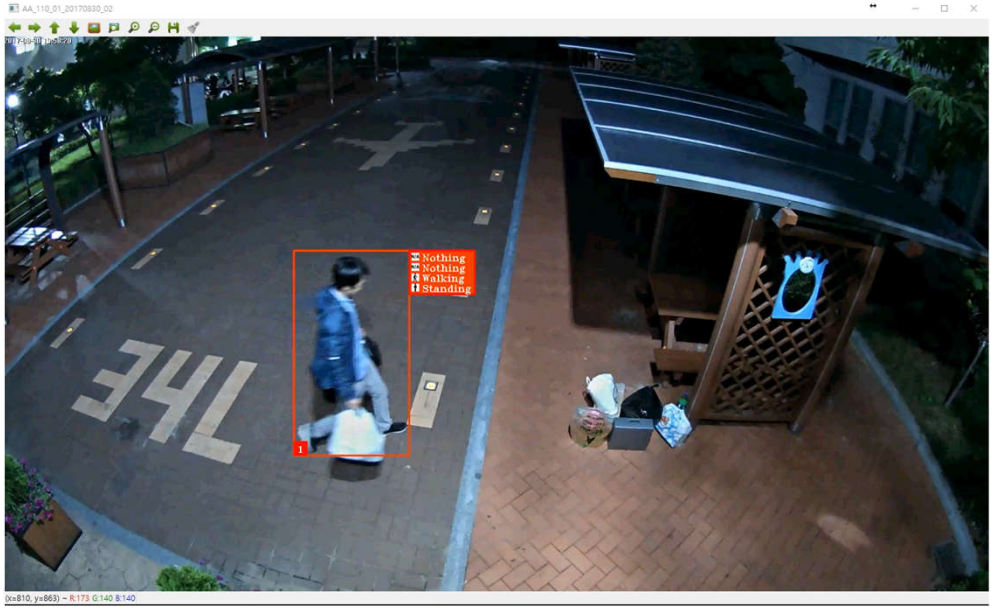
Lecture 1: Introduction and Historical Context

박사과정 김성빈 chengbinjin@inha.edu,
지도교수 김학일 교수 hikim@inha.ac.kr
인하대학교 컴퓨터비전 연구실

Instructors

Instructors	http://vision.inha.ac.kr	
Professors	Hale Kim (김학일) Professor, ICE Hitech Rm 514 (Tel: 7385) hikim@inha.ac.kr 	Xuenan Cui (최학남) Professor, ICE Hitech Rm 401 (Tel: 8427) xncui@inha.ac.kr 
TAs	Chengbin Jin (김 성빈) PhD student, ICE Hitech Rm 525 (Tel: 7385) chengbinjin@inha.edu 	Nguyen Thi Hai Binh (띠하이빈) PhD student, ICE Rm 209 Inha Dream Center binhnth12@gmail.com 

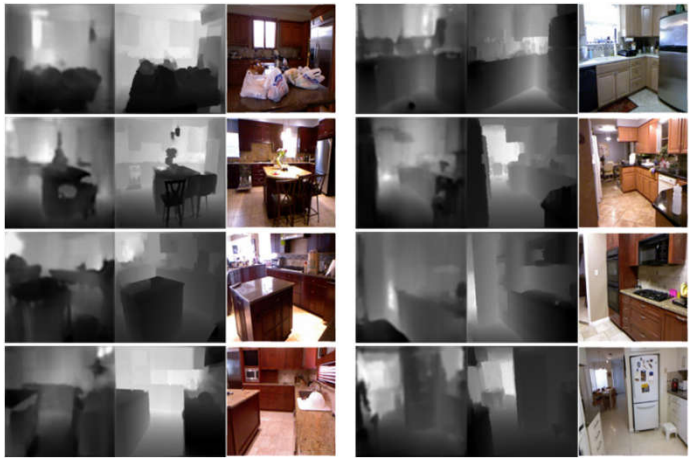
Who am I?



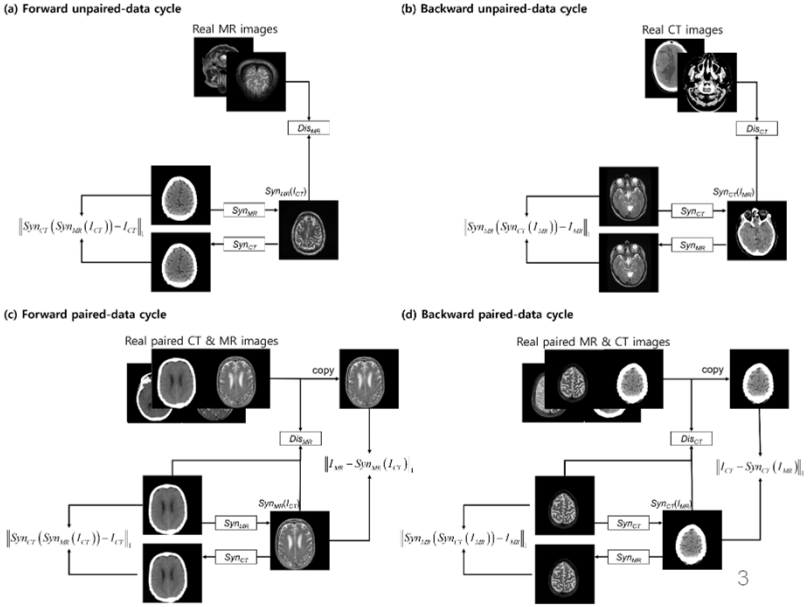
Action detection (video surveillance)

Github page:
https://github.com/ChengBinJin

Cheng-Bin Jin
(김성빈)



Depth estimation from a single image



Generative Adversarial Networks (GANs)

Class Overview

CS231n 2016

Homepage: http://cs231n.stanford.edu/2016/

Videos: https://www.youtube.com/watch?v=g-PvXUjD6qg&list=PLlJy-eBtNFt6EuMxFYRiNRS07MCWN5UIA

CS231n 2017

Homepage: http://cs231n.stanford.edu/2017/

Videos: https://www.youtube.com/watch?v=vT1JzLTH4G4&list=PLC1qU-LWwrF64f4QKQT-Vg5Wr4qEE1Zxk

Contents

- Schedule and grading policy
- More deep learning courses
- A brief history of computer vision
- Deep learning class overview

Introduce Yourself

- Name
- Where are you from
- Laboratory and department
- Advisor
- Interested domain
- Any other …

For the beginner project

- Final project will be evaluated base on team work (can not be individual)
- One team should include 2 or 3 persons
- Try to decide your team members by next class

Schedule & Grading Policy

Weekly Schedule (1)

WK	Lecture Subject Contents	Instructor
1 09/05	Orientation (1) Course Overview (2) Course logistics	Kim
2 09/12	Introduction to AI and Computer Vision (1) Historical context (2) Deep learning: Intro	C.-B.
3 09/19	Linear classification 1 (1) Image classification and data-driven approach (2) K-nearest neighbor	Cui
4 09/26	Linear classification 2 (1) Higher-level representations, image features (2) Optimization, stochastic gradient descent	Cui
5 10/03	Training neural networks - part 1 (1) Activation functions, weight initialization, gradient flow, batch normalization (2) Babysitting the learning process, hyperparameter optimization	Cui

Weekly Schedule (2)

WK	Lecture Subject Contents	Instructor
6 10/10	Training neural networks - part 2 (1) Parameter updates, ensembles, dropout (2) Convolutional neural networks: Intro	C.-B.
7 10/17	Convolutional neural networks (1) Architectures, convolution/pooling layers (2) Case study of ImageNet challenge wining ConvNets	C.-B.
8 10/24	Mid Exam Assignment of Term-Project	Kim
9 10/31	ConvNets for spatial localization (1) Object localization (2) Object detection	Binh
10 11/07	Understanding and visualizing convolutional neural networks (1) Backprop into image: visualizations (2) Deep dream, artistic style transfer (3) Adversarial fooling examples	C.-B.

Weekly Schedule (3)

WK	Lecture Subject Contents	Instructor
11 11/14	Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) (1) RNN language models (2) Image captioning	C.-B.
12 11/21	Generative models (1) Pixel RNN/CNN (2) Variational autoencoders (3) Generative adversarial networks	C.-B.
13 11/28	Deep reinforcement learning (1) Policy gradients, hard attention (2) Q-learning, Actor-Critic	Binh
14 12/05	Final Exam	Kim
15 12/12	Project Presentation	All
16 12/19	Project Presentation (If necessary)	All

Grading Policy

Evaluation							
Mid Exam	Final Exam	Assignments	Class Activity (be present)	Term Project	Quiz	Presentation	Total
10 %	10 %	10%	5 %	30 %	25 %	10 %	100 %
2018.10.24	2018.12.05	Every week	Every week	2018.12.12	Every week	Every week	

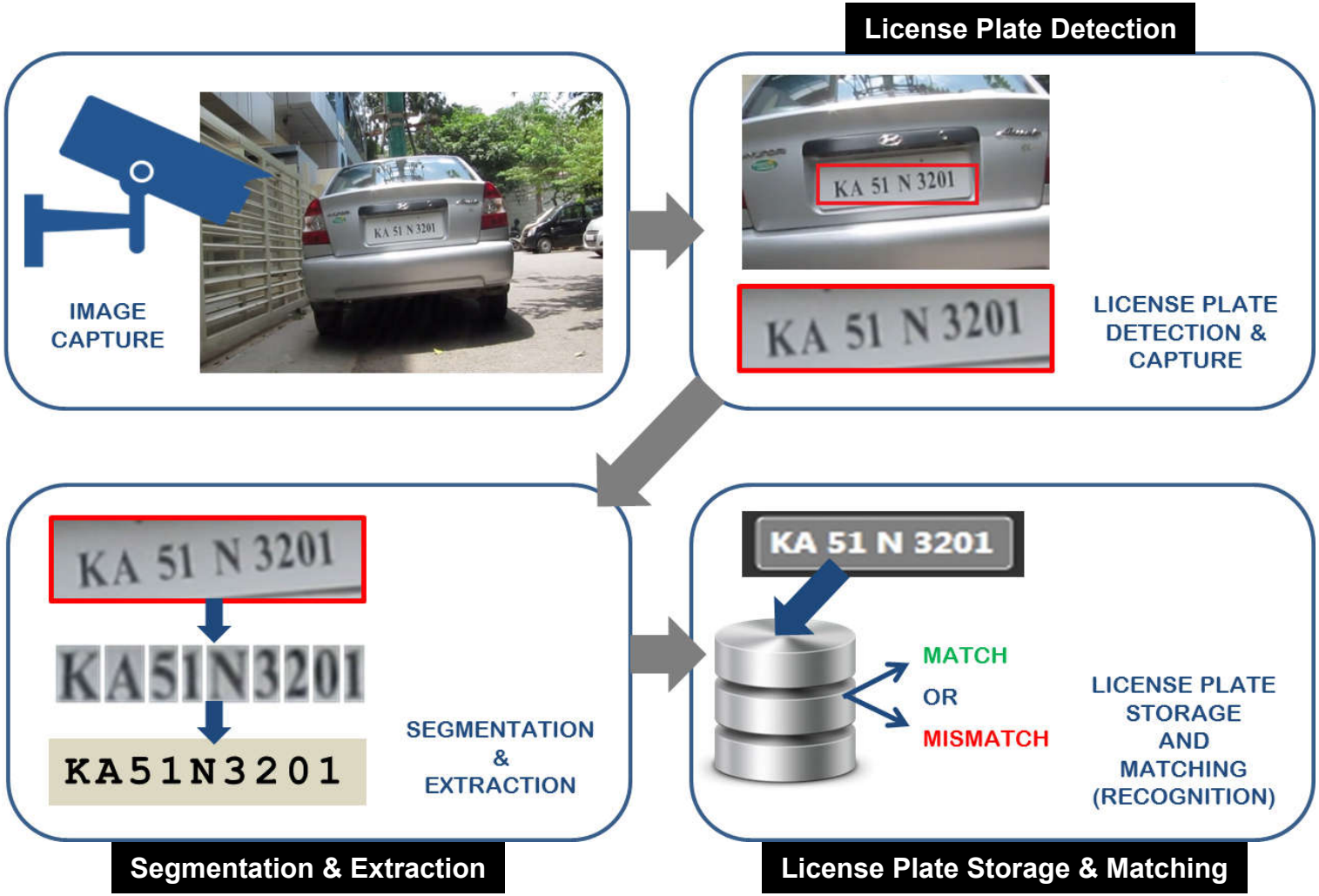
Note:

1. Mid & final exam: lectures (PPT) & coding parts in assignments
2. Assignments: assignment 1~3 of the CS231n 2016 http://cs231n.stanford.edu/2016/syllabus.html
3. Class activity: attendance check before class
4. Term project:
   - (1) license plate detection and recognition; (2) on-road object detection
   - Scores will be evaluated based on the team ranking
   - No open or commercial library is allowed.
5. Quiz: will be taken in the beginning of every class to check students' pre-studying the CS231n video
6. Presentation: 2 or 3 students every week

Python Numpy Tutorial:
http://cs231n.github.io/python-numpy-tutorial/

Final Project (Beginner)

- **License Plate Detection & Recognition**



Parking Lot Data

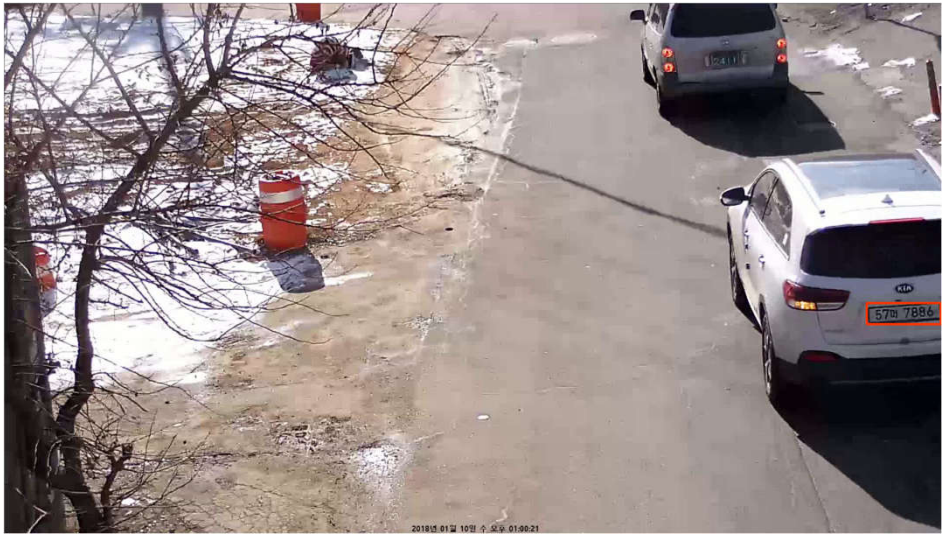
789 473 343 76
46고7080

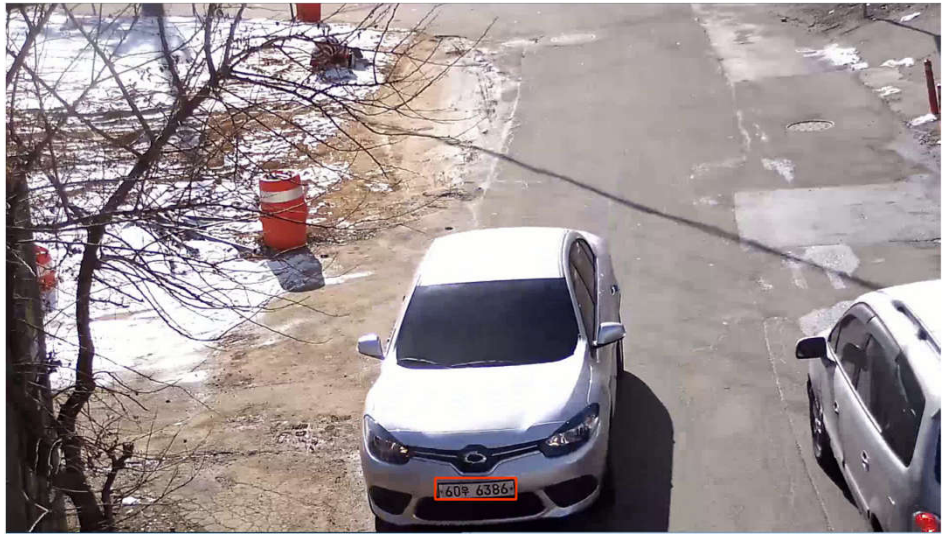
591 509 299 74
34구8107

756 419 310 86
62고9567

225 805 297 149
80조9523

CCTV Data





Competition Results

	Team	parking				cctv				
		Det acc (%)	Rec acc (%)	PT (ms)	Score	Det acc (%)	Rec acc (%)	PT (ms)	Score	Total
1		90.88	90.53	91.48	182.26	77.38	79.36	76.81	159.06	341.32
2		73.33	44.91	75.08	120.74	70.51	48.17	164.56	112.22	232.96
3		50.18	0.00	50.02	55.17	77.83	0.00	89.67	78.86	134.03
4		50,53	0.00	91.67	51.36	14.86	0.00	49.29	19.93	71.29
5		22.81	16.49	575.37	-8.24	2.00	0.00	825.02	-70.51	-78.75
6		96.49	0.00	1106.94	-4.20	7.10	0.00	1194.71	-102.38	-106.58
None		None	None	None	None	None	None	None	None	None

- Undergraduate students (last semester)

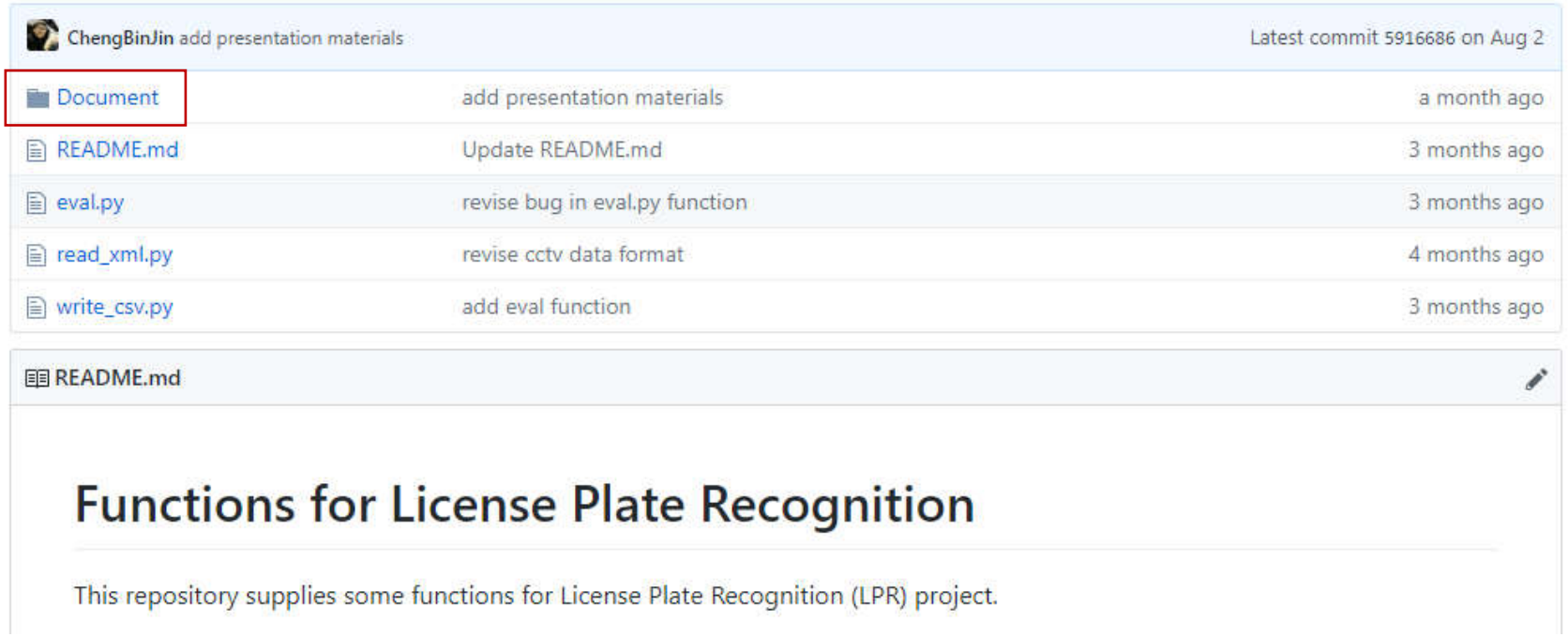
Parking: #285

CCTV: #451

More Information

https://github.com/ChengBinJin/License_plate_recognition

- Check “Document” file to see undergraduate students presentations and evaluated results.



More Deep Learning Courses

CS231n Overview

CS224d: Deep Learning for Natural Language Processing **(Winter 2017)**

- http://cs224d.stanford.edu/

CS20SI: Tensorflow for Deep Learning Research **(January-March, 2018)**

- http://web.stanford.edu/class/cs20si/

CS294-112: Deep Reinforcement Learning **(Fall 2018)**

- http://rail.eecs.berkeley.edu/deeprlcourse/calendar/

CS231n: Convolutional Neural Networks for Visual Recognition **(Spring 2018)**

- http://cs231n.stanford.edu/

모두를 위한 딥러닝 (김성훈 교수님) – 딥러닝 Part: https://hunkim.github.io/ml/

v2 is coming soon!

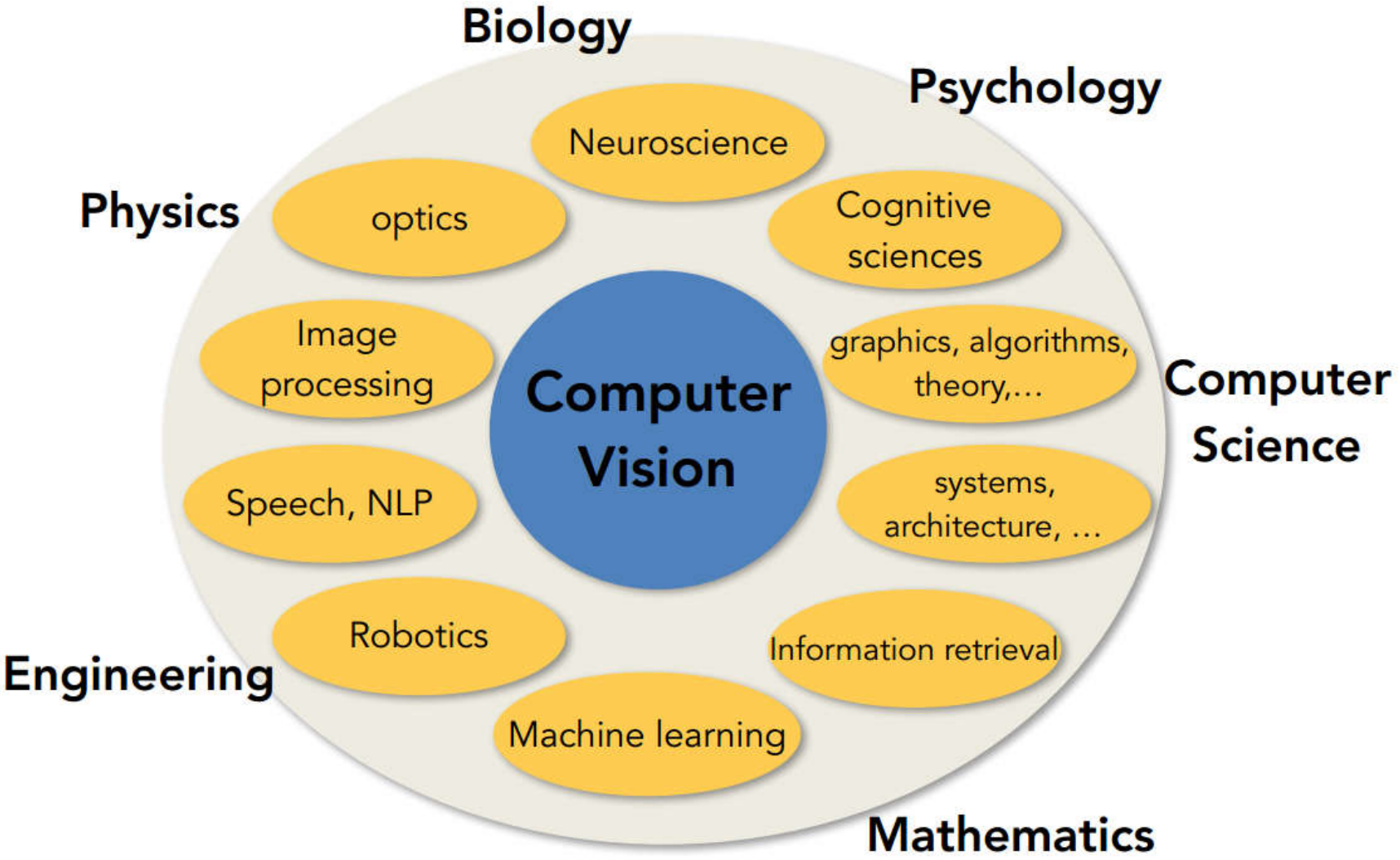
Welcome to Deep Learning Class

- **More sensors** than the number of owners
- Every 60s, we have 150 hours of videos uploaded YouTube server
- **Dark matter** on the Internet

Slide from CS231n

Computer Vision

- Compute vision is a field that really a truly **interdisciplinary**



Slide from CS231n

Requirements

Undergraduate course

- Digital Image processing (Prof 김학일 & Prof 최학남, 정보통신학과)

Graduate Courses

- Computer Vision (Prof 박인규, 정보통신학과)
- Computer Vision (Prof 송병철, 전자공학과)
- Special Topic in Image Processing (Pro. 최학남, 정보통신학과)
- Probabilistic Inference for Computer Vision (Pro. 이상철, 컴퓨터공학과)

A Brief History of Computer Vision

Brief History but Not Short History

- **Big bang of evolution:** 543million years
- Animals **start to diversify** and they can start to use all kind of tools to survive

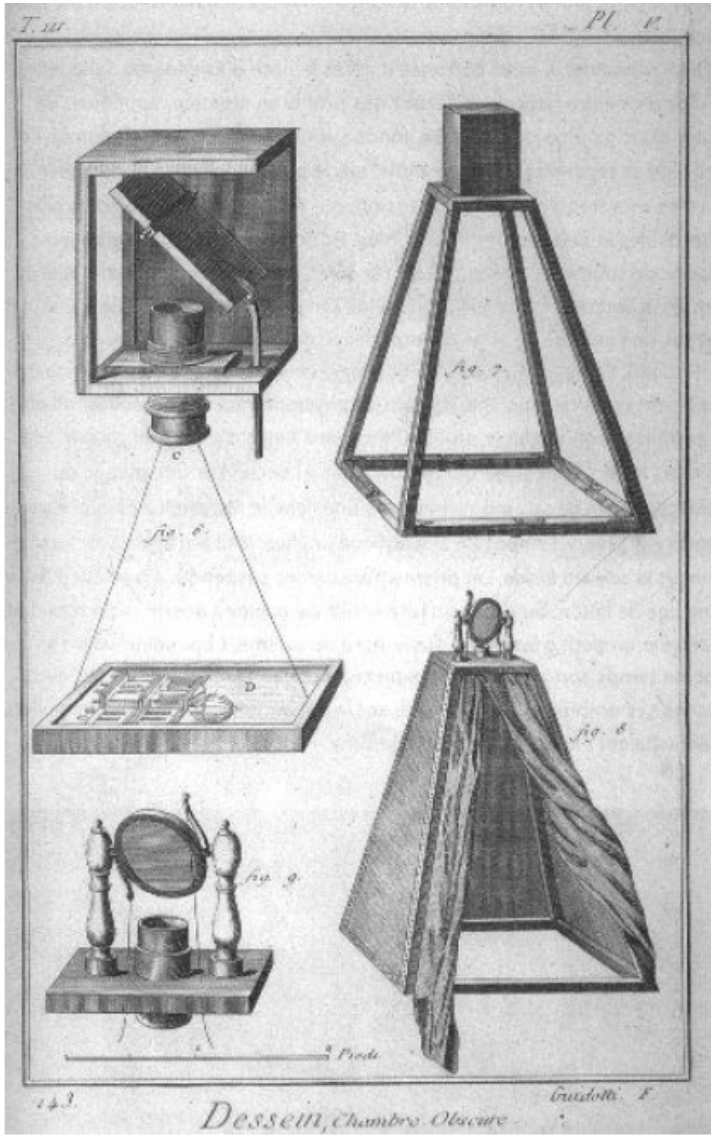
Slide from CS231n

Camera Obscura

Leonardo da Vinci (16th Century, A.D.)

- **First engineering of vision** happened around Renaissance

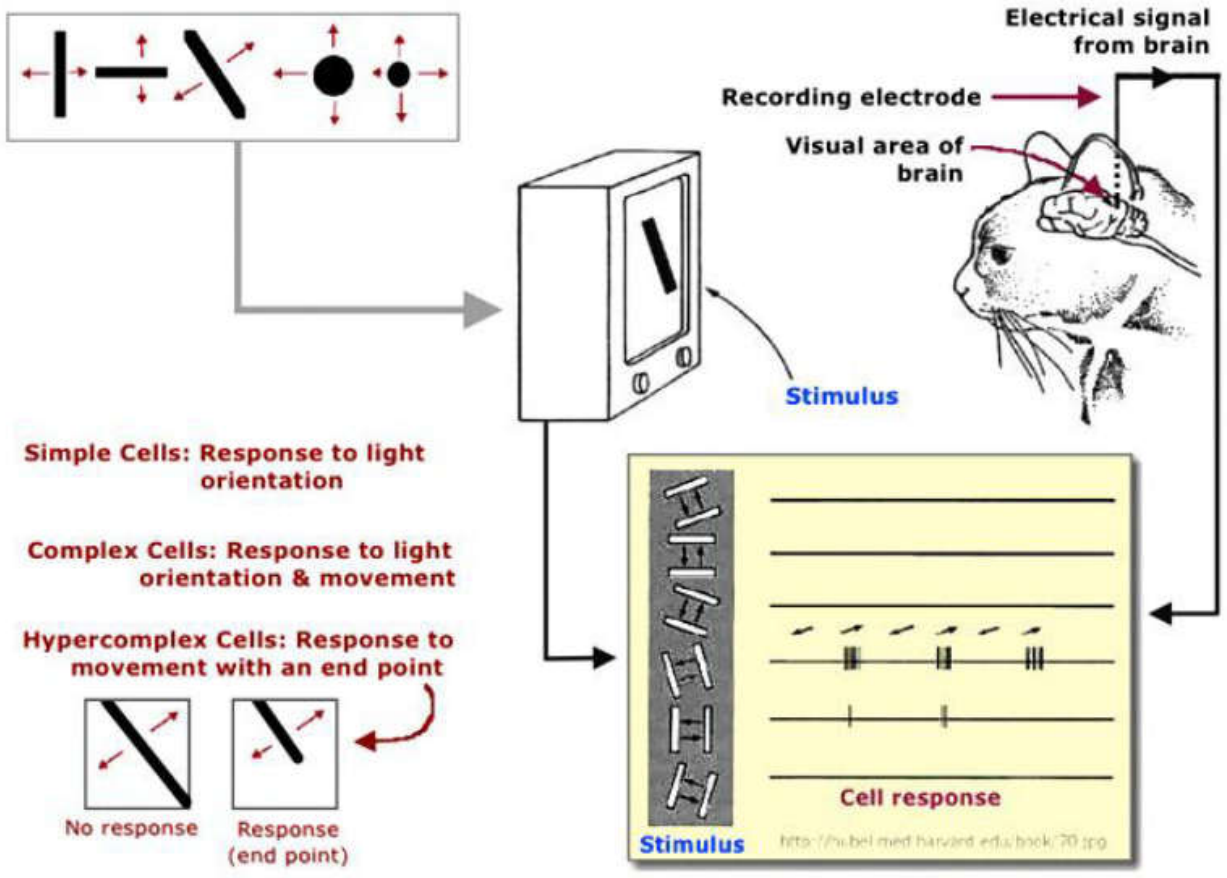
- Capture information of real-world image



Slide from CS231n

Hubel & Wiesel, 1959

- Visual cortex is the earliest process stage for visual processing (back of brain)



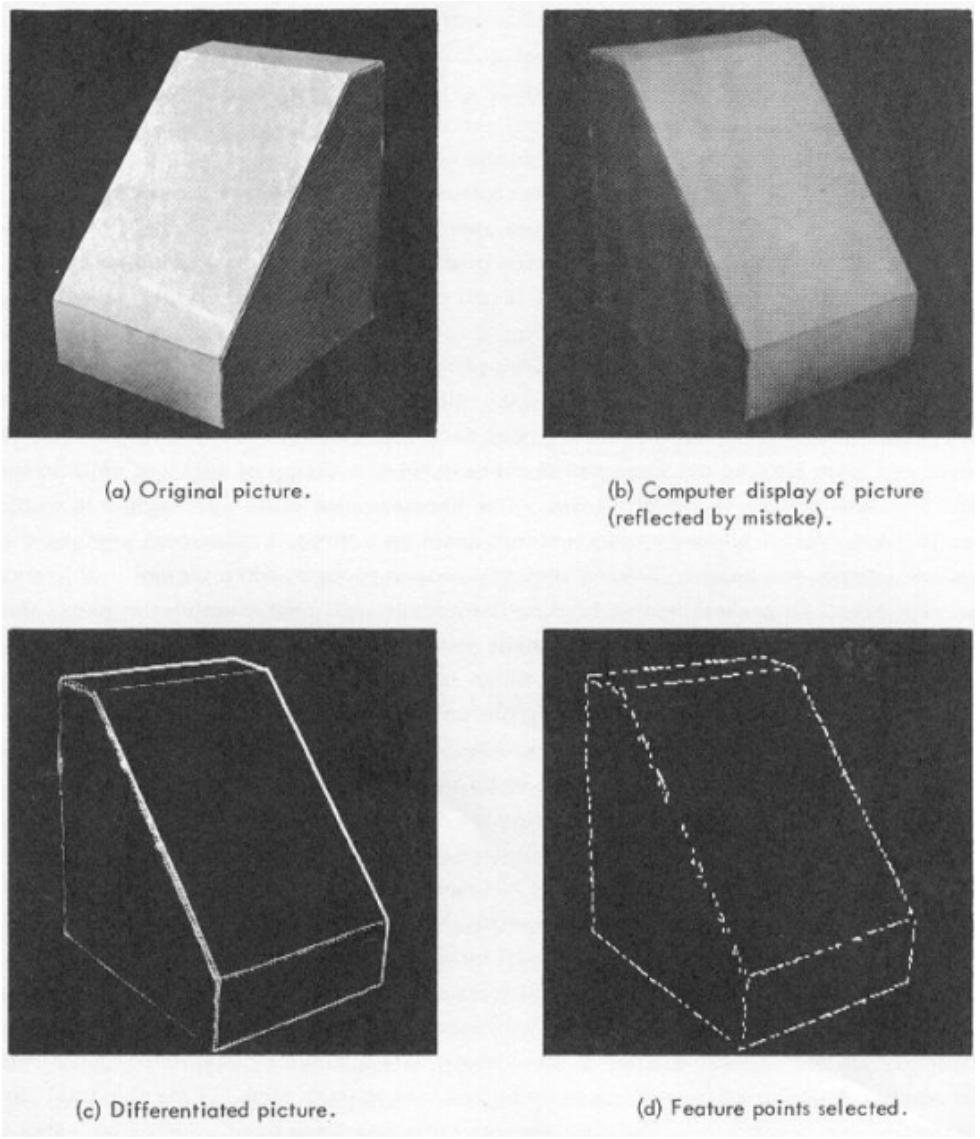
- The beginning of visual processing is not a holistic **fish or mouse**, the beginning of visual processing is simple structures of the world, edges

Slide from CS231n

Block world

Larry Roberts, 1963 (computer science, PhD student)

- Wrote a PhD dissertation just extracted edges
- It was the first precursor of computer vision PhD thesis
- One of the inventors of the Internet



Slide from CS231n

Summer Vision Project 1966

- It's time to solve vision problem

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

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

THE SUMMER VISION PROJECT

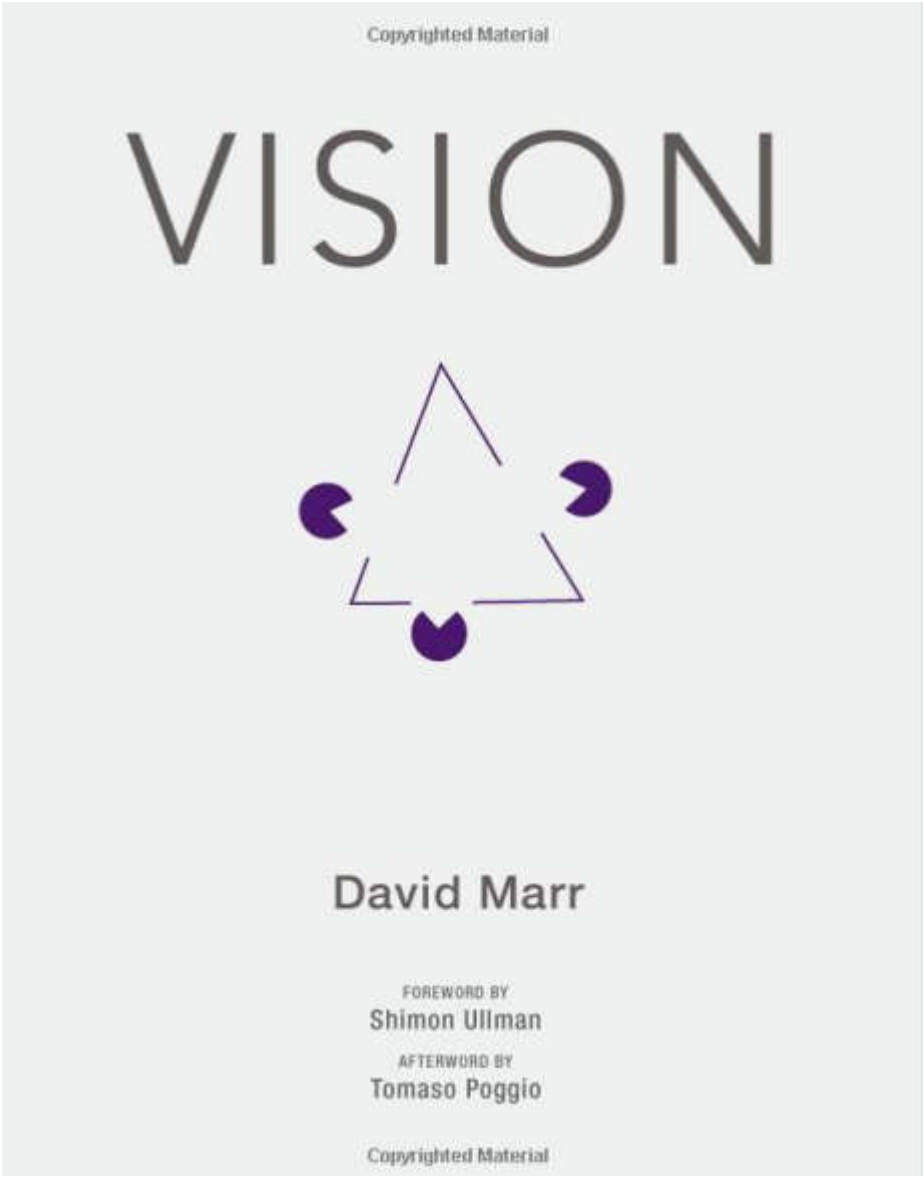
Seymour Papert

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".

Slide from CS231n

David Marr, 1970

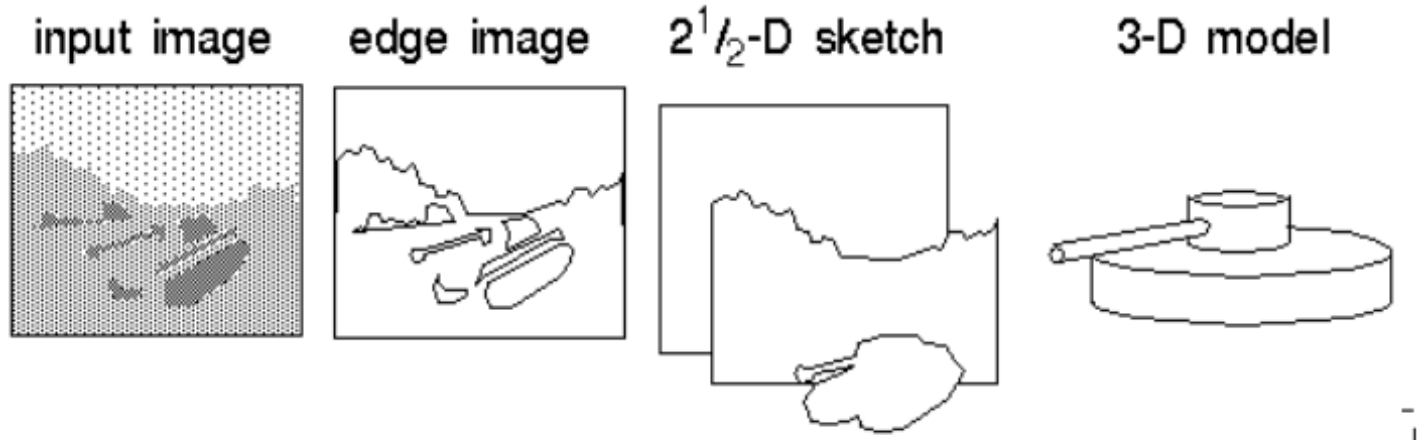
- Wrote a very influential book called “vision”
- Gave a lot of insights from computer science
- Vision is hierarchical (beginning of deep learning)



Slide from CS231n

Stages of Visual Representation, David Marr, 1970s

- Build a hierarchical model

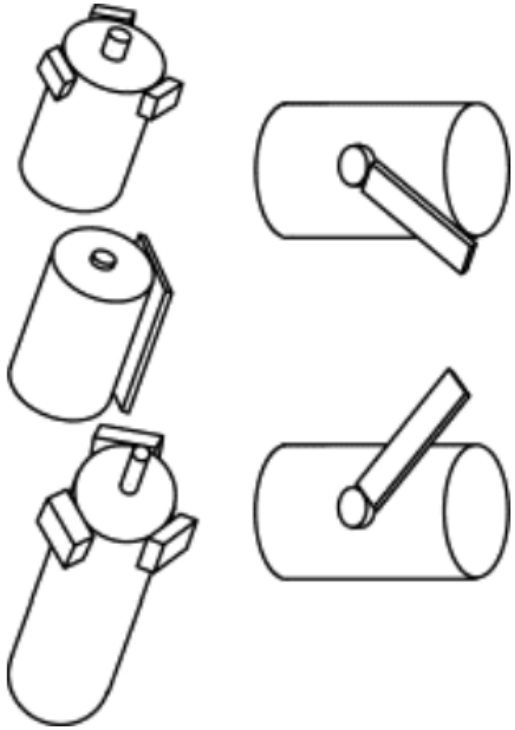


Input Image	Primal Sketch	2 1/2-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs,edges, bars, ends, virtual lines, groups, curves boundaries.	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organised in terms of surface and volumetric primitives

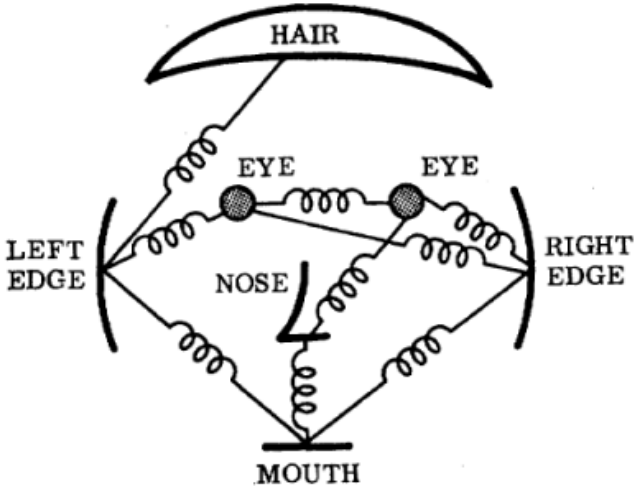
Slide from CS231n

Generalized Cylinder & Pictorial Structure

- The world is composed of simple **cylinder blocks**
- Any real world object is combination of these simple shapes given a particular viewing angle
- A person's head is made of some parts, and they are connected by spring for some deformation



Generalized Cylinder
- Brooks & Binford, 1979

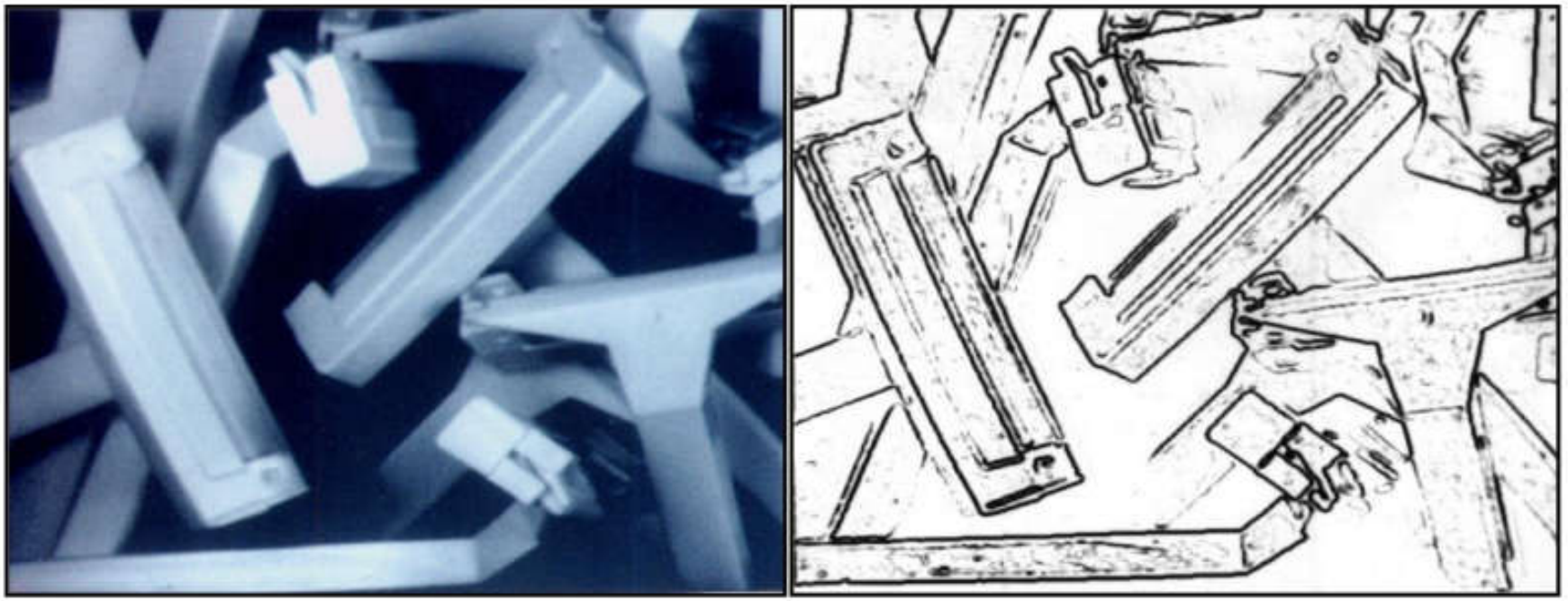


Pictorial Structure
- Fischler and Elschalager, 1973

Slide from CS231n

David Lowe, 1987

- Recognizing real world object
- Shaving razors
- Using **edges and simple shapes formed by edges** to recognize object



These are ancient works

- **Black and white images**
- **Synthetic images**

Slide from CS231n

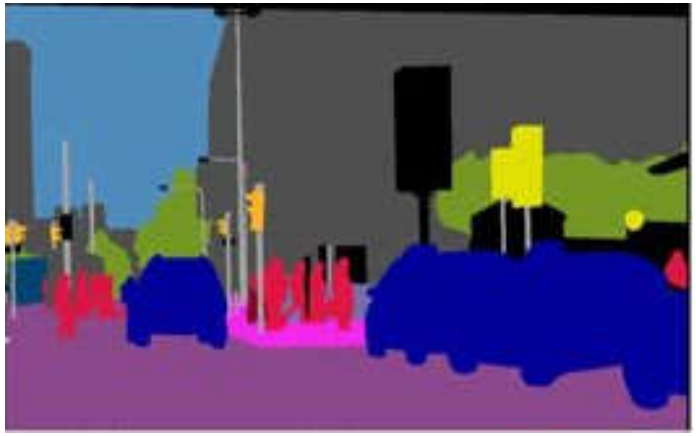
Shi & Malik, 1997

- **Perceptual grouping** is one of the most important problem in vision
- The fundamental problem **is still not solved** in compute vision

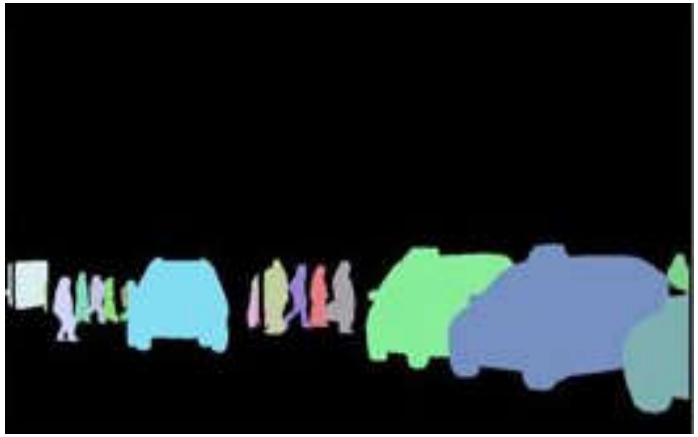
Slide from CS231n

Segmentation

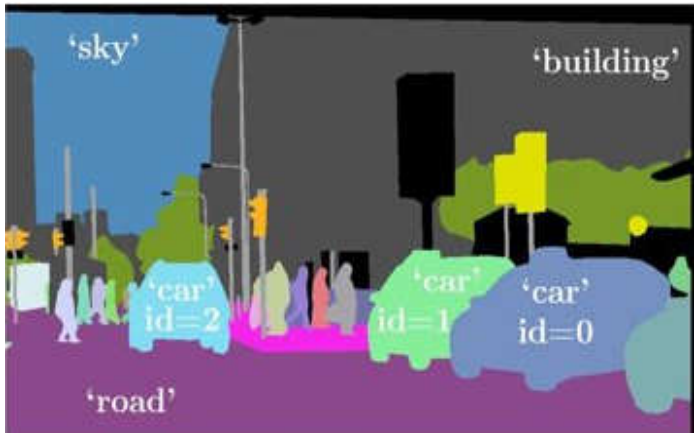
Input image



Semantic segmentation



Instance segmentation

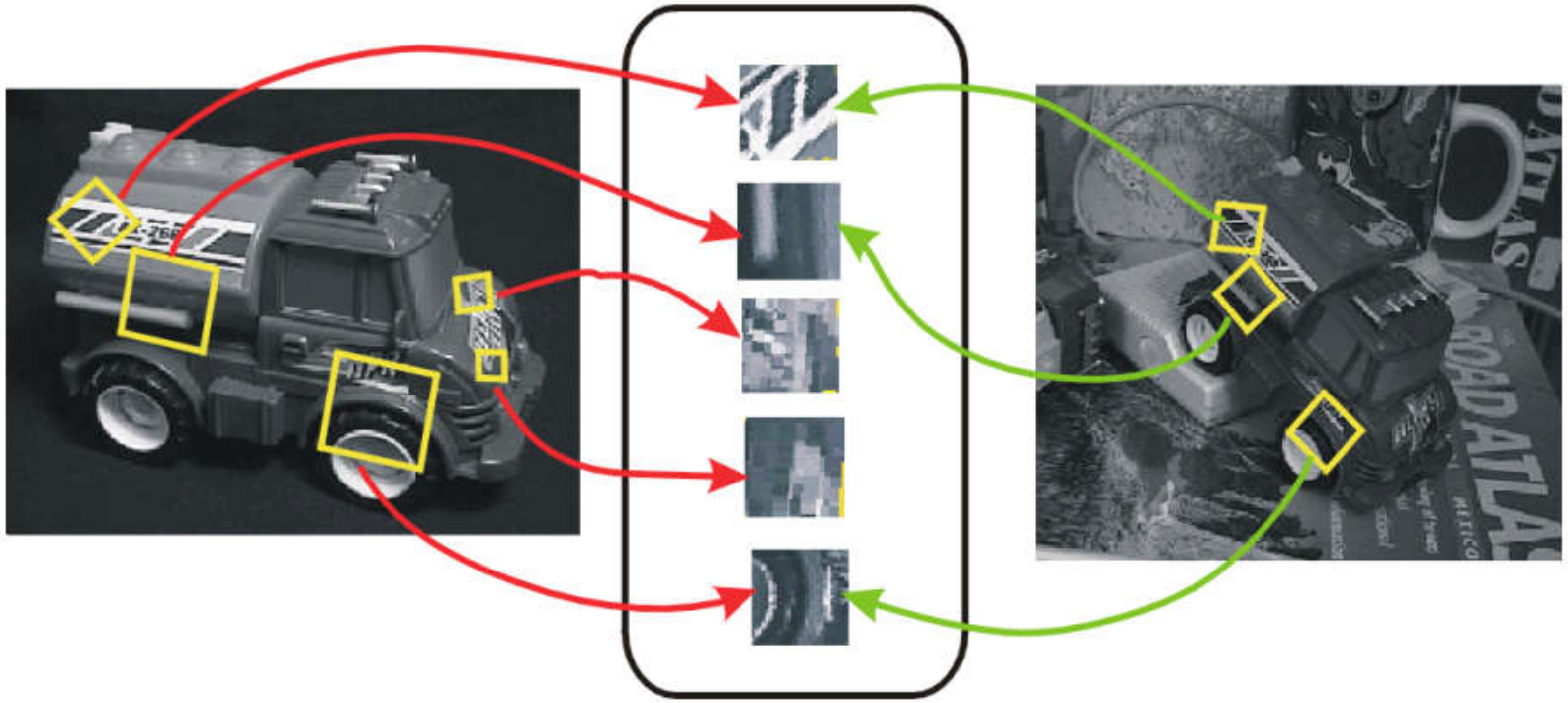


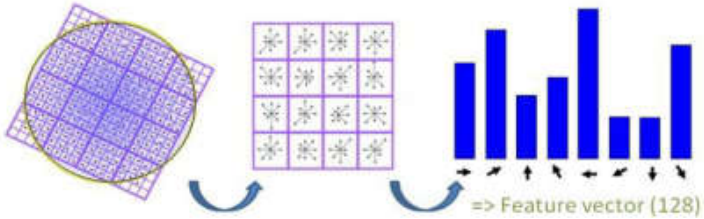
Panoptic segmentation

Alexander Kirillov, **Kaiming He**, **Ross Girshick**, Carsten Rother, Piotr Dollar, "Panoptic segmentation," arxiv:1801.00868

David Lowe, 1999

- Scale-Invariant Feature Transform **(SIFT)**
- An important work started to **focus on features**
- Learn important features from one object and recognize the object in any angle

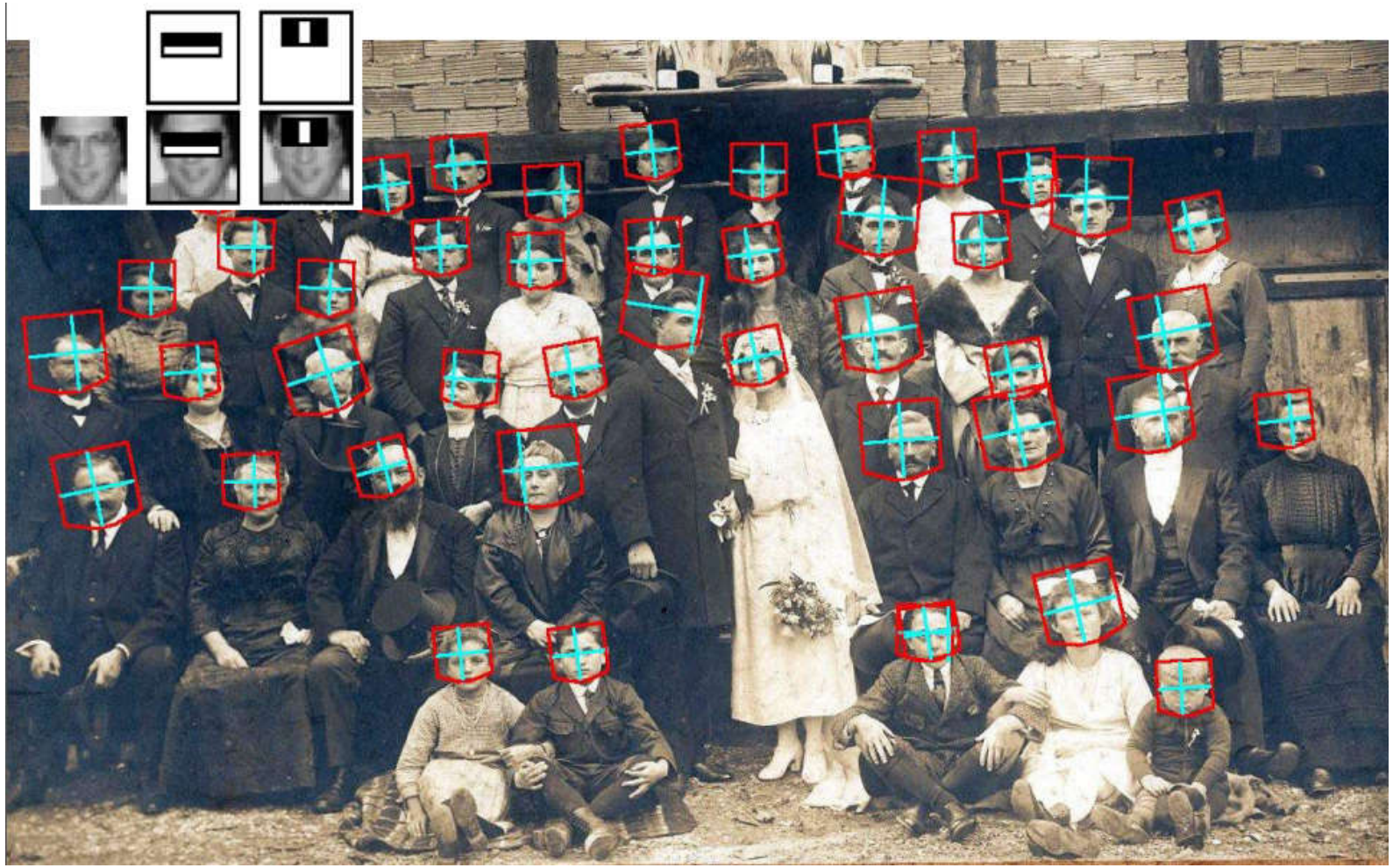




Slide from CS231n

Face Detection, Viola & Jones, 2001

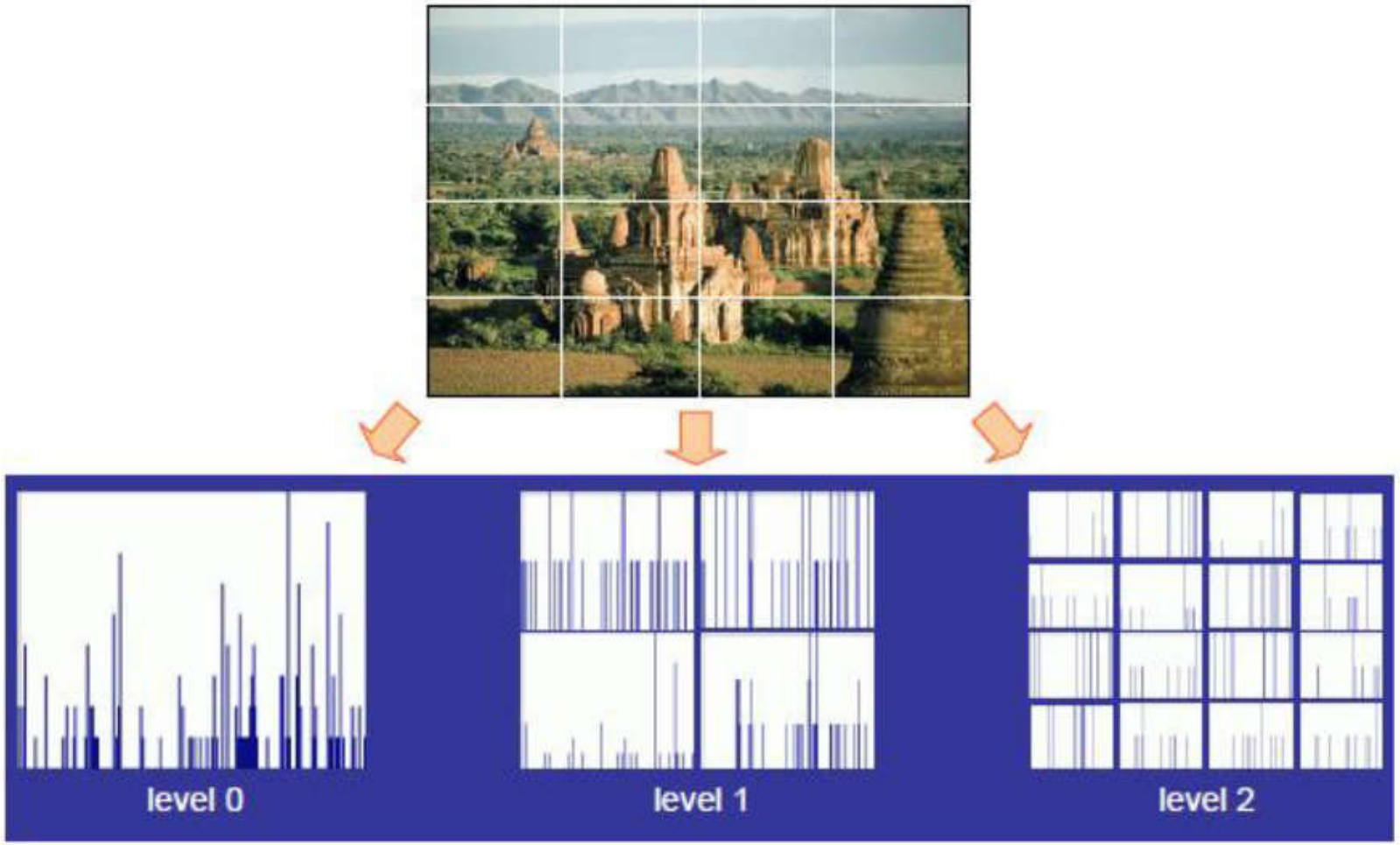
- **First smart digital camera** by Fujifilm in 2006
- First digital camera has a **face detector**
- One of the first computer vision work that deployed on PC can work in **real-time**



Slide from CS231n

Lazebnik, Schmid & Ponce, 2006

- Scene recognition
- Hand-craft features were designed by brilliant engineers

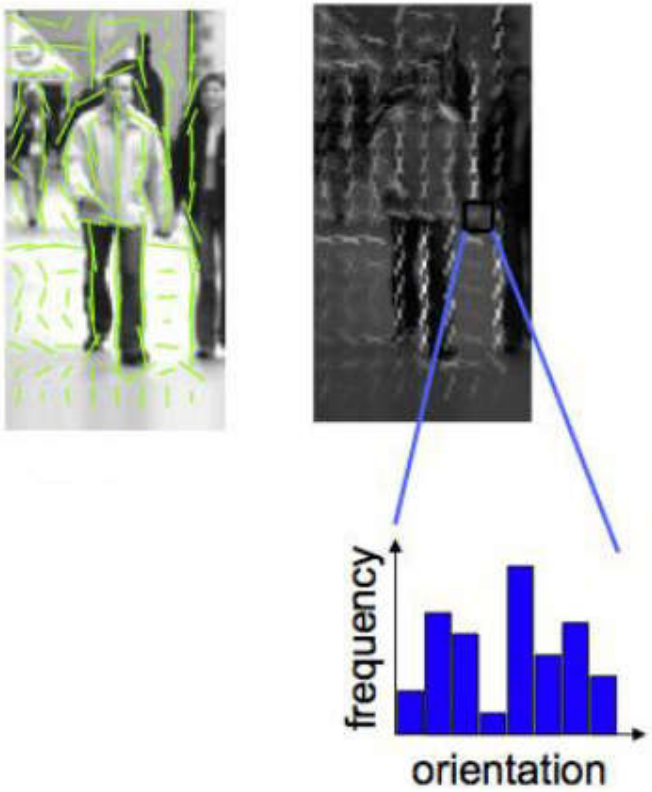


Spatial Pyramid matching

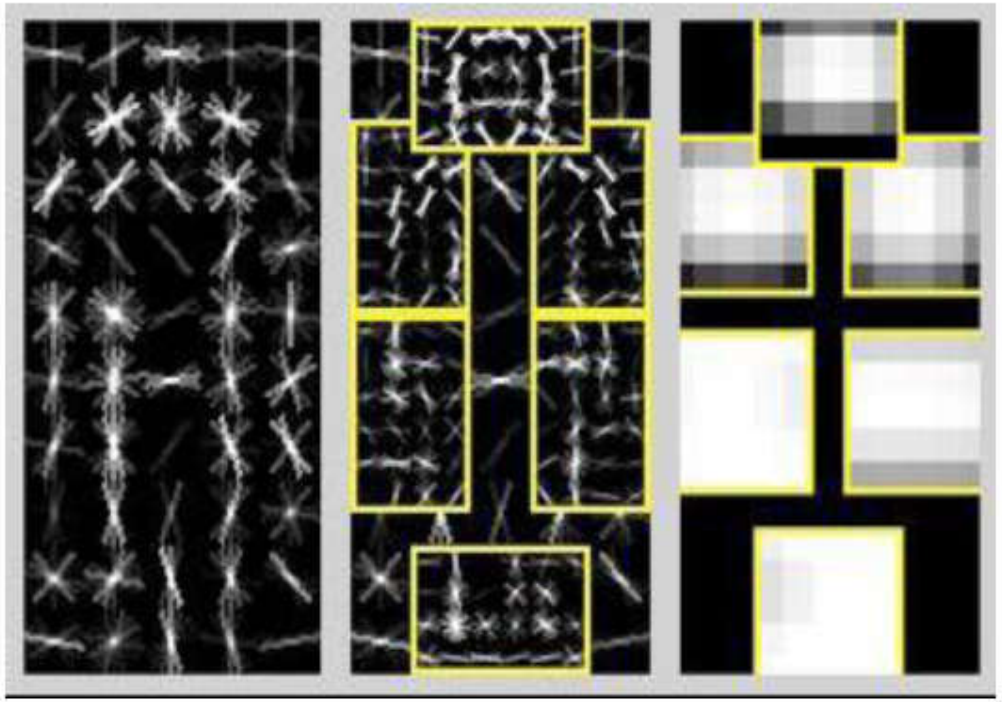
Slide from CS231n

HOG 2005 & DPM 2009

- Learn parts of the object and learn how they configure each other in space
- About 2009~2010, **computer vision community** found that **they need a benchmark**



Histogram of Gradients (HoG)
Dala & Triggs, 2005

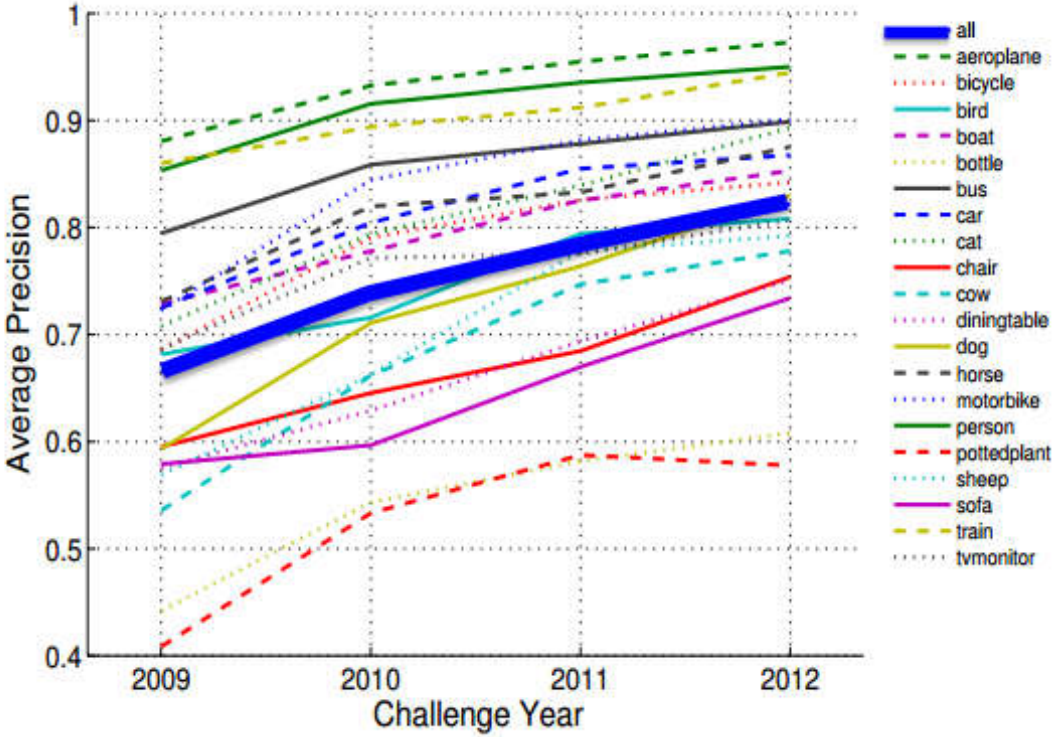


Deformable Part Model (DPM)
Felzenswalb, McAllester,
Ramanan, 2009

Slide from CS231n

PASCAL VOC (Visual Object Challenge)

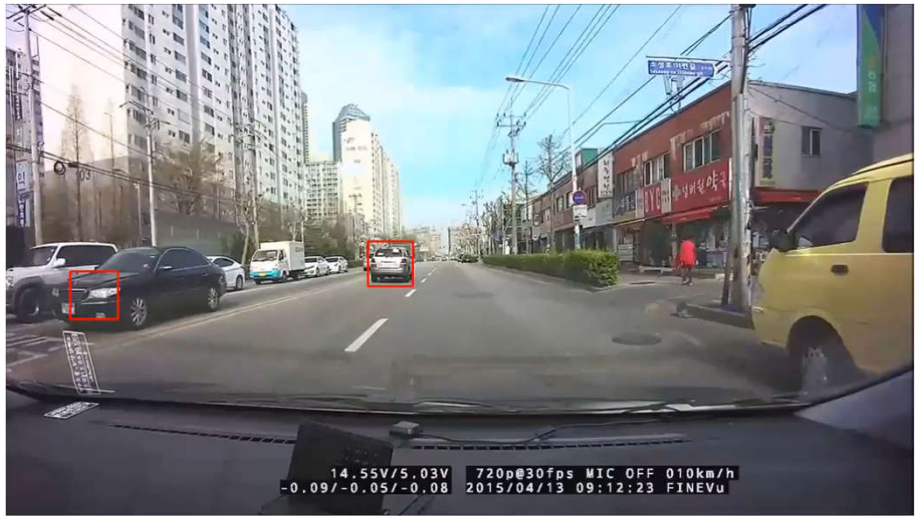
- **Object recognition** benchmark by European
- **Tens of thousands** of images for **20 object categories**
- Through the years, performance just keeps increasing



[Everingham et al. 2006-2012]

Slide from CS231n

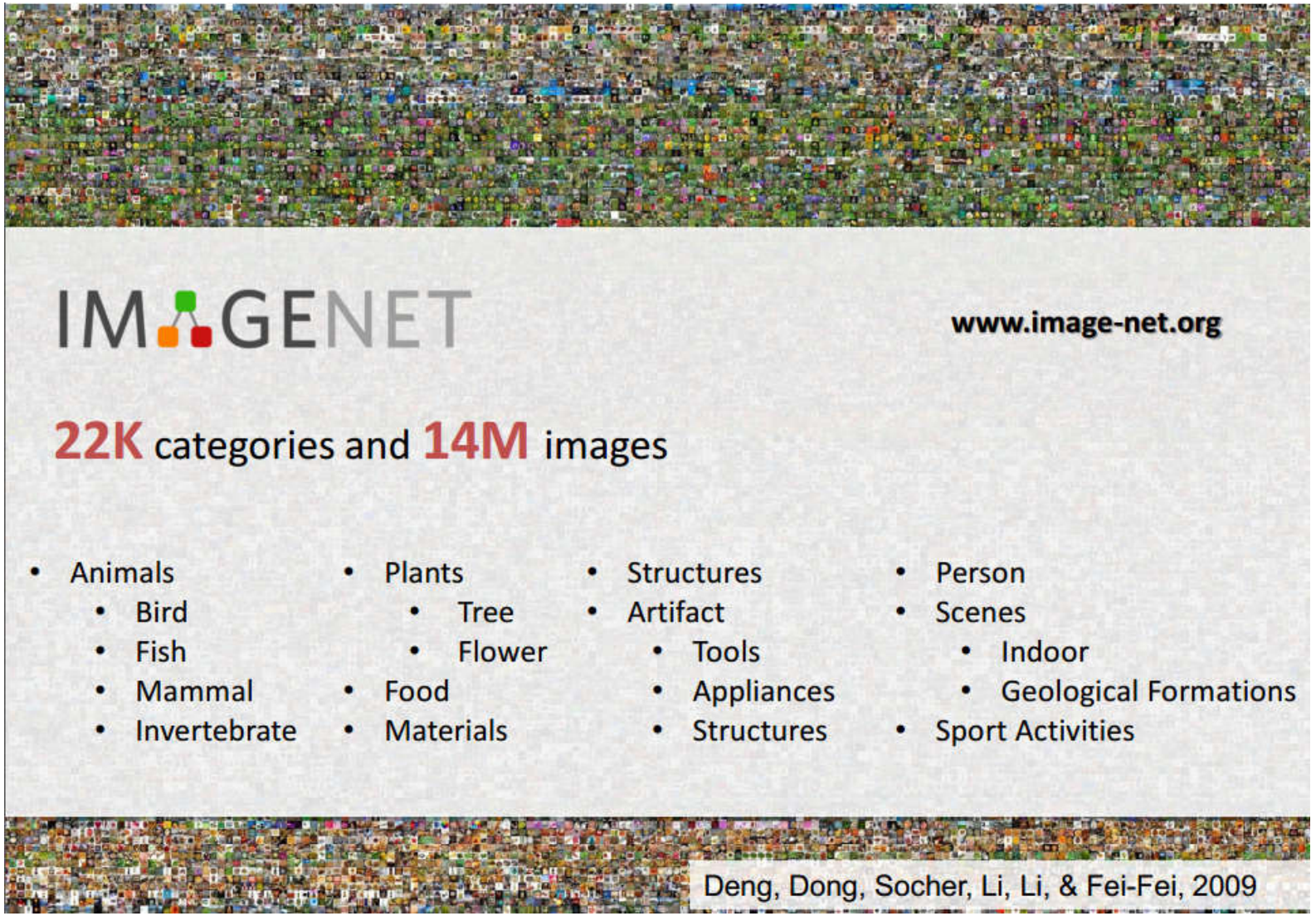
VD/PD using Conventional Vision



VD: vehicle detection PD: pedestrian detection

IMGENET Challenge

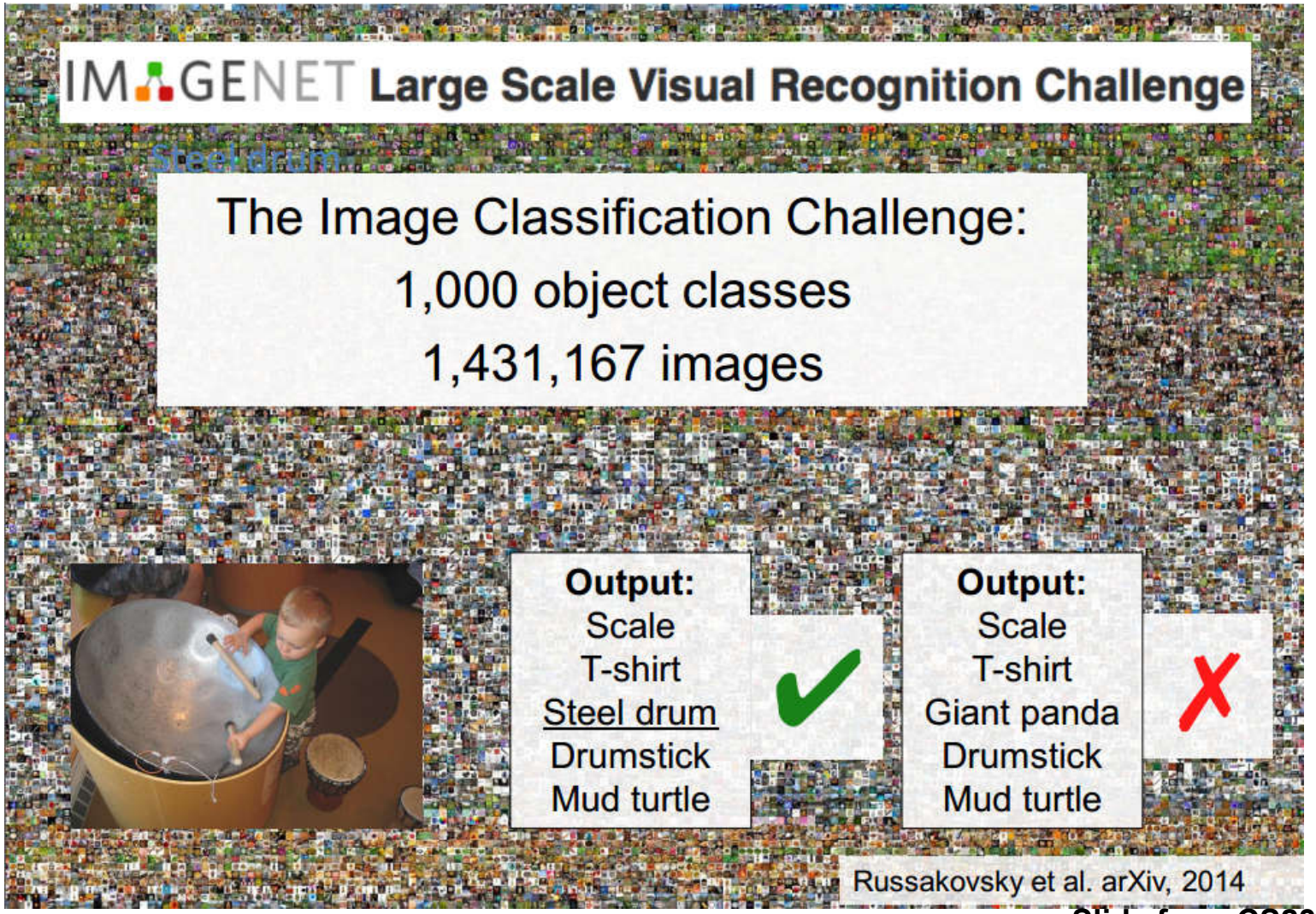
- Amazon Mechanical Turk platform



Slide from CS231n

IMGENET Challenge

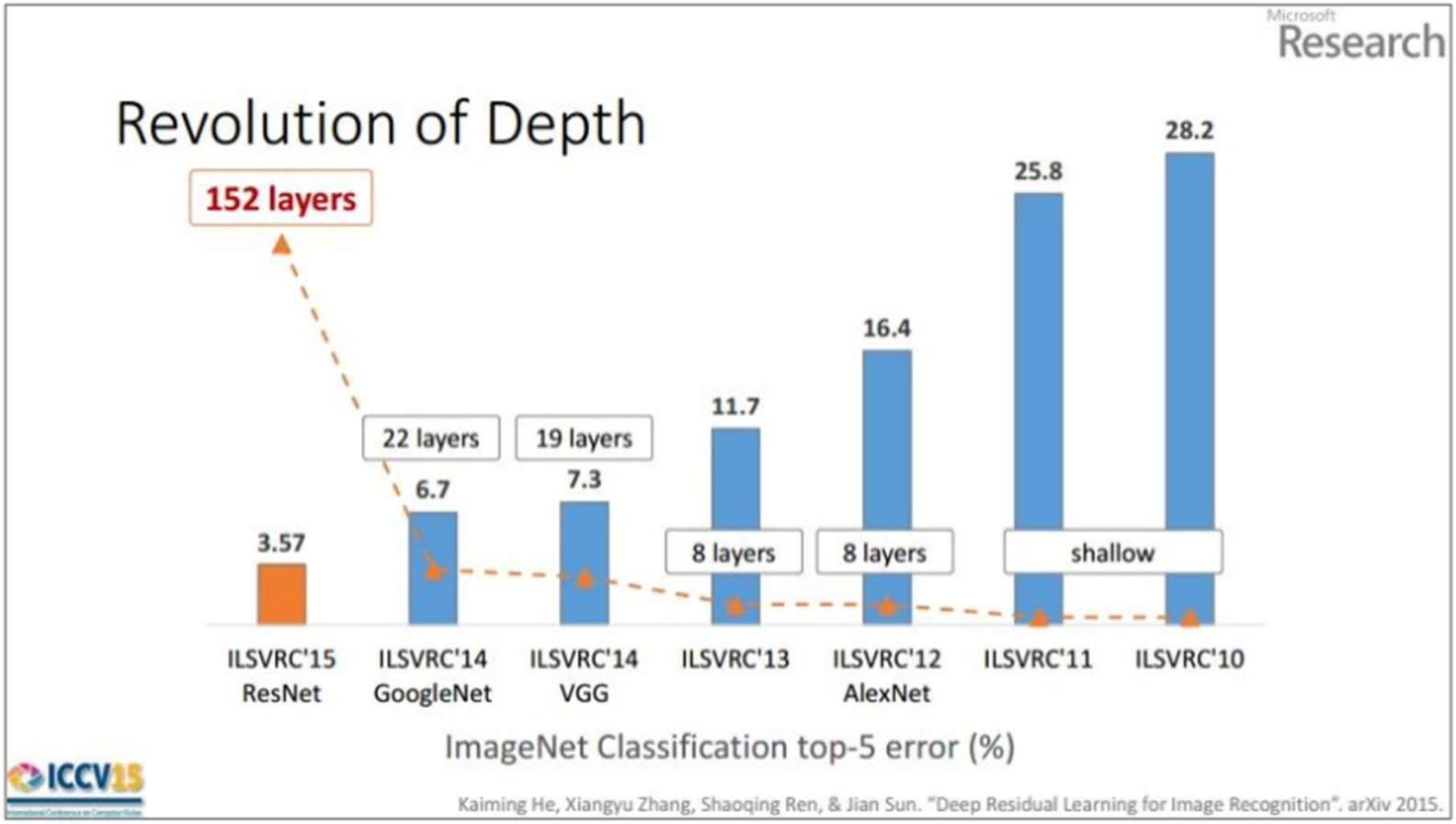
- **1000 classes**, and about **1.5 million images**
- **Olympics** of computer vision challenging



Slide from CS231n

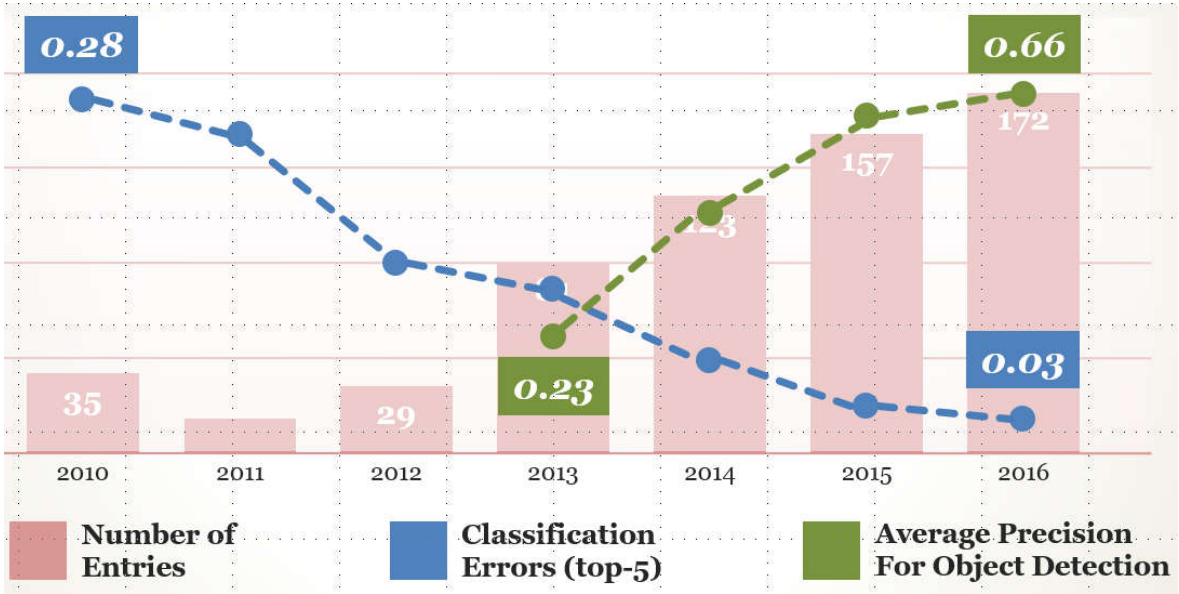
IMGENET Challenge

- **2012** is the year that the architecture of the winner was a convolutional neural network model
- This is the beginning of **deep learning revolution**



Source: Lecture slide 13 of "CS231n : Convolutional Neural Networks for Visual Recognition"

Visual Recognition using DL



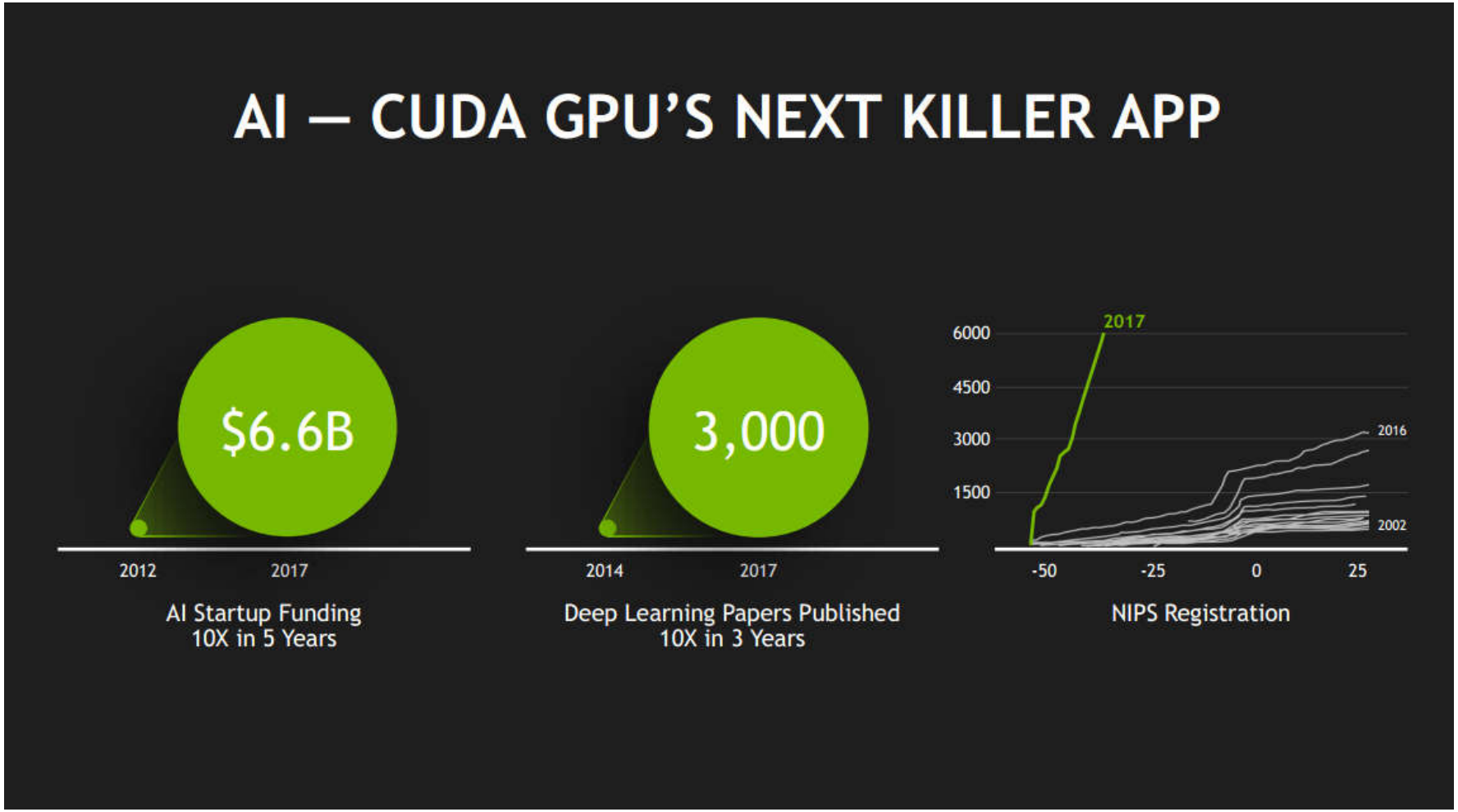
Relative Confusion	A1	A2
Human succeeds, GoogLeNet succeeds	1352	219
Human succeeds, GoogLeNet fails	72	8
Human fails, GoogLeNet succeeds	46	24
Human fails, GoogLeNet fails	30	7
Total number of images	1500	258
Estimated GoogLeNet classification error	6.8%	5.8%
Estimated human classification error	5.1%	12.0%

Table 9 Human classification results on the ILSVRC2012-2014 classification test set, for two expert annotators A1 and A2. We report top-5 classification error.

Comparison between human and GooleNet

Source: Li Fei-Fei et. al., “ImageNet Large Scale Visual Recognition Challenge”, 2014

What's the Catalyst for One Research Community?



- NIPS2018 received a record-breaking 4856 submissions, of which 1011 submissions were accepted

NIPS: Neural Information Processing System

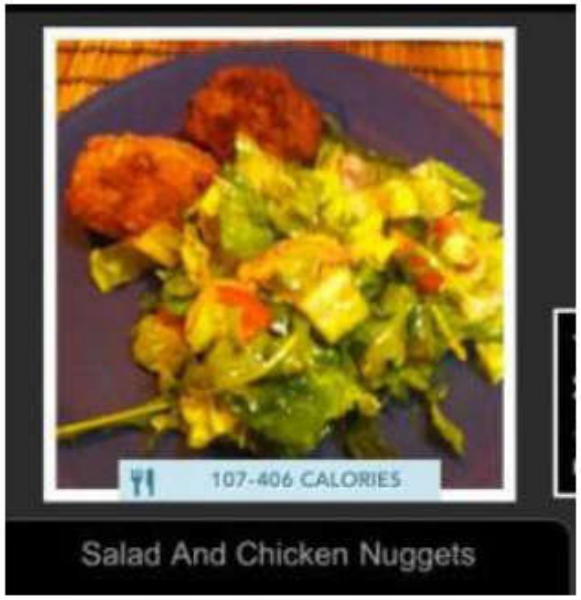
Deep Learning Class Overview

Class Overview

- This class focuses on one of the most important problems of visual recognition
  - **Image Classification**

- Visual recognition **is not just** image classification
  - Segmentation, object detection, 3D modeling, and image captioning, etc.

Image Classification App



Slide from CS231n

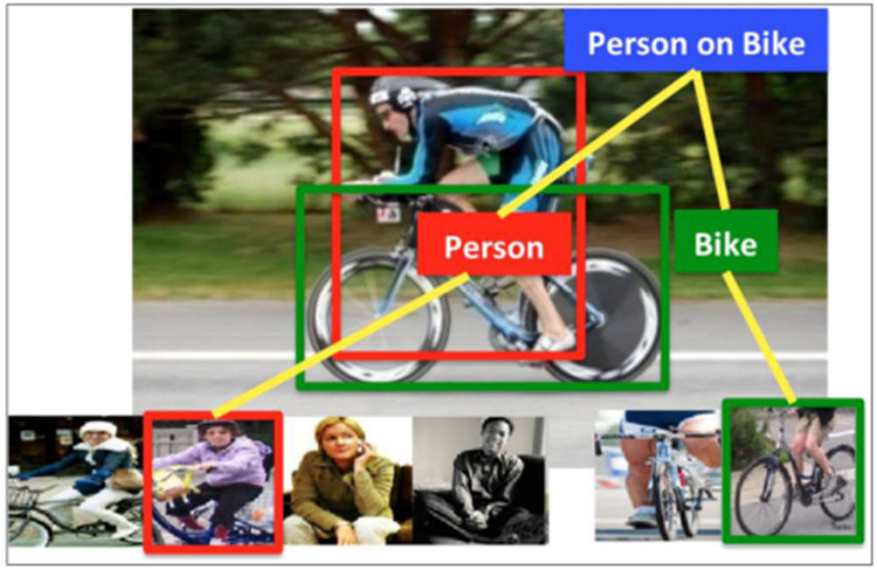
Class Overview

- There is a number of visual recognition problems that are related to **image classification**, such as **object detection**, and **image captioning**

- **Convolutional Neural Network (CNN)** has become **an important tool** for object recognition

Class Overview

- **Image classification**
- Object detection
- Action classification
- Image captioning
- Segmentation
- Video classification



Slide from CS231n

IMAGENET Challenge

Year 2010

NEC-UIUC

Dense grid descriptor: HOG, LBP

Coding: local coordinate, super-vector

Pooling, SPM

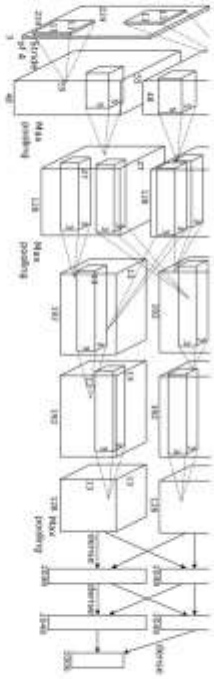
Linear SVM

[Lin CVPR 2011]

Hierarchical

Year 2012

SuperVision



[Krizhevsky NIPS 2012]

8 layers

Year 2014

GoogLeNet

[Szegedy arxiv 2014]

22 layers

VGG

[Simonyan arxiv 2014]

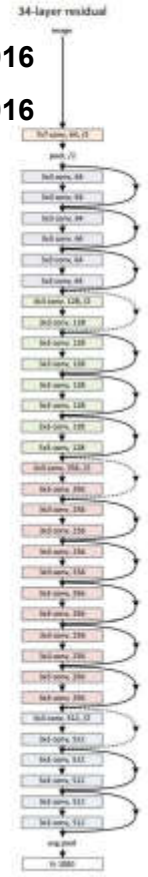
19 layers

Year 2015

MSRA

CVPR 2016

ECCV 2016

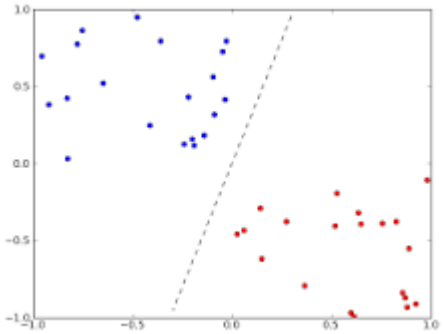


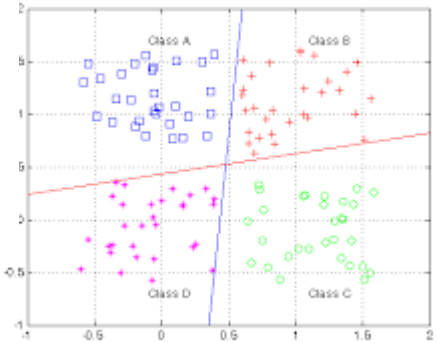
152 layers

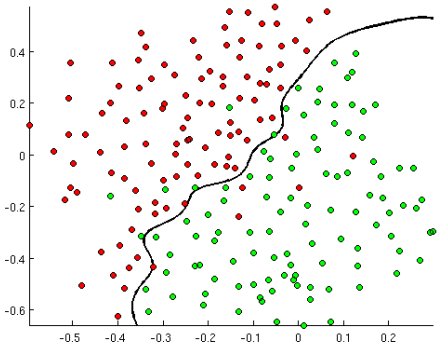
Slide from CS231n

Machine Learning 101

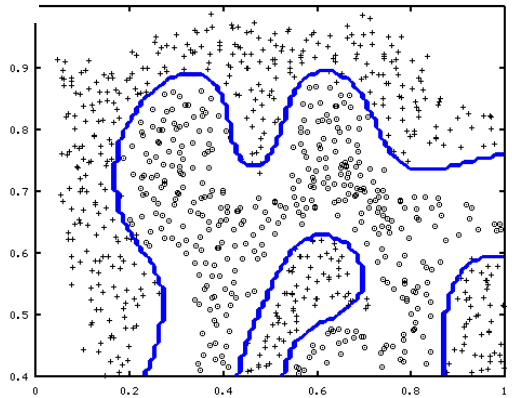
- Depth vs. Nonlinearity





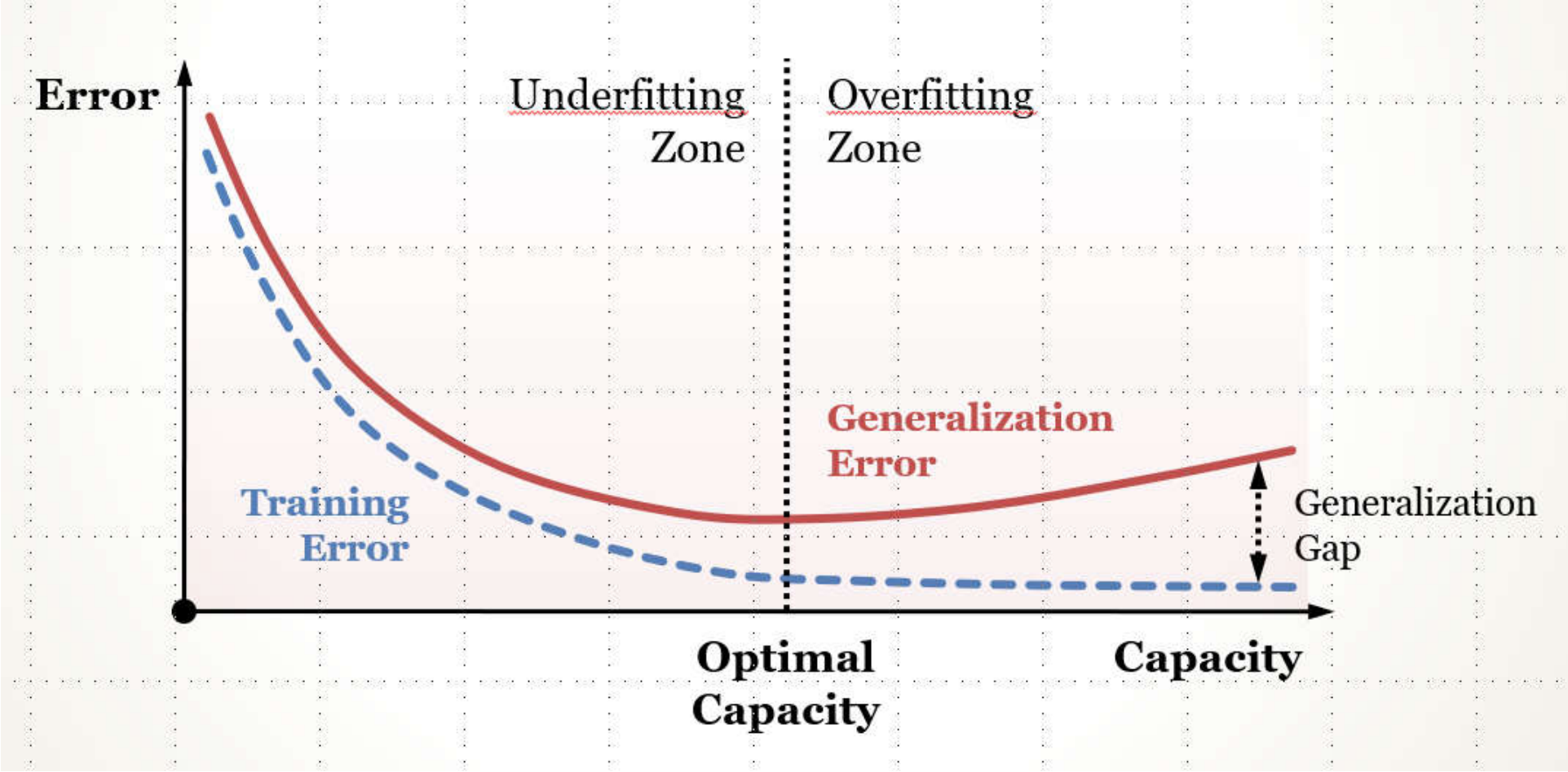


Depth (# of layers)



Machine Learning 101

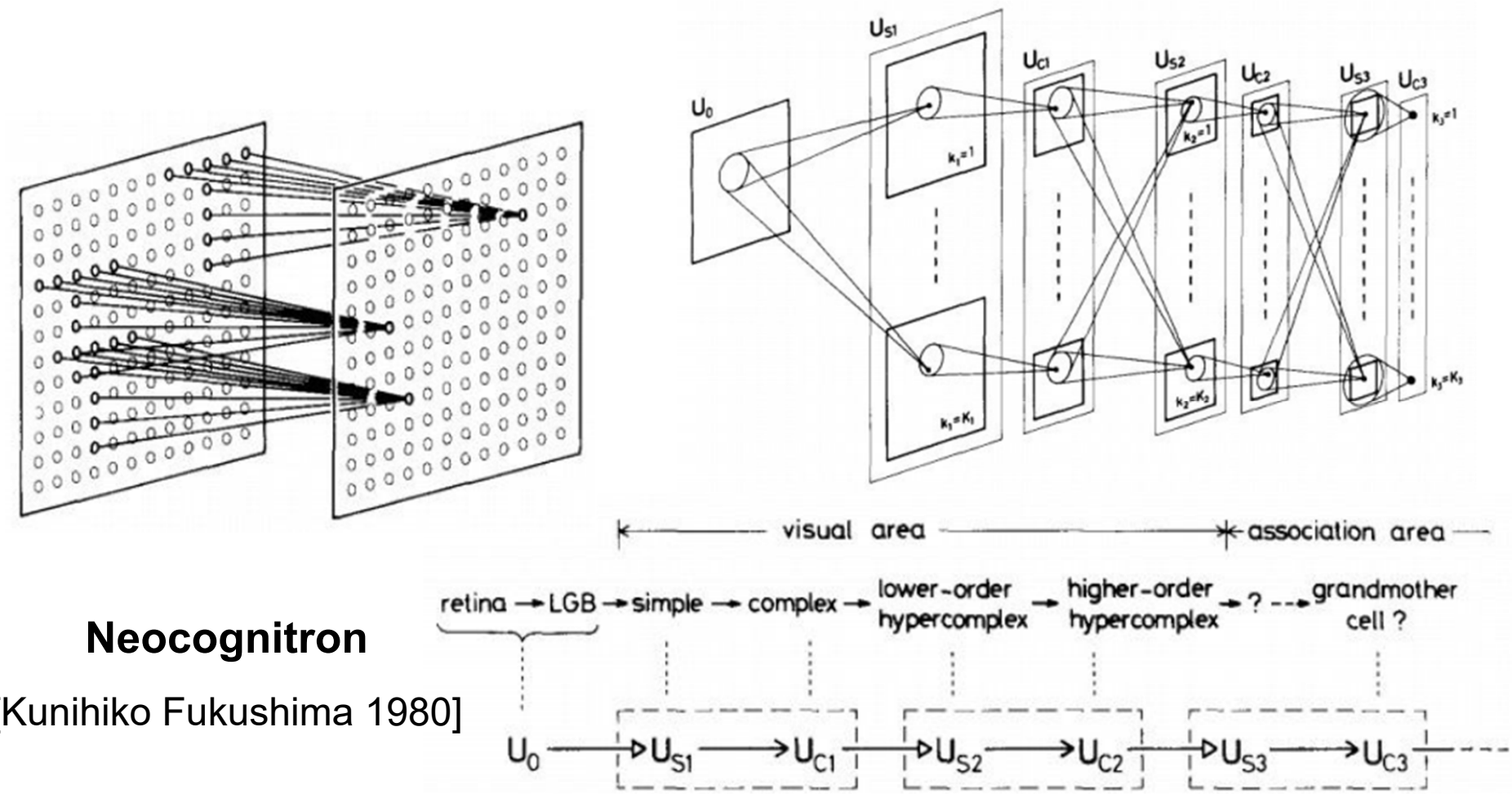
- Complexity, Generalization, Overfitting



Source: ACM Webinar on ImageNet by Fei Fei Li_20170921

Class Overview

- **CNN is not invented overnight**
- Beginning of the convolutional neural network architecture



Neocognitron

[Kunihiko Fukushima 1980]

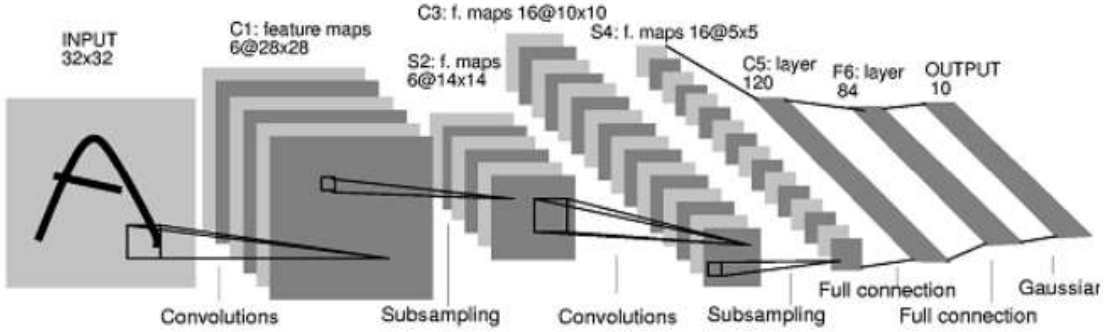
Slide from CS231n

Class Overview

- What's the differences between **LeNet and AlexNet?**

1998

LeCun et al.



# of transistors

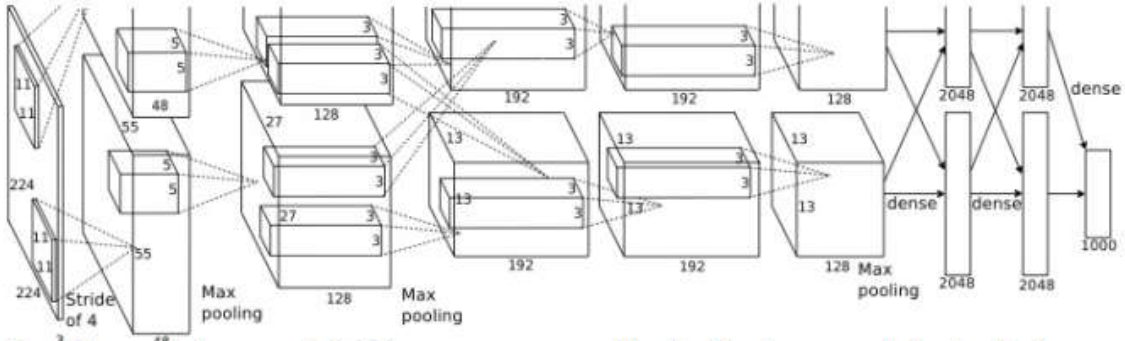
10^6

# 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

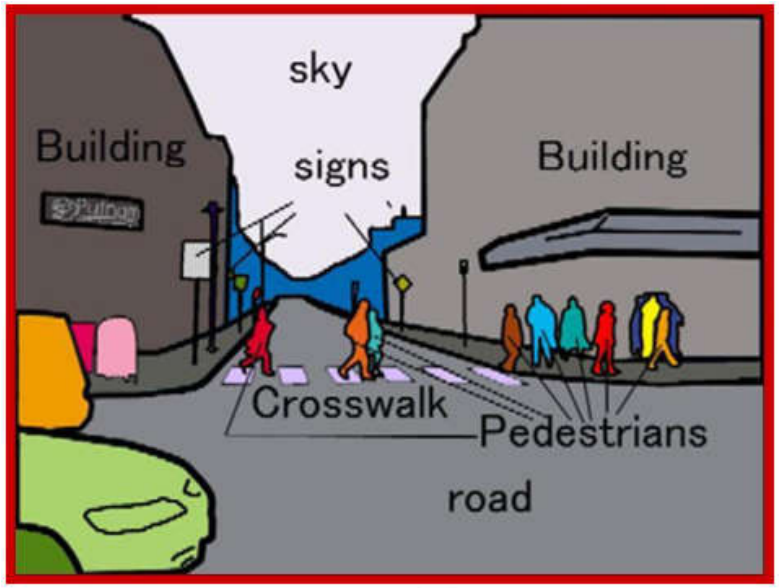
Slide from CS231n

Class Overview

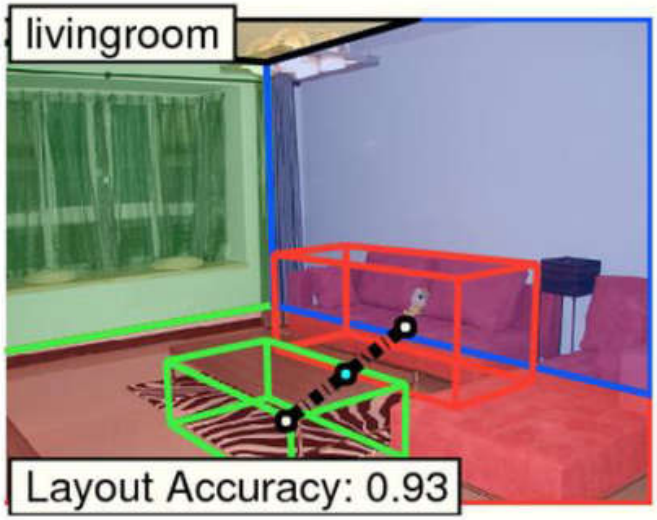
- The **quest for visual intelligence** goes far beyond object recognition…
- We have not done all of the things…
- **IMGENET challenges**
  - Object localization
  - Object detection
  - Object detection from video
  - Scene classification
  - Scene parsing
- We still have a lot of cool problems to solve

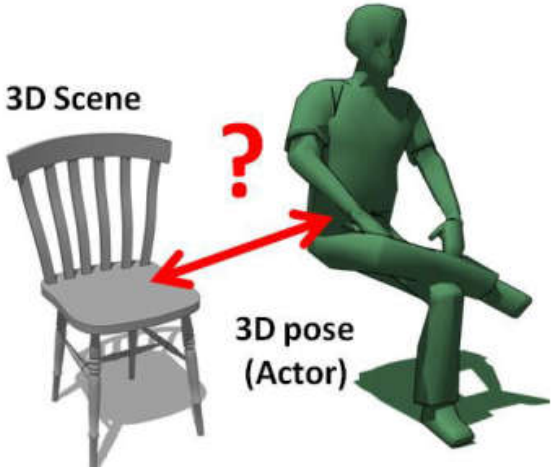
Class Overview

Perceptual grouping

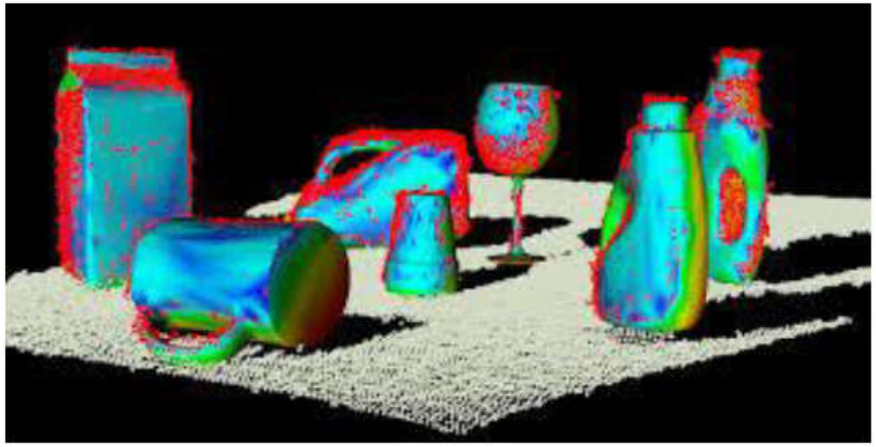


Recognition with 3D





Pose estimation

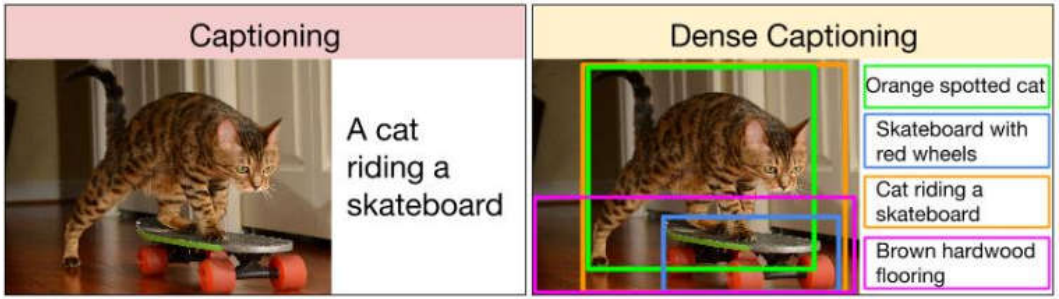


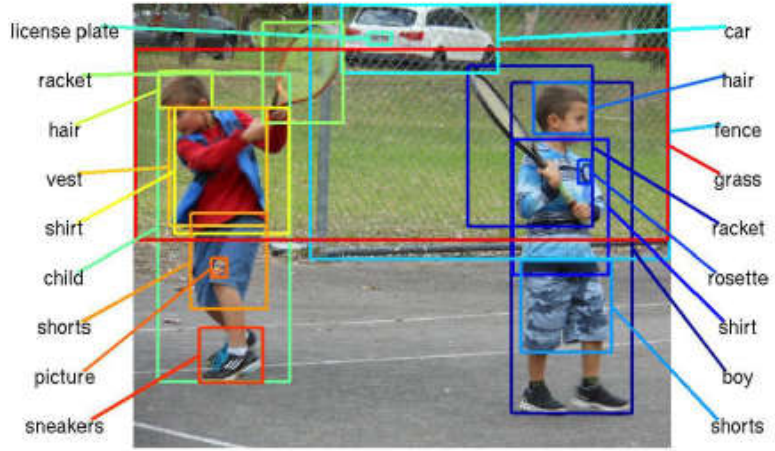
3D reconstruction

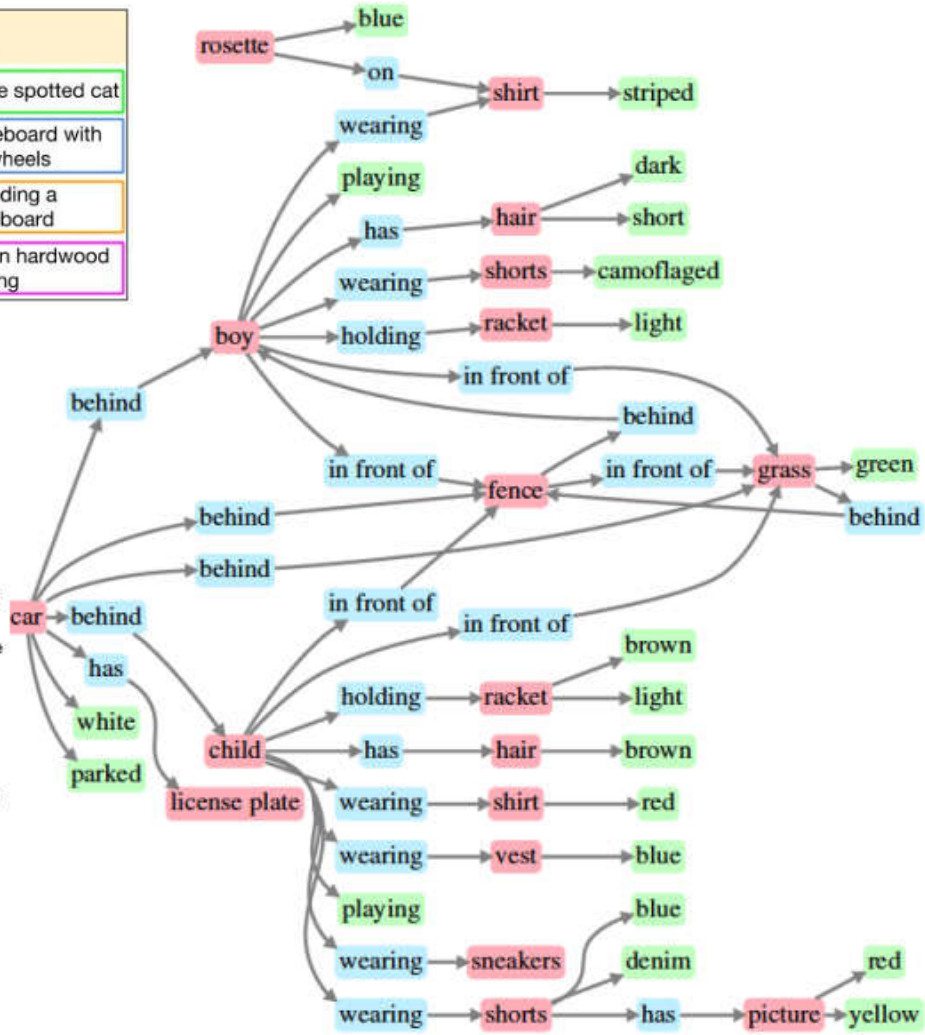
Slide from CS231n

Class Overview

- Far beyond image classification







- Image captioning
- Dese captioning

Slide from CS231n

Fun: Deep Dream





Fun: Neural Art (ECCV2016)

Johson, et al., “Perceptual losses for real-time style transfer and super-resolution,” **ECCV2016**

https://github.com/ChengBinJin/Real-time-style-transfer

Style Transfer

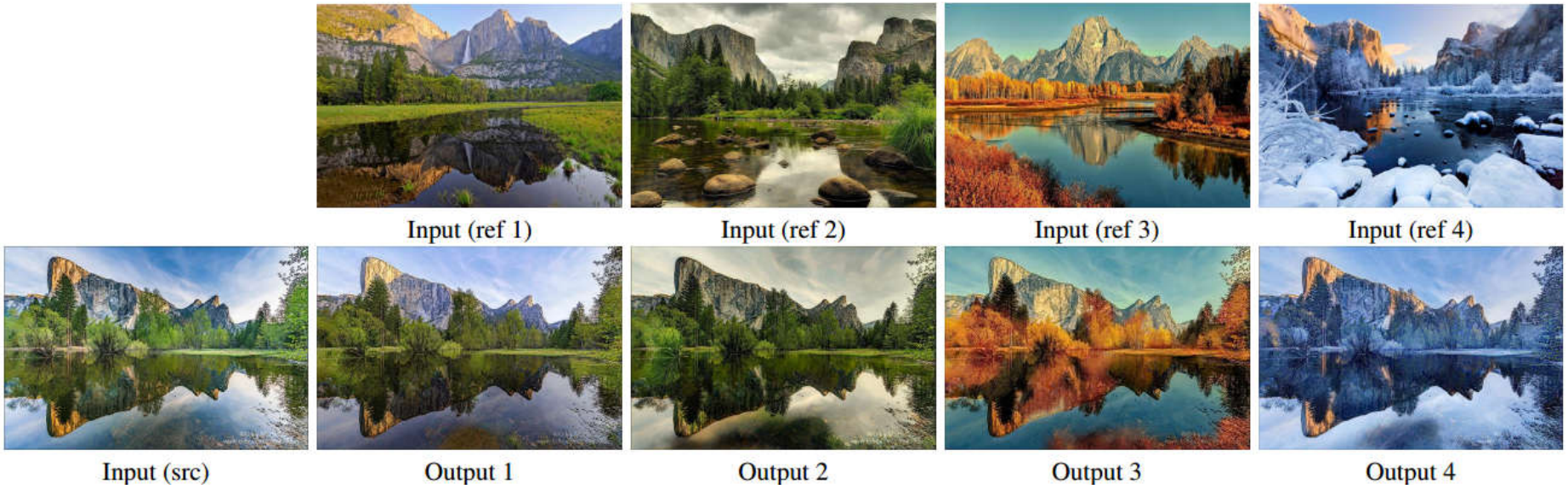
Deep image analogy

Input image

Reference image

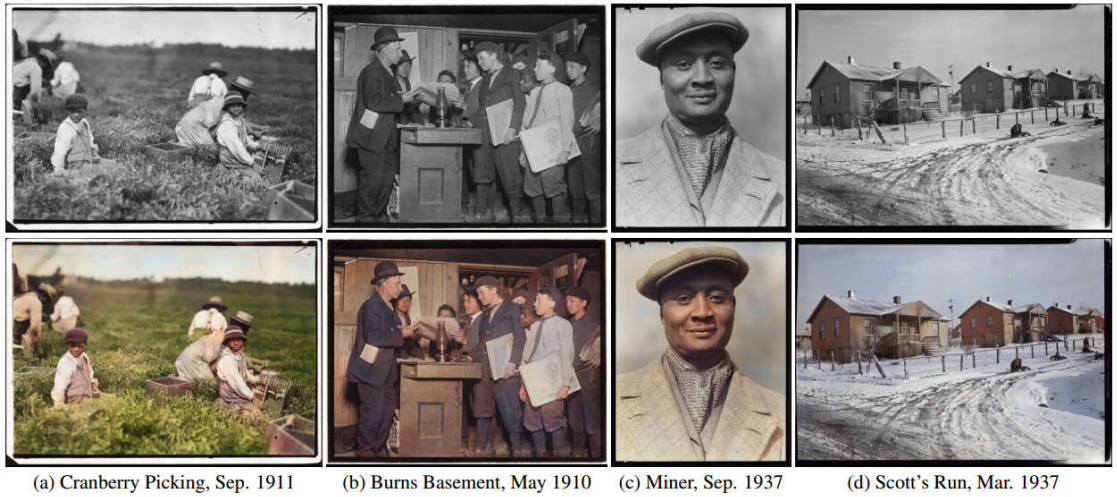
Deep photo style

Deep image analogy



J. Liao, et al., Visual attribute transfer through deep image analogy," **arXiv2017**

Image Colorization



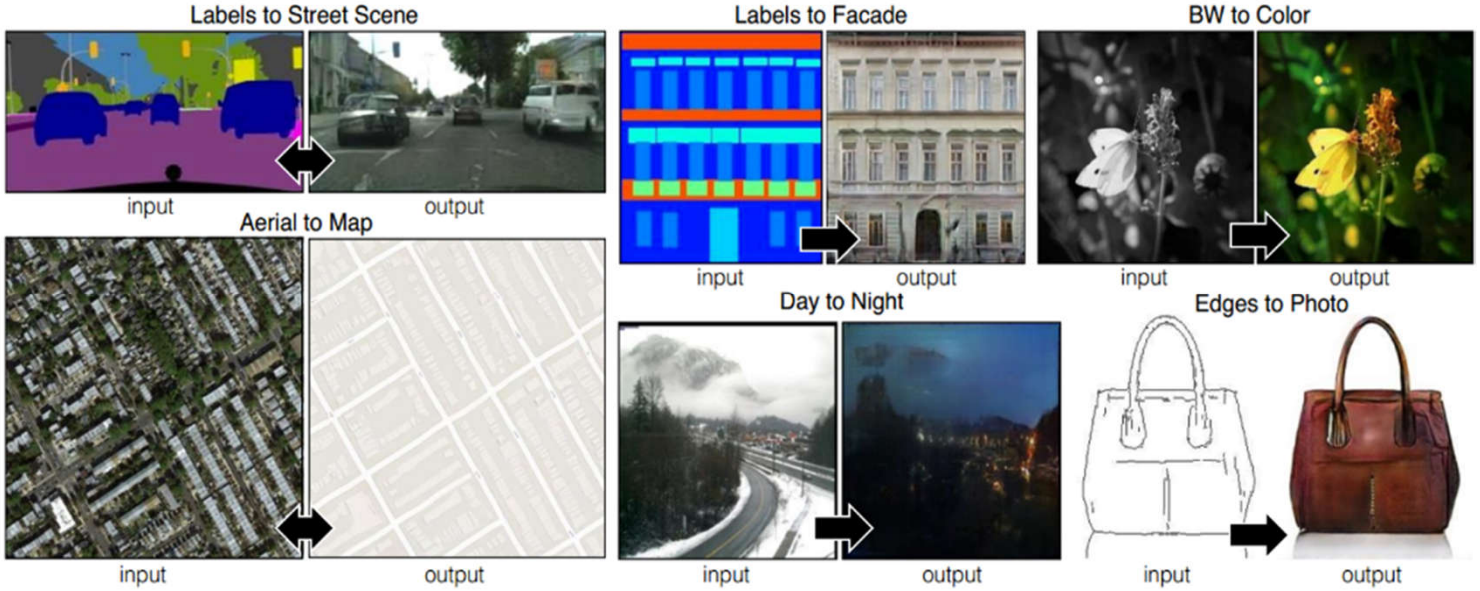
SIGGRAPH: Special Interest Group on Computer Graphics)

R. Zhang, et al., "Real-time image colorization with learned deep priors," **SIGGRAPH2017**

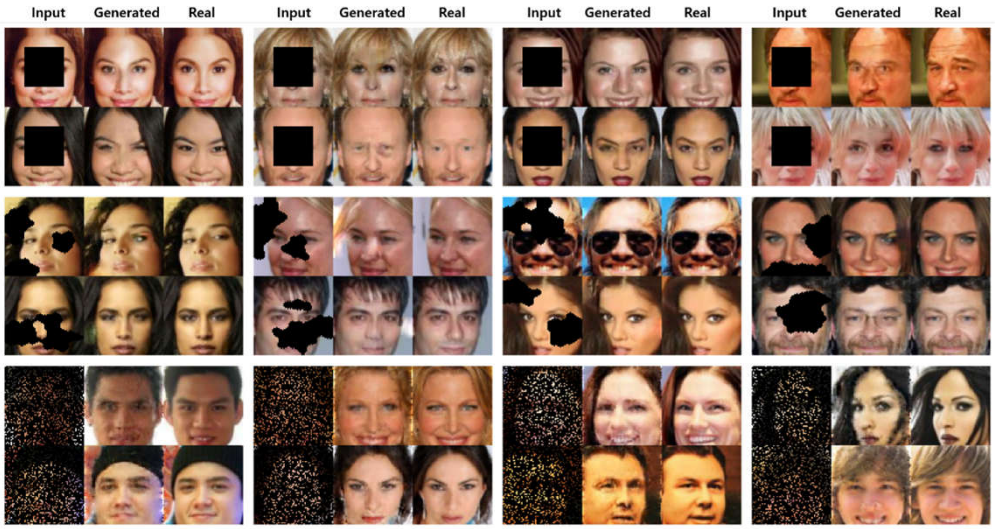
Visual Question and Answering (VQA)

Slide from Richard Socher

Generative Adversarial Networks (GANs)



Pix2px
https://github.com/ChengBinJin/pix2pix-tensorflow

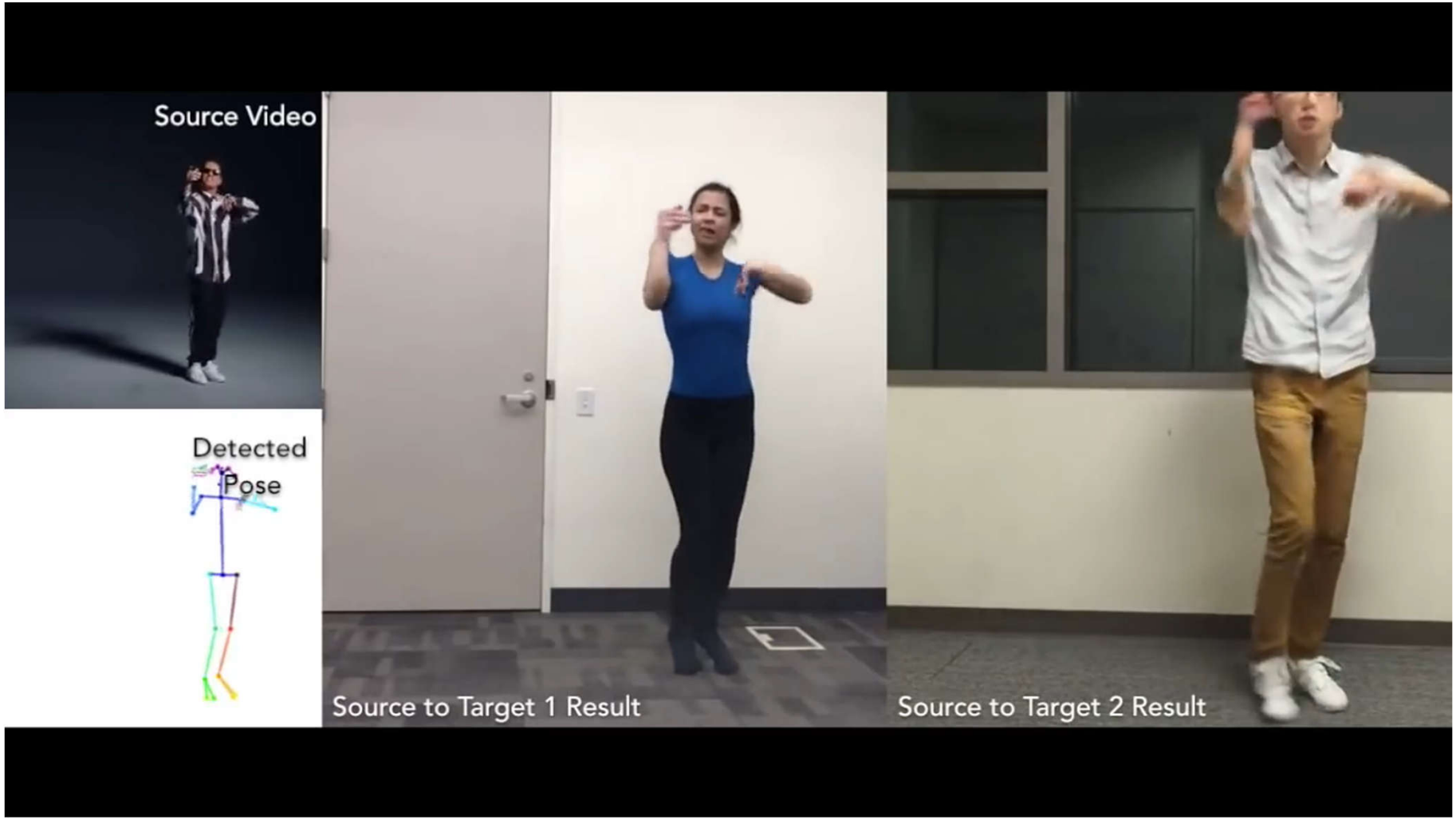


Semantic image inpainting
https://github.com/ChengBinJin/semantic-image-inpainting

CycleGAN

Generative Adversarial Networks (GANs)

https://www.youtube.com/watch?v=PCBTZh41Ris



Caroline Chan, et al., “Everybody dance now,” arXiv1808.07371

Class Overview

Slide from CS231n

Q & A