EE452 Computer Vision (Spring 2023)

Introduction

Lectures: Tuesdays and Thursdays,
11:30 am – 12:45 pm, in E-012

Instructor:

Dr. Muhammad Farhan

Faculty Pod, C-214

Phone Ext: 5261

E-mail: muhammad.farhan@sse.habib.edu.pk

Course TAs: tbd

Office Hours: Wed 2:30-3:30 pm and Thu 3:30-4:30 pm + by appointment

Course Description

- Have you ever wondered how a machine or computer is made capable of understanding, interpreting and giving semantics to an image/video?
- Have you ever thought how image/video could be used to automate processes in a wider application domain ranging from industry to biomedicine?
- Computer Vision has emerged as a revolutionary field which employs Image Processing, Pattern Recognition, and Machine Learning to imitate human vision in order to automate processes in a wide range of real-world applications.
- The ever so rapid growth and hence increasing applications has made Computer vision ubiquitous and it, along with Artificial Intelligence, is transforming the way humans approach their daily routine life.
- In this course, the aim is to explore the field of computer vision and pattern recognition from an application perspective where the main focus will be on visual recognition and classification using deep neural networks. The students will learn and implement the state-of-the-art algorithms and techniques for gaining high-level understanding from images and videos.

Credit Hours: 3 (3 + 0)

Prerequisite: MATH205 Linear Algebra

Course Objectives

The course aims to equip students with:

- the components as well as techniques of computer vision and image analysis;
- the concepts as well as ingredients of traditional machine learning and convolutional neural networks based deep learning for visual recognition;
- an ability to utilize the tools of computer vision to solve real world problems, e.g., in industry and biomedicine.

Course Learning Outcomes (CLOs)

Upon completion of this course, students will be able to

describe methods for image modeling and representation as well as for image-based measurements and analysis;	Cog2	PLO1
apply traditional pattern classification algorithms as well as the ones based on deep neural networks on visual data for object and scene recognition and understanding;	Cog3	PLO1
implement computer vision algorithms in MATLAB/OpenCV/**Python**; and	Psy5	PLO5
develop computer vision solutions for applications ranging from industry to medicine and routine life.	Cog5	PLO3

Course Format (Important)

- The material in each class will build up on the previous classes and students are required to read the relevant sections for each week before the lecture.
- Some of the topics are taught employing flipped classroom learning where the students will be required to watch a given video or read slides or material and the lecture session would be utilized to discuss those topics in detail, in group as well as in combined manner.
- Class participation is highly encouraged through questions and answers, and discussion. If a student has any ambiguity during the lecture s/he is supposed to kill that then and there through questions, no matter a silly one. The students are therefore expected to respect their peers.
- Class participation also includes formative assessment through unannounced short quizzes at the end of few randomly selected lectures which requires student to submit a few sentences identifying the main point(s) of a lecture.
- Since there is no lab component so a couple of workshops might be conducted in two working Saturdays in the middle of the semester to get the students required hands-on skills for the assignments as well as for the project.

Required and Reference Texts

Selected Text

- R. Szeliski, “Computer Vision: Algorithms and Applications,” 1st Edition Springer-Verlag, 2010.
- D. Forsyth and J. Ponce, “Computer Vision: A Modern Approach,” 2nd Edition, Prentice Hall, 2012.
- I. Goodfellow, Y. Bengio, and A. Courville “Deep Learning,” MIT Press, 2016. (http://www.deeplearningbook.org/)
- R. Duda, P. Hart, and D. Stork, “Pattern Classification,” 2nd Edition, Wiley, 2001.
- C.M. Bishop, “Pattern Recognition and Machine Learning,” Springer-Verlag, 2006..
- Selected research papers

Marks Distribution

- Quizzes (4x5) 20 %
- Class Participation* 10 %
- Assignments (2 → 8+7) 15 %
- Midterm exam 20 %
- Project and Paper** 35 %
  - Proposal 2%,
  - Detailed Literature Review 6%
  - Mid progress report and demo 4%
  - Final Working demo and presentation 5%
  - Viva (Maturity Test) 5%
  - Research Paper 8%
  - Peer Review 5%

* Random unannounced end-of-lecture quizzes (min 7)

**** NO flexibility esp in project deliverable, 10% deduction on 48 hours delay**

Grading Scale

Letter Grade	GPA Points	Percentage Marks
A+	4.00	95 – 100
A	4.00	90 – 94
A-	3.67	85 – 89
B+	3.33	80 – 84
B	3.00	75 – 79
B-	2.67	70 – 74
C+	2.33	67 – 69
C	2.00	63 – 66
C-	1.67	60 – 62
F	0.00	0 – 59

Tentative Course Schedule

Week	Topic(s)	Reading(s)	Remarks
Week - 1 January 9 – 13, 2023	Intro to Course, Overview of Computer vision, Intro to Object detection and Machine Learning for object and scene recognition		
Week - 2 January 16 – 20, 2023	Pattern Classification (Supervised, Unsupervised, Linear and Non-linear methods)	PCA, LDA for Visual recognition	
Week - 3 January 23 – 27, 2023	Neural Networks: Intro, perceptron for two classes, recap of linear classifier, loss function, optimization	Features (SIFT, HOG) and Object detection (BoW and DPM)	
Week - 4 January 30 – February 3, 2023	Optimization, stochastic gradient descent, back propagation, two layer and multilayer perceptron		Quiz 1: PCA, LDA, SIFT, HOG, BoW, DPM Assignment 1 released
Week - 5 February 6 – 10, 2023	two layer and multilayer perception (Cont'd) Convolutional Neural Networks (CNNs): CNNs intro and basics, CNN layers		

Tentative Course Schedule

Week - 6 February 13 – 17, 2023	CNNs intro and basics, CNN layers (Cont'd)		Quiz 2
Week - 7 February 20 – 24, 2023			**Midterm exam**
Week -8 February 27 – March 3, 2023	Training neural networks, Deep learning software and CNN Architectures for Computer Vision		Project Proposal due Assignment 2 released
Week - 9 March 6 – 10, 2023	Generative models		Assignment 1 due
Week - 10 March 13 – 17, 2023		Recurrent Neural Networks	Quiz 3
Week - 11 March 20 – 24, 2023	Object detection and Segmentation, Semantic segmentation Conference Days: March 23 – 26, 2023 **(No Classes)**	Transformers	Project: Detailed Literature Review due
Week - 12 March 27 – 31, 2023	Semantic segmentation (Cont'd)		Assignment 2 due

Tentative Course Schedule

Week – 13 April 3 – 7, 2023	Self-supervised learning		Project: Mid-progress report/Demo due
Week - 14 April 10 – 14, 2023	Self-supervised learning (Cont'd) Performance assessment, Hyperparameter selection, Visualization and understanding for image data, Texture Models		Quiz 4
Week – 15 April 17 – 20, 202	Multitopic Overview: Color Vision, 3D Representation and Stereo vision		
April 21 – 25, 2023	**Eid ul Fitr†**		
Week – 16 April 26 – 28, 2023	Object Tracking and Motion estimation **Last Day of Classes: April 28, 2023**		Project Finals and Paper due

Papers from Past Offerings

- *M. Farhan et al.*, ”Data extraction and processing pipeline for SDSS DR9 UGRIZ collection,” under finalization and submission to Journal.
- *M. Farhan et al.*, “Age progression in dogs as a two domain image to image translation problem,” under review process.
- *M. Farhan et al.*, "Detection of underground water pipelines for infrastructure mapping of underground water utilities," under finalization and submission to Journal.
- M. A. Khan, K. G. Alamdar, A. Junaid, ***M. Farhan***, "Mitigating the Zero Biased Steering Angles in Self-driving Simulator Datasets," in Proceedings of 17th International Conference on Computer Vision Theory and Applications (VISAPP) 2022.
- S. M. A. Alam, M. Ul Huda, and ***M. Farhan***, "Alpha-trimmed mean filter and XOR based image enhancement for embedding data in image," in Proceedings of IEEE International Conference on Visual Communications and Image Processing.
- S. T. Wasim, S. N. Hasany, M. H. Shaikh, A. A. Ahmed, H. F. Ahmed, K. Abbasi, and ***M. Farhan***, "Sim-to-Real Transfer for Object Detection and Localization on Animals," in CV4Animals Workshop, IEEE International Conference on Computer Vision and Pattern Recognition, Virtual, 2021.
- H. F. Ahmed, H. Jamal, ***M. Farhan***, “Satellite image future landscape prediction using conditional adversarial networks,” in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Brussels, 2021.
- N. Zehra, S. H. Azeem, and ***M. Farhan***, “Human Activity Recognition through ensemble learning of multiple Convolutional Neural Networks,” in *Proceedings of 55th IEEE Annual Conference on Information Sciences and Systems*, USA, 2021.
- M. Haris, M. A. Moin, F. A. Rehman, and ***M. Farhan***, “Vehicle crash prediction using Vision,” in *Proceedings of 55th IEEE Annual Conference on Information Sciences and Systems*, USA, 2021.

What is (computer) vision?

Image (or video)

Sensing device

Interpreting device

Interpretations

garden, spring, bridge, water, trees, flower, green, etc.

The goal of computer vision?

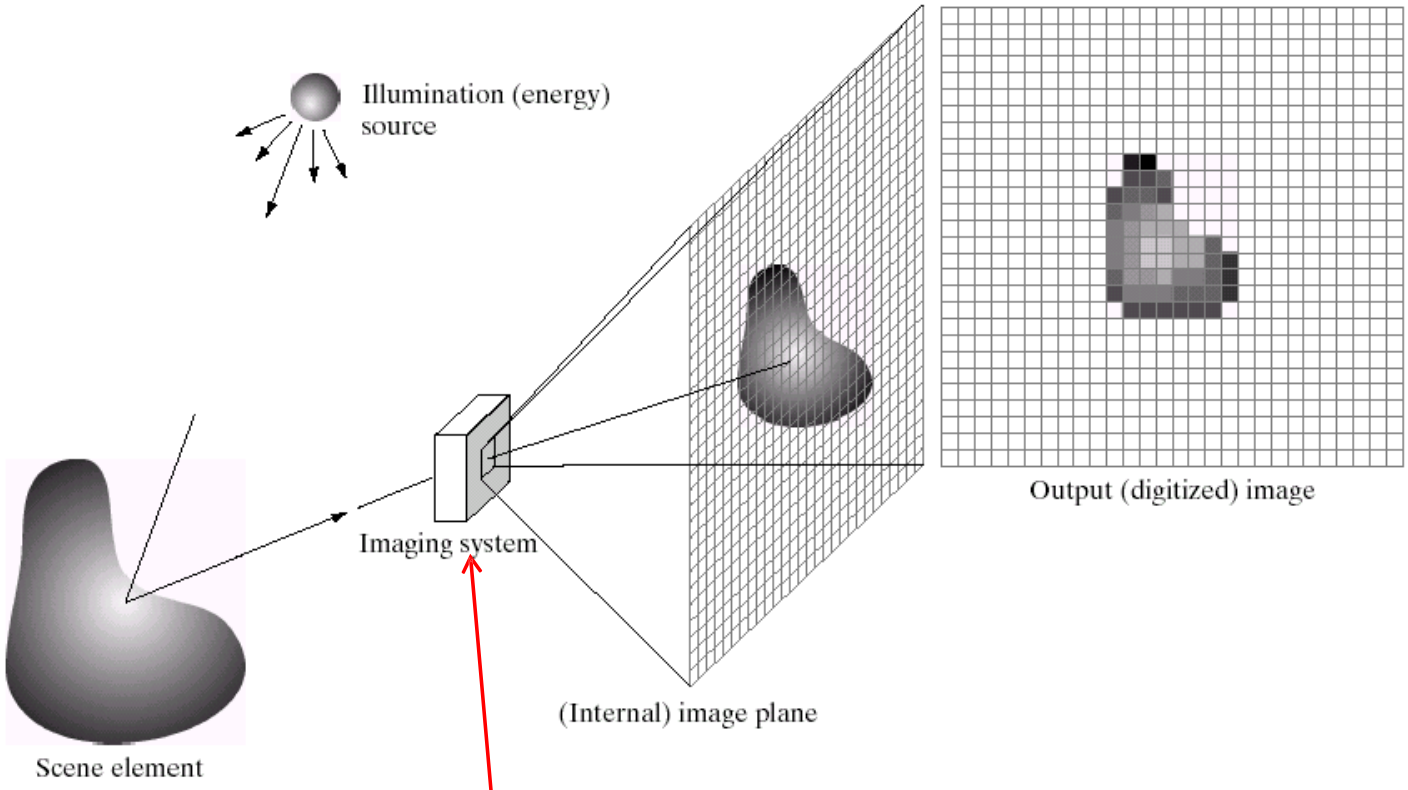
What we see

?

What a computer sees

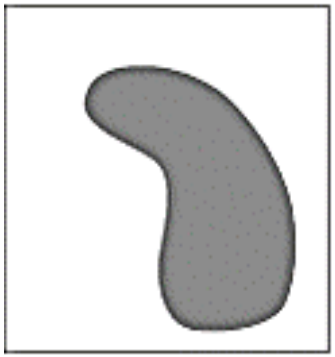
Imaging Acquisition System

Images typically generated by ***illuminating a scene*** and absorbing energy reflected by or transmitted through scene objects

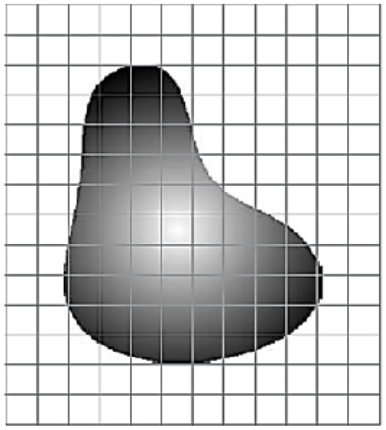


Example: a camera
Converts light to image

Image Discretization and Resolutions



Original image and its spatial Sampling using finer to coarse grid (Spatial Resolution decreasing)



Original image with a sampling grid is Quantized using more and less gray levels (Intensity Resolution decreasing)

The goal of computer vision?

- To bridge the gap between pixels and “meaning”

What we see

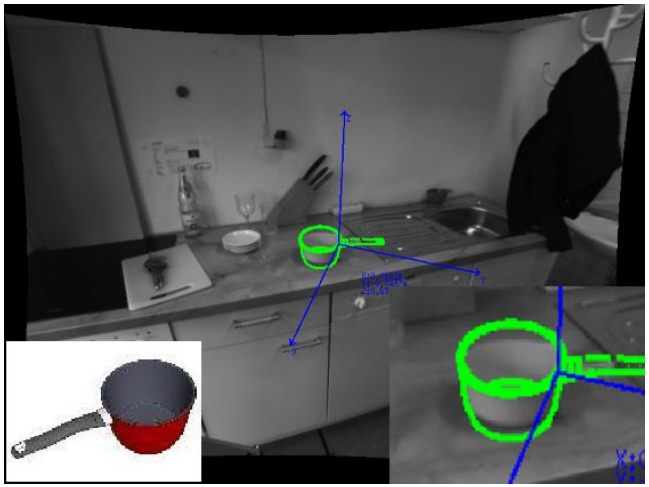
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

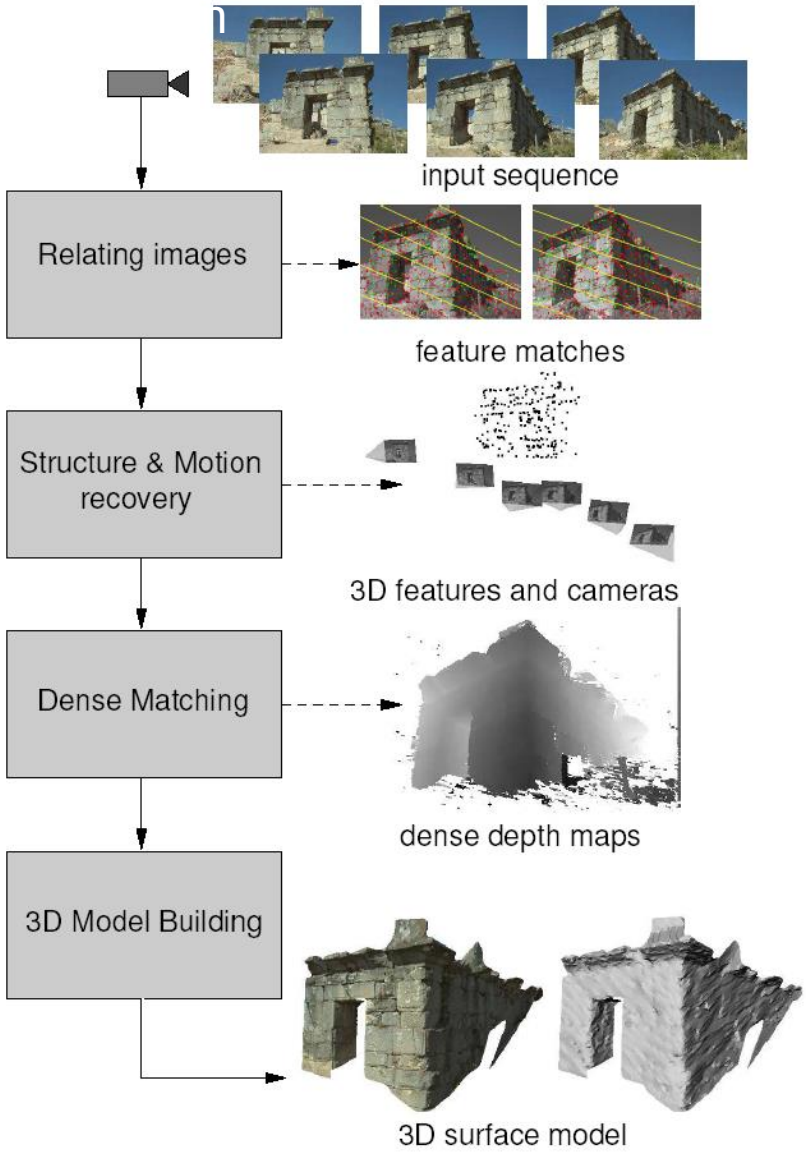
What kind of information can we extract from an image?

- Metric 3D information
- Semantic information

Vision as measurement device



Pollefeys et al.



Vision as a source of semantic information

Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions…

- outdoor
- city
- …

Why study computer vision?

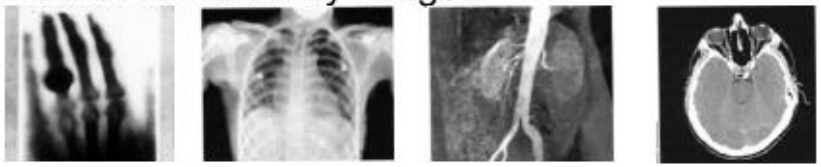
- Vision is useful: Images and video are everywhere!

Manage image archives

- …
- Search for particular image (by its smaller version, by its fragment)
- Search for similar images (landscape paintings, sea views, paintings by the same author)
- Search for a painting with particular colors ("I want a sea view painting to my bedroom with an orange carpet and yellow walls")
- Search for group photos of my family
- Search for an image that will be a good illustration to my article/presentation

Medical diagnosis

- Collection of X-ray images



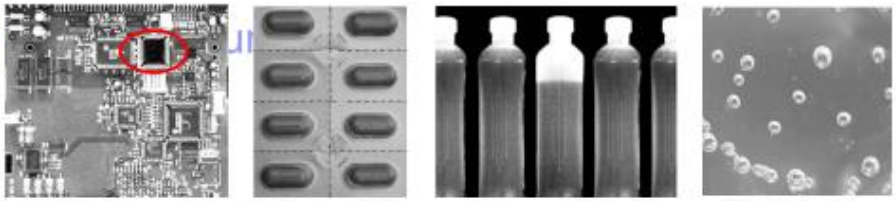
- Search for similar past cases
- Is it similar to the "healthy" case?
- Classification of X-ray images

Security issues

- Video surveillance material
- Faces, fingerprints, retina images

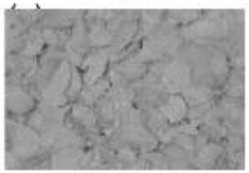
- Detect suspicious objects during the video surveillance
- Detect "wanted" faces during the video surveillance
- Grant or deny access based on fingerprints/retina scanning

Industrial inspection



(a) CD-ROM controller (b) Pack of pills (c) Level of liquid (d) Air-bladders in plastic

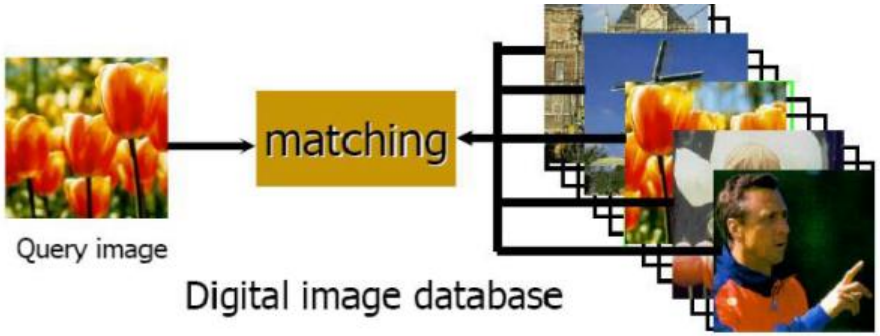
- Control that all parts of the product are on place
- Control if all places in pill pack are filled (b)
- Control the level of liquid in bottles (c)
- Control the quality of plastic details (d)
- And even control the corn flakes! (e)



(e) Corn flakes

Google Image Search · Google Photos · flickr · webshots · picsearch · YouTube Broadcast Yourself

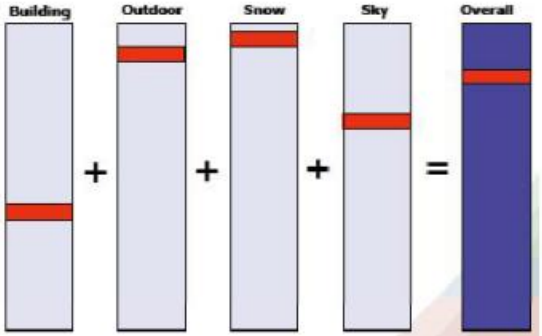
Visual understaning (image retreival)

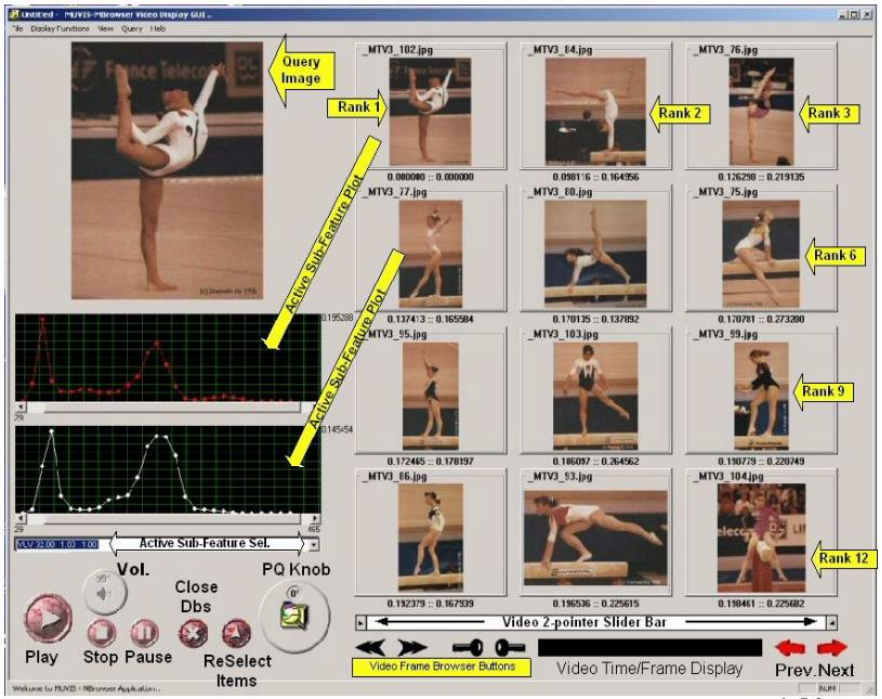


FEATURE DETETCION

0.2	Indoor
0.8	Outdoor
0.7	CityScape
0.3	Landscape
0.1	People
0.0	Face
0.8	Sky
0.2	Vegetation
0.7	Building

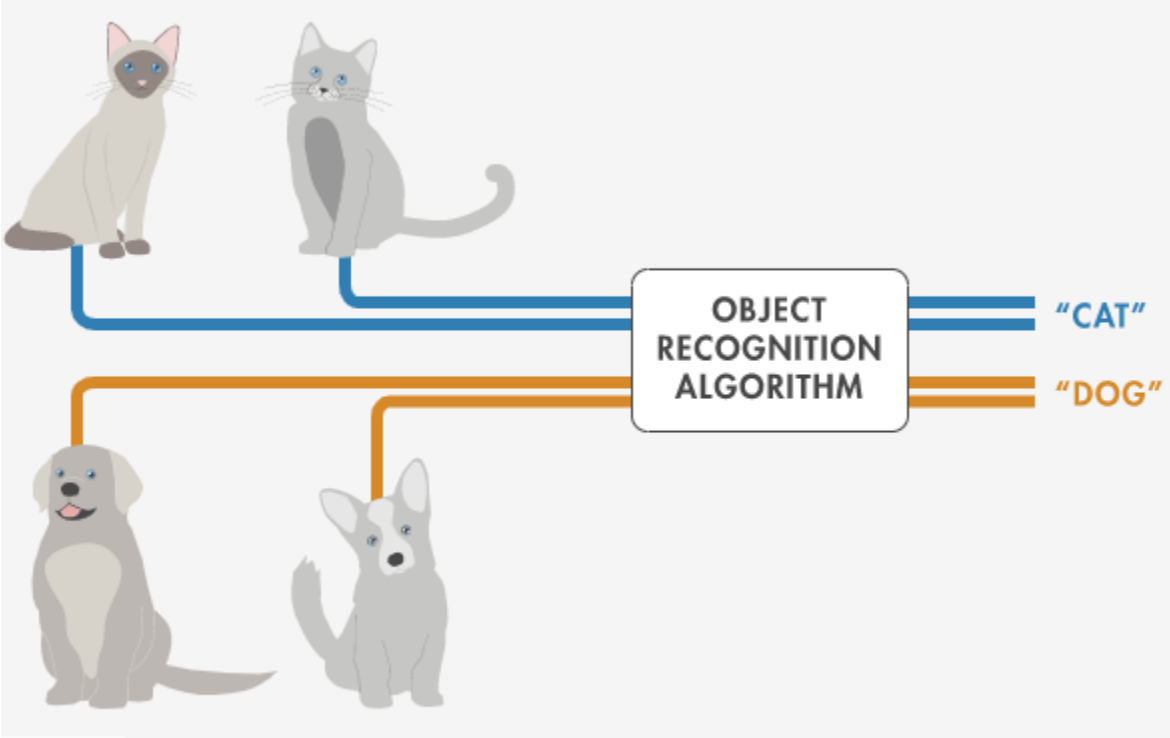
Building :	0.4
Outdoor :	0.8
Snow :	0.9
Sky :	0.6



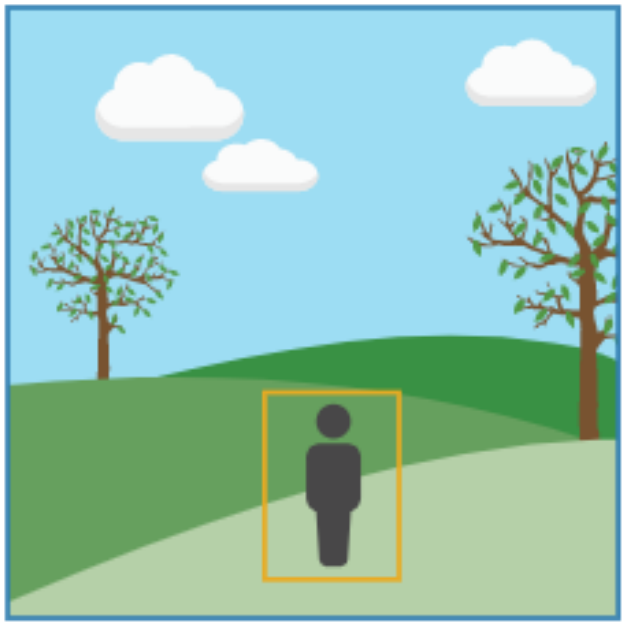


Object Recognition

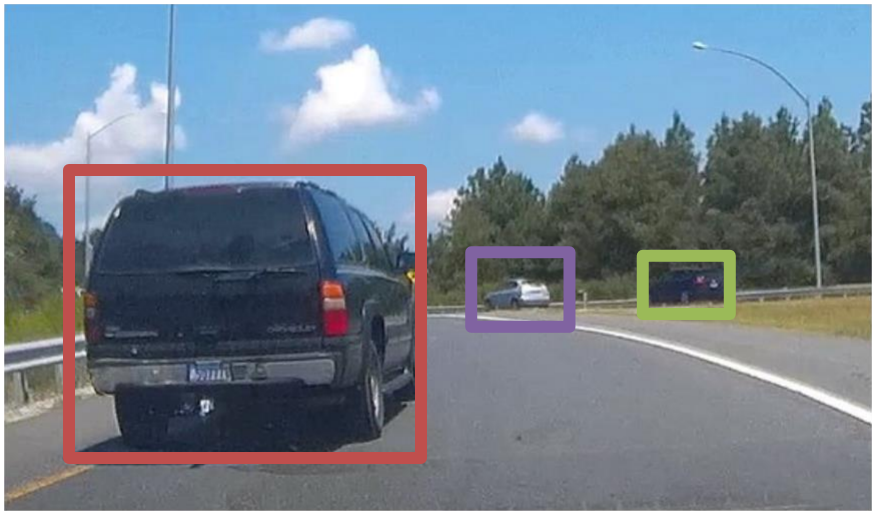
- Object vs Image Classification/Recognition
- Object recognition vs Object detection(recognition + localization)



Using object recognition to identify different categories of objects.

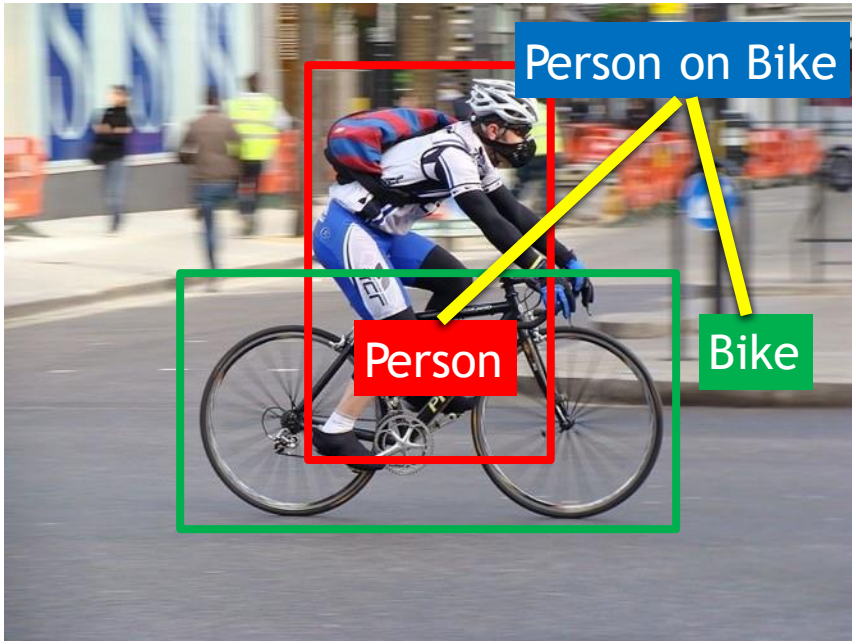


Object Detection

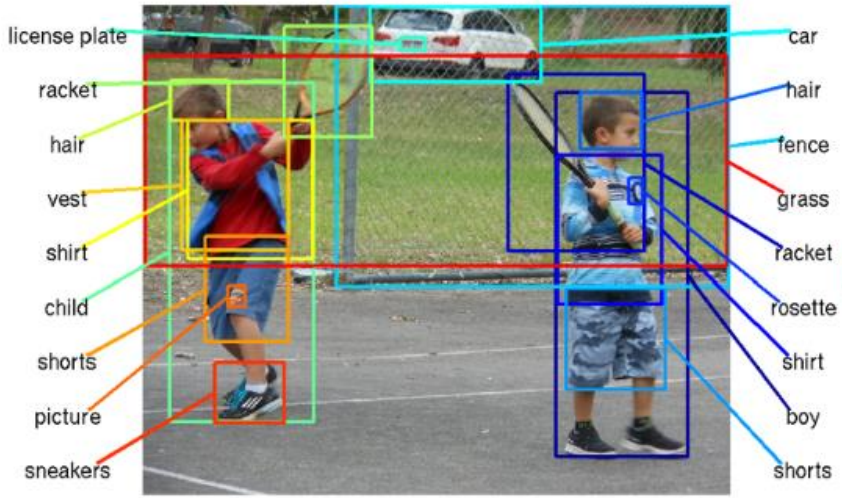


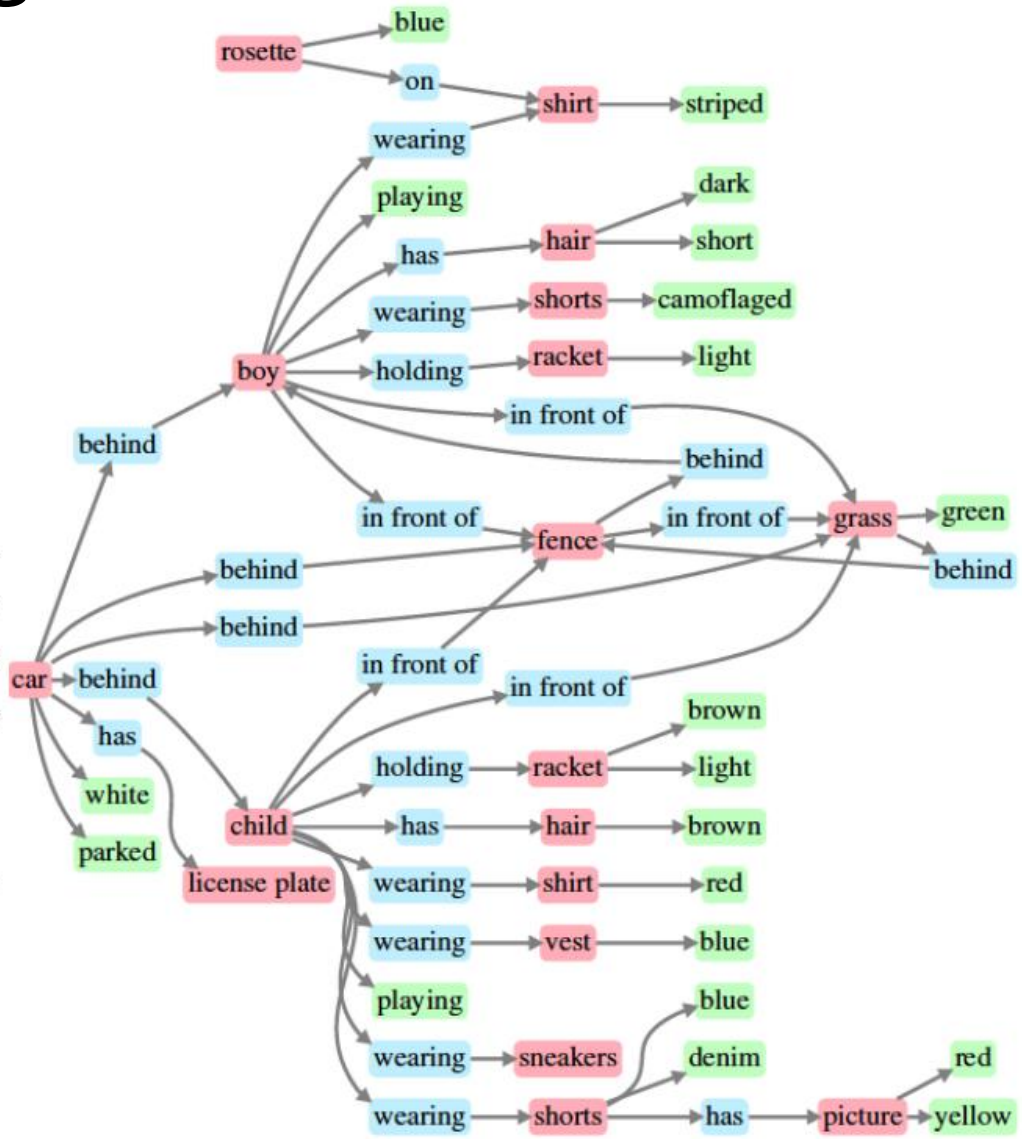
Person

Hammer



Scene/Image Recognition





Johnson *et al.*, "Image Retrieval using Scene Graphs", CVPR 2015

Scene/Image Recognition

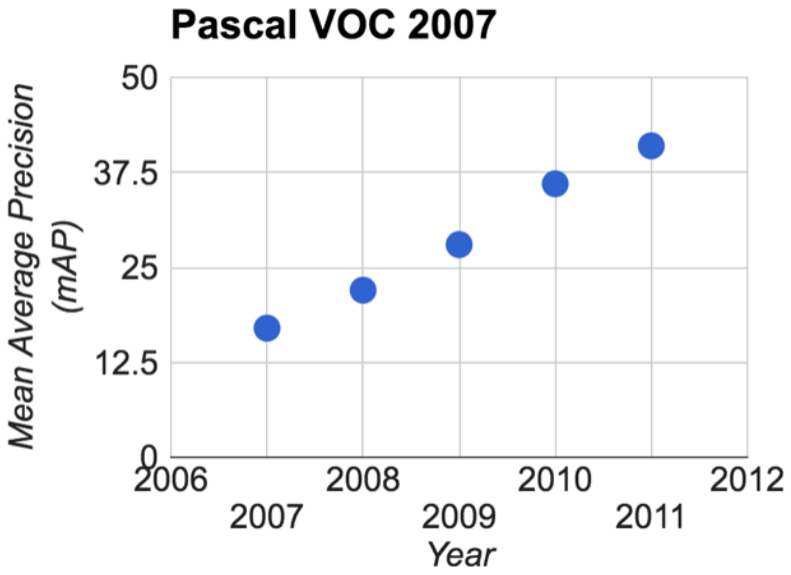
PT = 500ms

Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background. (Subject: SM)

Fei-Fei, Iyer, Koch, Perona, JoV, 2007

PASCAL Visual Object Challenge (20 object categories)

[Everingham et al. 2006-2012]



IMAGENET

www.image-net.org

- 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

Steel drum

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

✔

Output:
Scale
T-shirt Giant panda
Drumstick
Mud turtle

✗

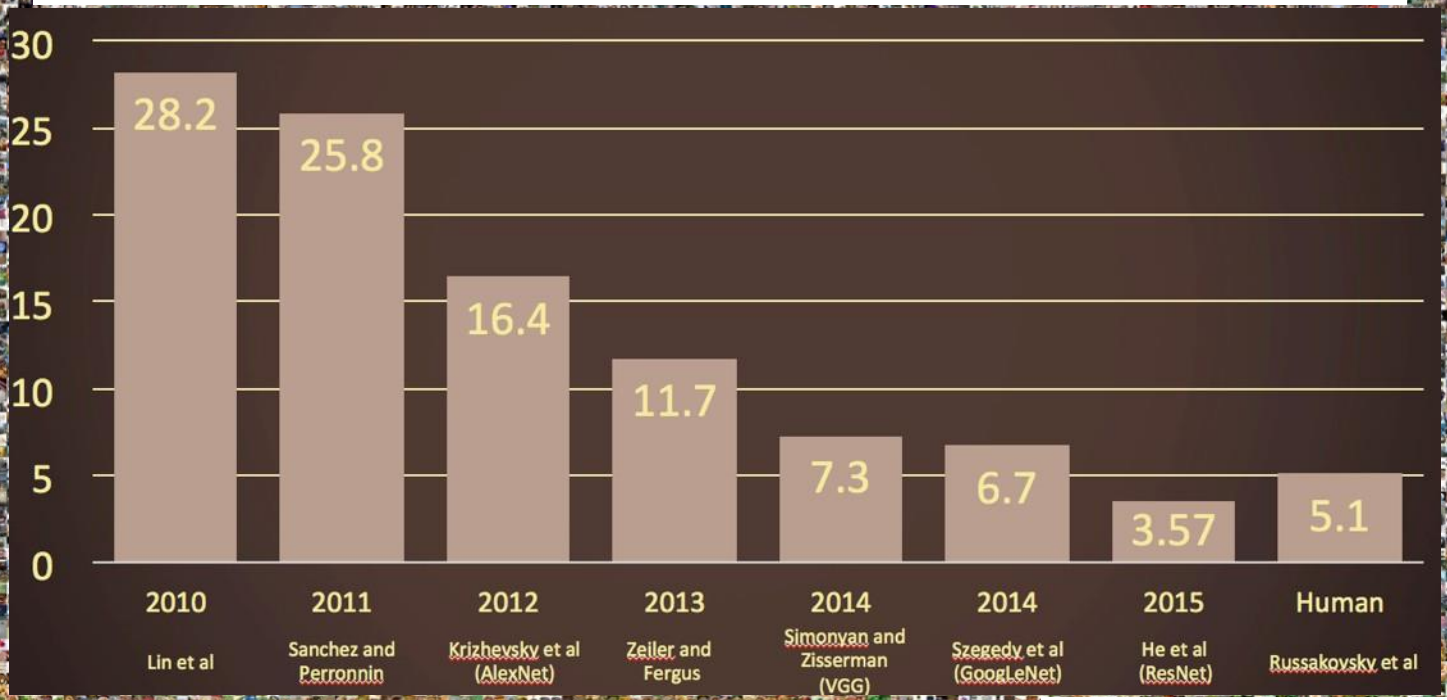
Russakovsky et al. arXiv, 2014

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images



Russakovsky et al. arXiv, 2014

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

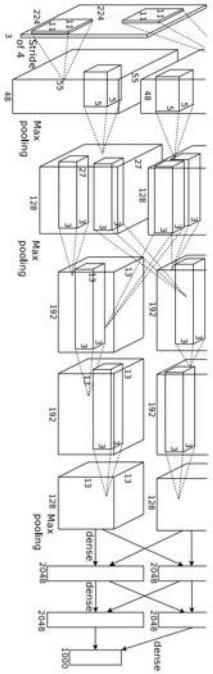
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

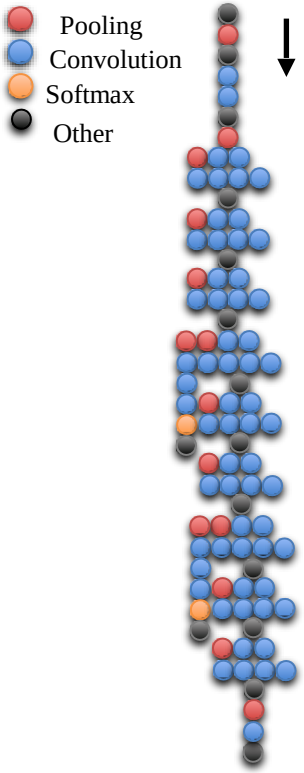
SuperVision



[Krizhevsky NIPS 2012]

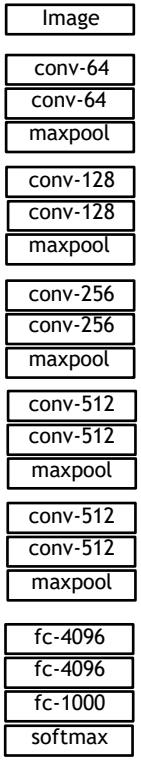
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

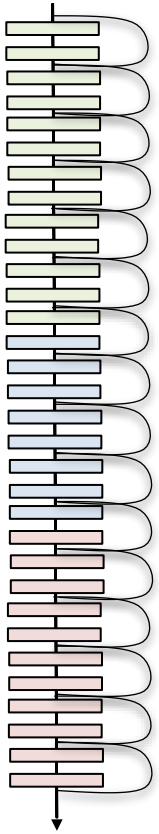
VGG



[Simonyan arxiv 2014]

Year 2015

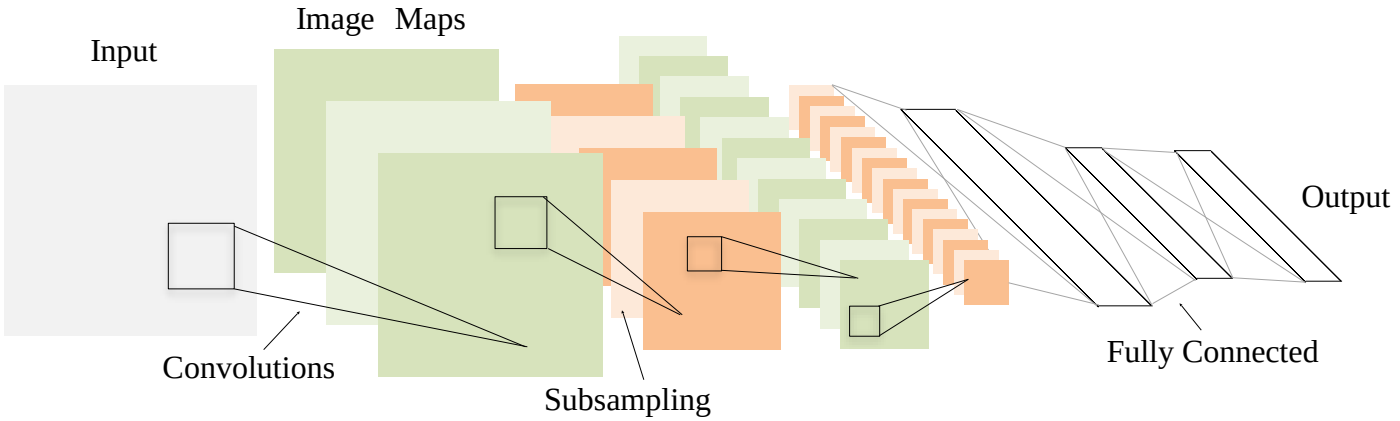
MSRA



[He ICCV 2015]

1998

LeCun et al.



# of transistors

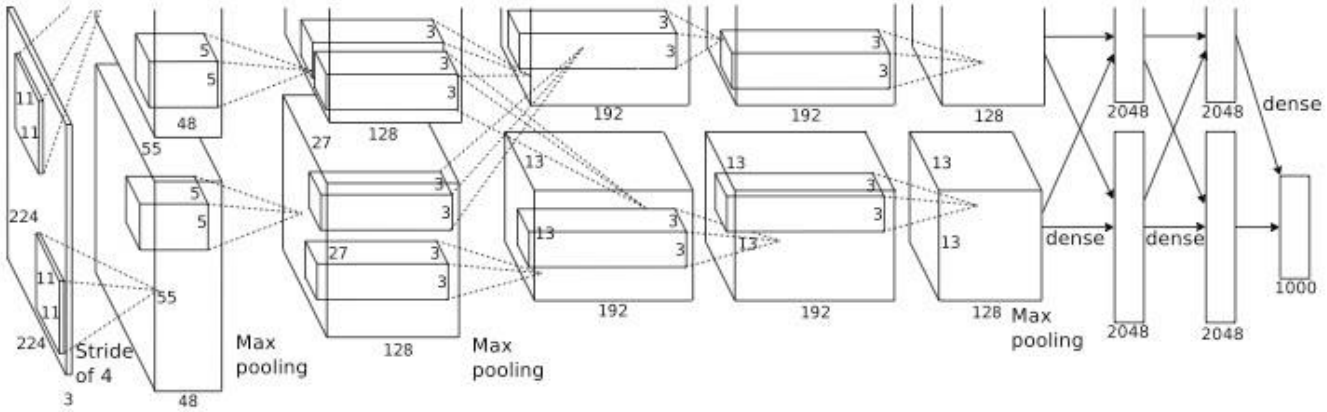
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



# of transistors

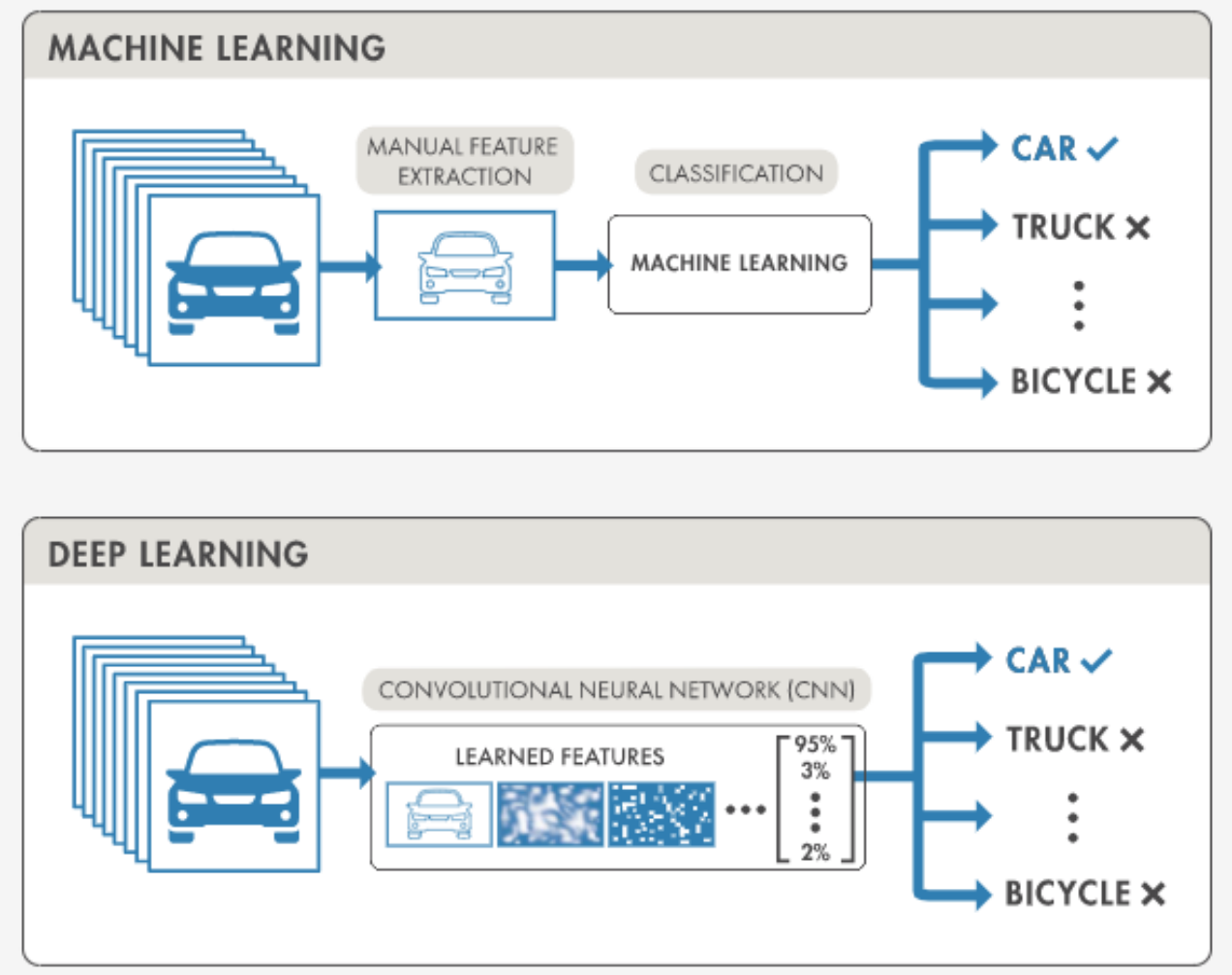
10^9

GPUs

# of pixels used in training

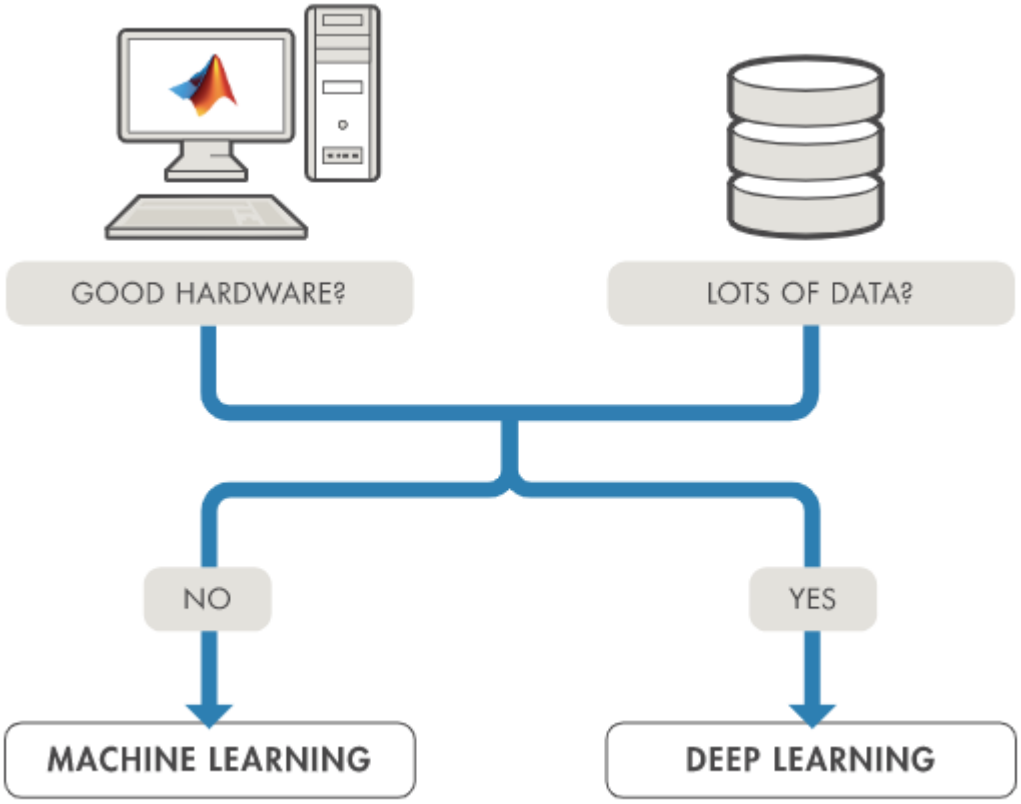
10^{14} IMAGENET

Machine Learning for Vision



Machine learning and deep learning techniques for object recognition.

Object Recognition



Key factors for choosing between deep learning and machine learning.