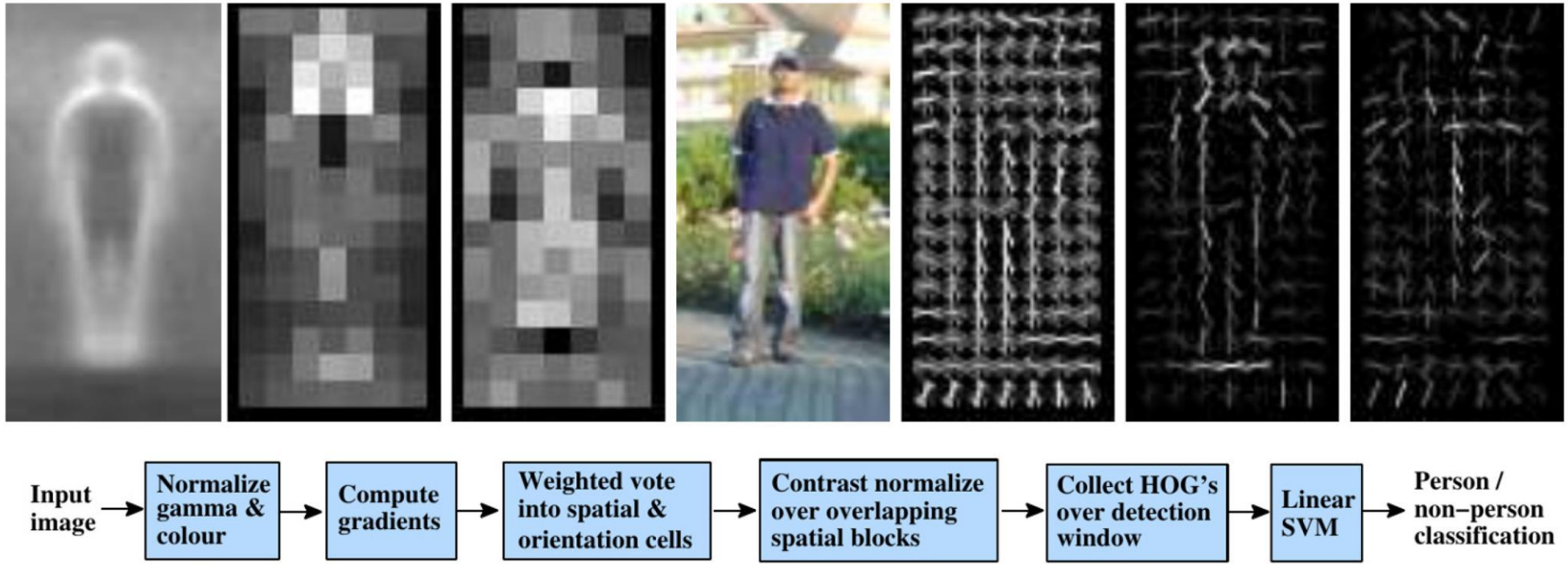
- 2000s
  - Data-driven and learning approaches
  - Cascaded classifiers for object detection



Viola and Jones: Robust Real-time Object Detection. IJCV, 2001.

A Brief CV History and My Chosen Milestones

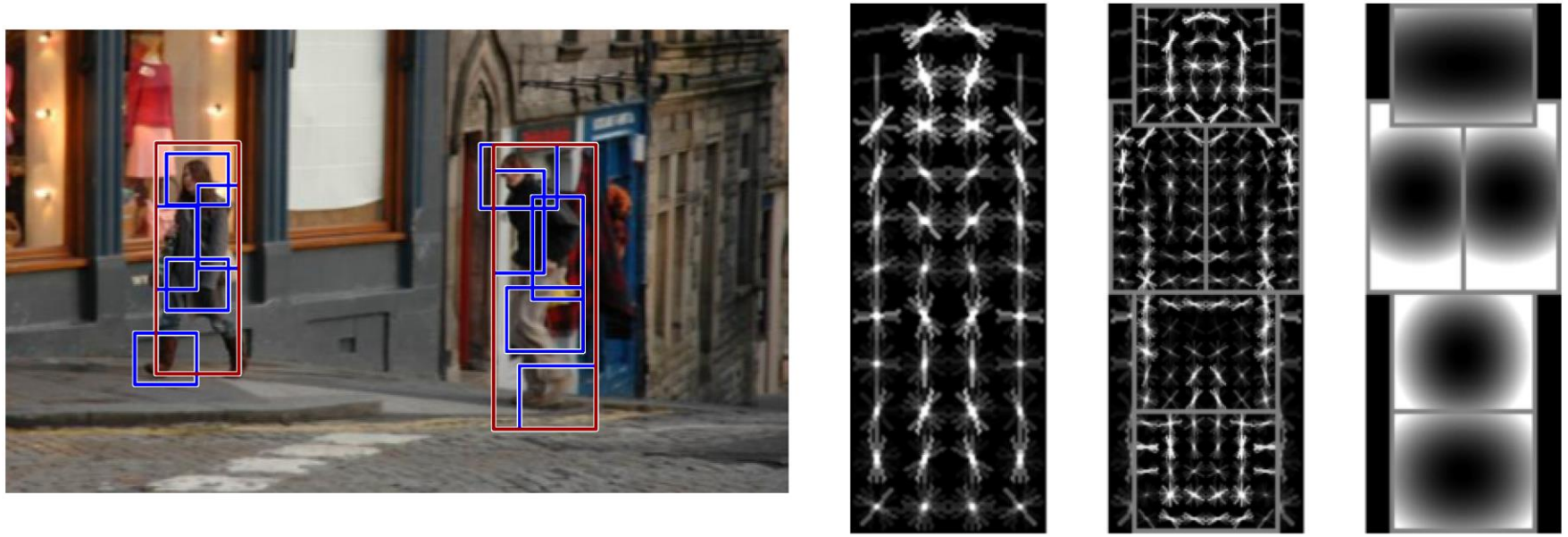
- 2000s
  - Histogram of Oriented Gradients for object detection



Dalal and Triggs: Histograms of Oriented Gradients for Human Detection. CVPR, 2005.

A Brief CV History and My Chosen Milestones

- 2000s
  - Deformable parts models for object detection



Felzenszwalb et al. Object detection with discriminatively trained part-based models . TPAMI, 2009.

A Brief CV History and My Chosen Milestones

- 2000s
  - Datasets

PASCAL VOC, Everingham et al., 2005 - 2012

ImageNet, Deng et al., 2009

A Brief CV History and My Chosen Milestones

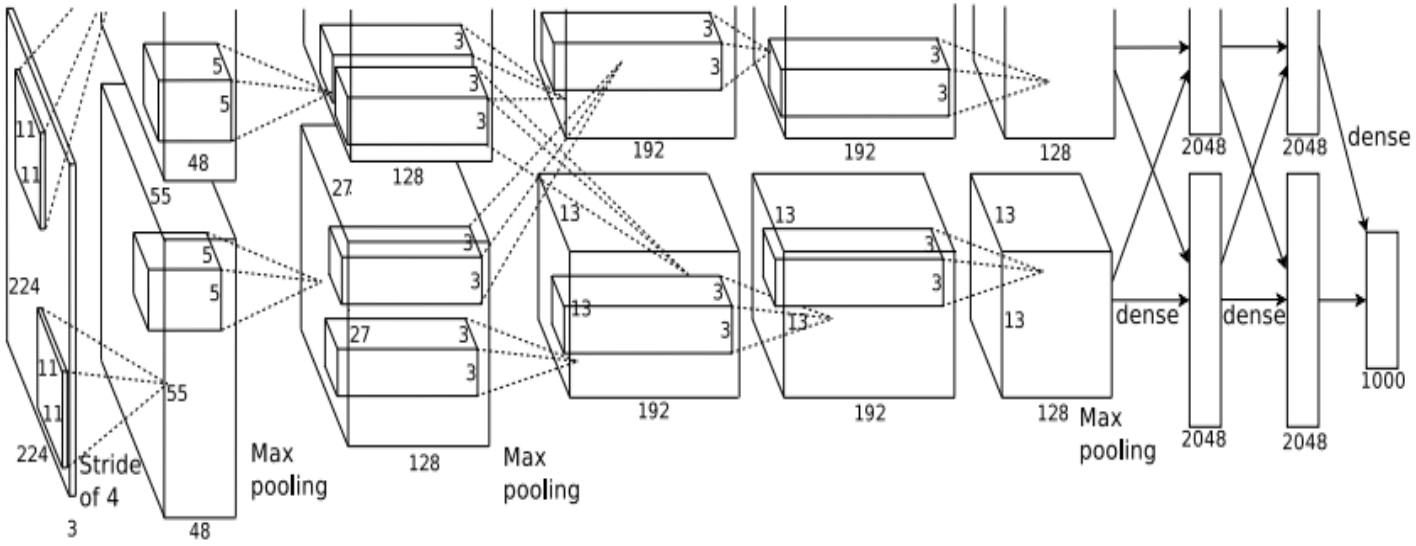
- 2000s
  - Large-scale structure from motion

San Marco Square: 13,699 images, 4,515,157 points

Agarwal et al. Building Rome in day. ICCV, 2009.

A Brief CV History and My Chosen Milestones

- 2010s
  - Deep Learning in CV

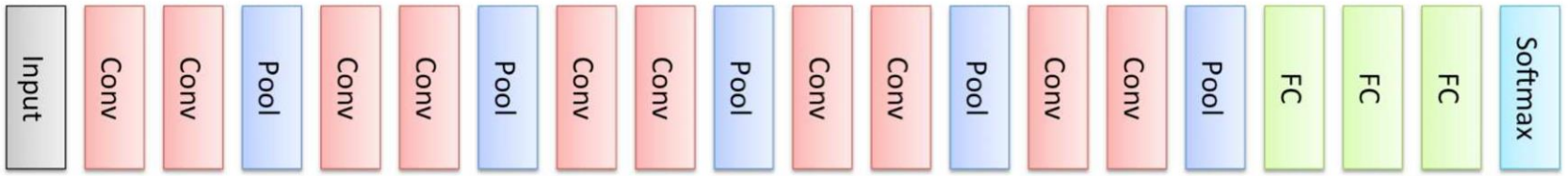


AlexNet. Krizhevsky et al., 2012, designed for ImageNet classification

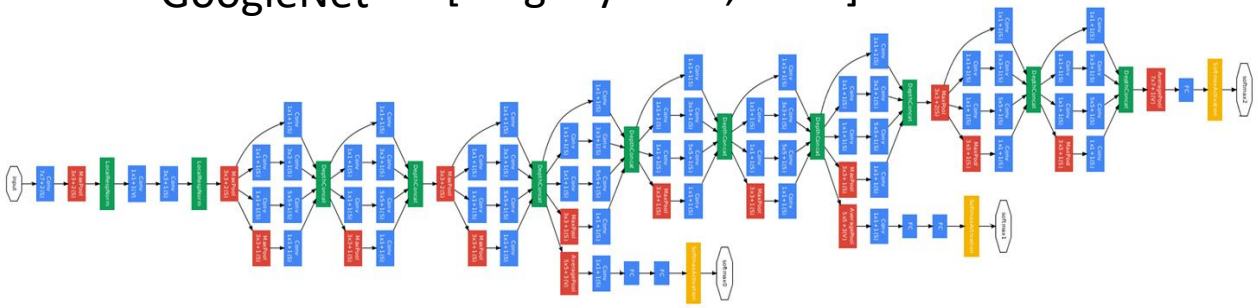
A Brief CV History and My Chosen Milestones

- 2010s
  - Deeper and wider networks

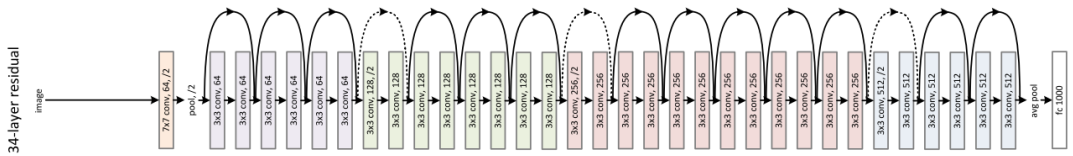
VGGNet [Simonyan and Zisserman, 2014]



GoogleNet [Szegedy et al., 2014]

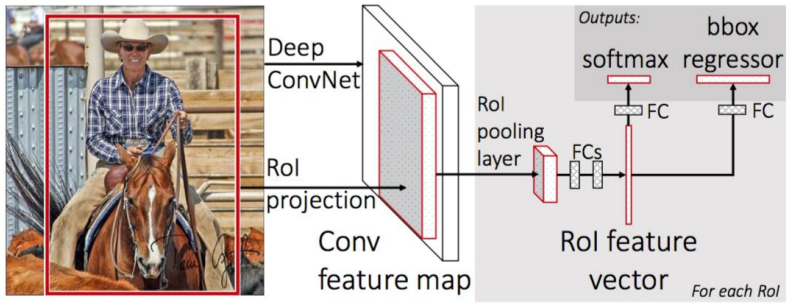


ResNet [He et al., 2015]

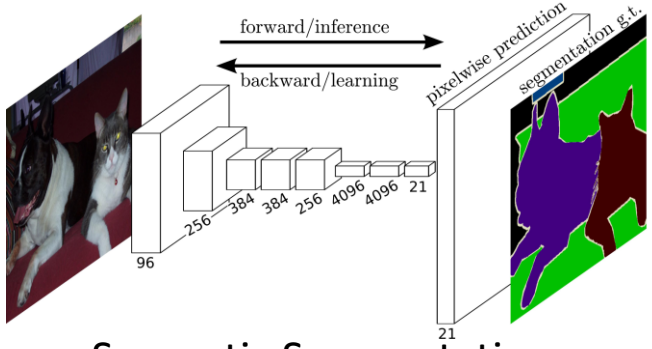


A Brief CV History and My Chosen Milestones

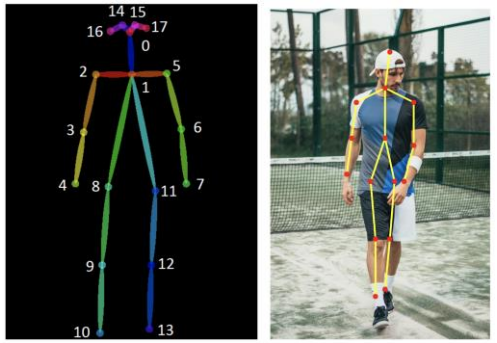
- 2010s
  - Neural networks for recognition



Object Detection (Fast RCNN, Girshick, 2015)

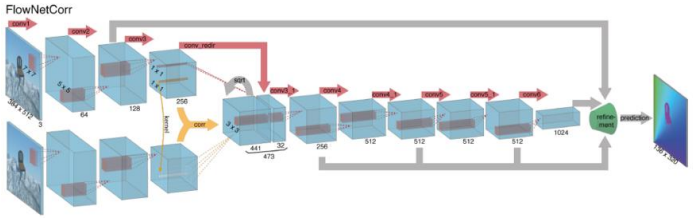


Semantic Segmentation
(FCN, Long et al., 2014)

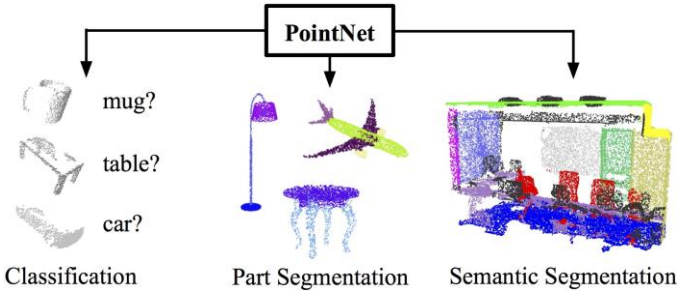


Human Pose Estimation
(OpenPose, Cao et al., 2017)

Depth Estimation (Eigen et al. 2014)



Optical Flow
(FlowNet Fischer et al. 2015)

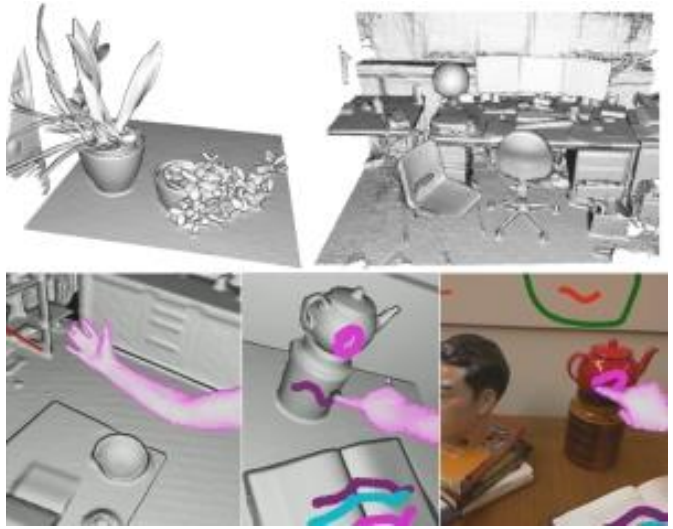


Point Cloud Recognition
(PoinetNet, Qi et al., 2016)

A Brief CV History and My Chosen Milestones

- 2010s
  - Depth sensing and 3D vision

Microsoft Kinect, 2010



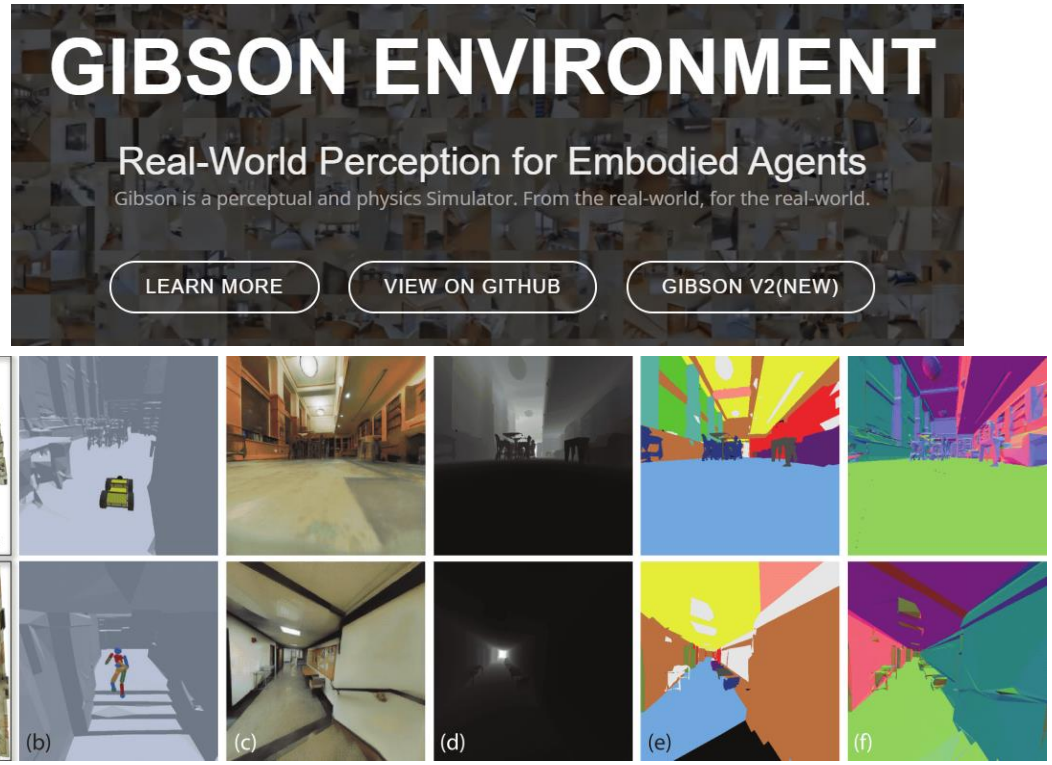
KinectFusion, Newcombe et al., 2011

DynamicFusion, Newcombe et al., 2015

A Brief CV History and My Chosen Milestones

- 2010s
  - Autonomous driving and embodied AI

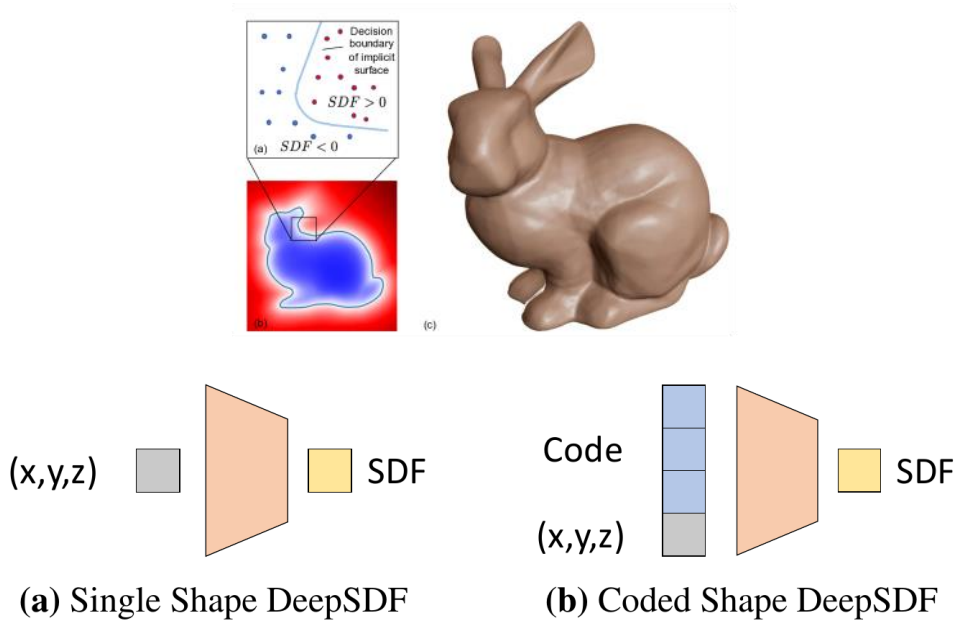
The KITTI dataset, Geiger et al., 2012



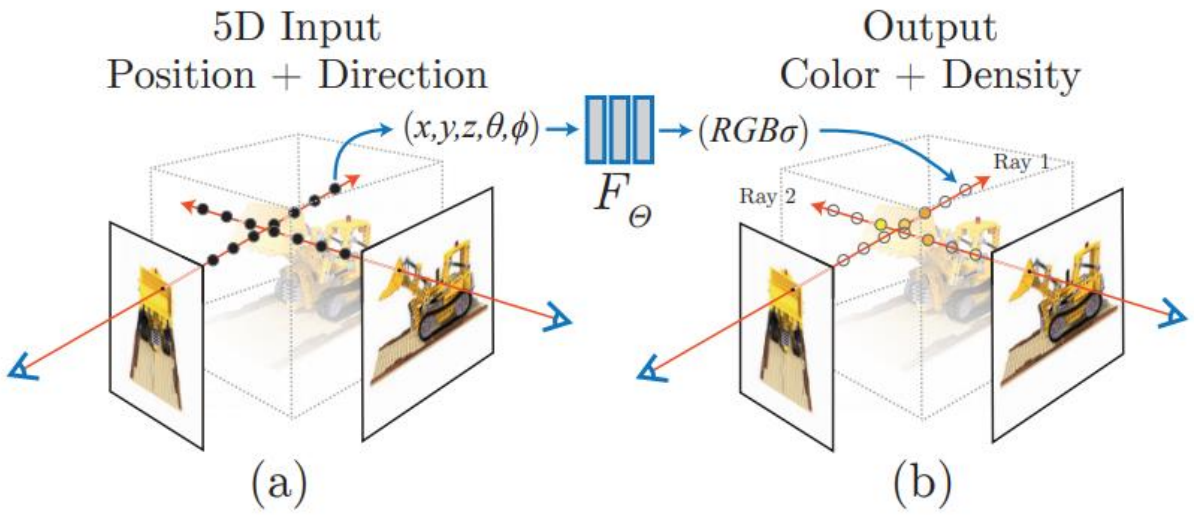
The Gibson environment, Xia et al., 2018

A Brief CV History and My Chosen Milestones

- 2010s
  - Neural implicit representations



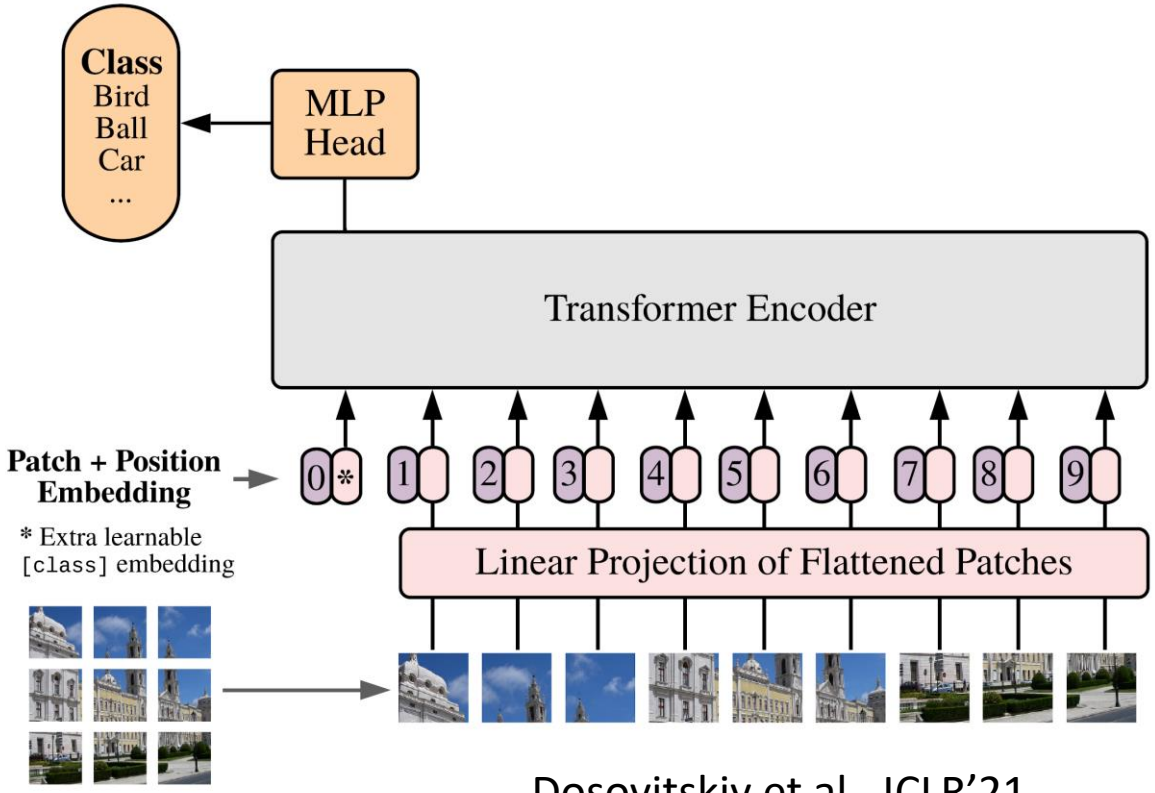
DeepSDF, Park et al., 2019



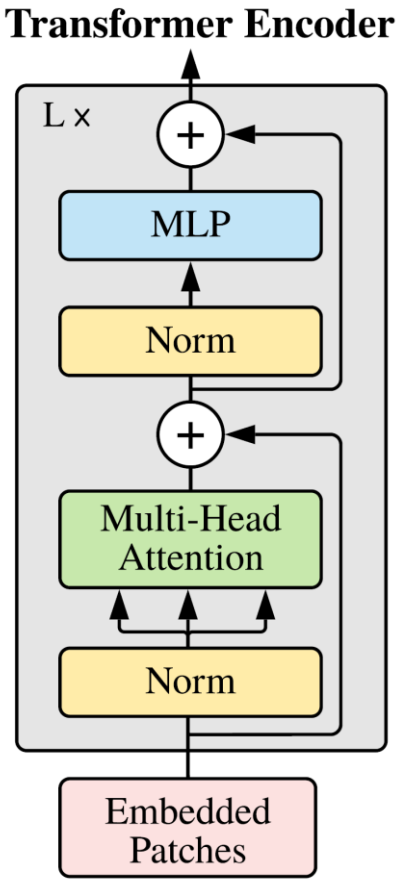
NeRF: Neural Radiance Fields. Midenhall et al. 2020

A Brief CV History and My Chosen Milestones

- 2020s
  - Vision transformers



Dosovitskiy et al., ICLR'21



A Brief CV History and My Chosen Milestones

- 2023
  - Large Multi-modal Model
  - GPT-4, etc

User What is funny about this image? Describe it panel by panel.

Source: hmmm (Reddit)

GPT-4 The image shows a package for a "Lightning Cable" adapter with three panels.

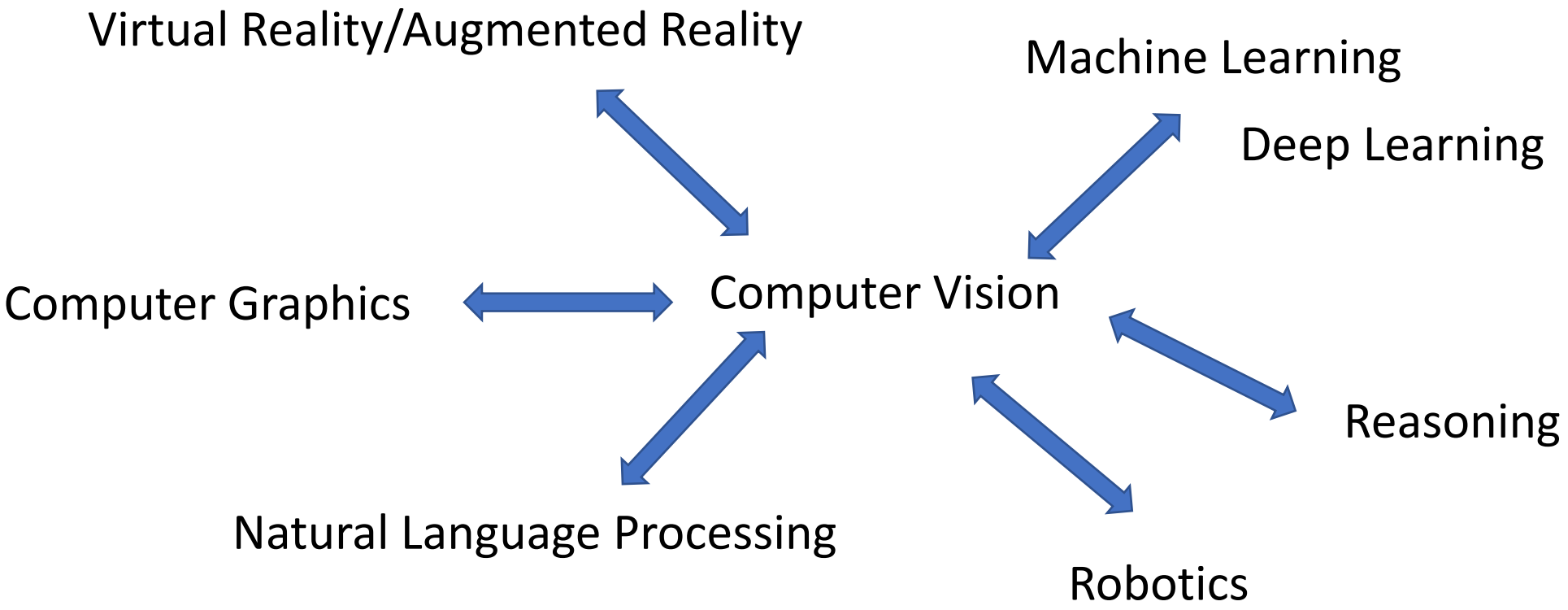
Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

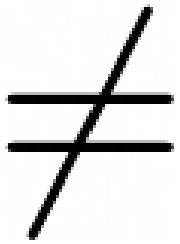
Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

Computer Vision in AI



Computer Vision in AI

Datasets  Real World

Test your algorithms in the real world, e.g., with a camera

What will you learn in this course?

- Geometry in computer vision
  - Camera model, stereo geometry, multi-view geometry, etc.
- Image Features
  - Point features, edges, contours, etc.
- Deep learning in computer vision
  - Convolutional neural networks, recurrent neural networks, generative networks, etc.
- Visual recognition
  - Object detection, semantic segmentation, human pose estimation, images and languages, etc.

Grading Policy

- Homework (50%)
  - 5 homework in total
  - Individual submission
- Midterm Exam (20%)
- Final Exam (25%)
- In-class Activity (5%)

Course Details

- Textbook
  - Richard Szeliski. **Computer Vision: Algorithms and Applications**. 2nd Edition. Springer. Available online https://szeliski.org/Book/
  - David Forsyth, Jean Ponce. **Computer Vision: A Modern Approach**, 2nd Edition. Pearson, 2011. (Optional)
  - Richard Hartley. **Multiple View Geometry in Computer Vision**, 2nd Edition. Cambridge University Press, 2004. (Optional)
- My office hour

  Tuesday & Thursday 2:00PM – 3:00PM
  ECSS 4.702
- TA office hour: TBD
- Course access and navigation: eLearning

Questions?