INSAID

OVERVIEW OF COMPUTER VISION

Introduction

The very likeable computer scientist, entrepreneur, and venture capitalist Paul Graham once mentioned, in his essay "Write Like you Talk":

"Informal language is the athletics clothing of ideas".

With that in mind, let's discuss what Computer Vision "even is", starting from not really knowing anything about it.

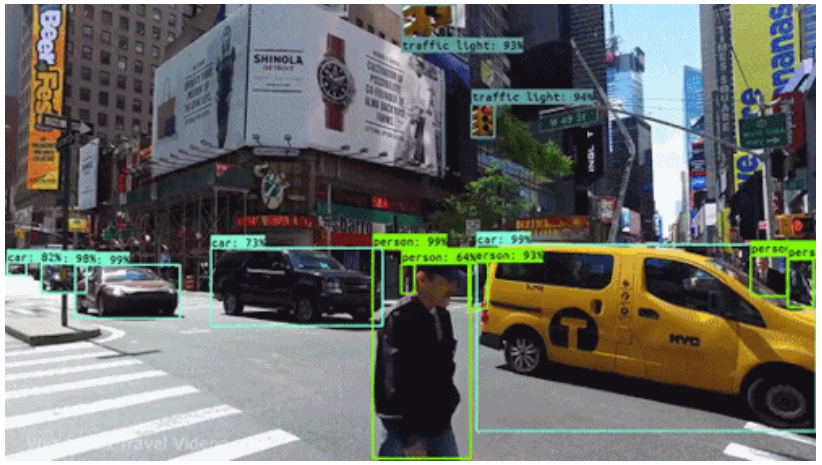
Definition

Computer vision is an interdisciplinary

field that deals with how computer can

be made to gain high-level understanding

from digital images or videos.



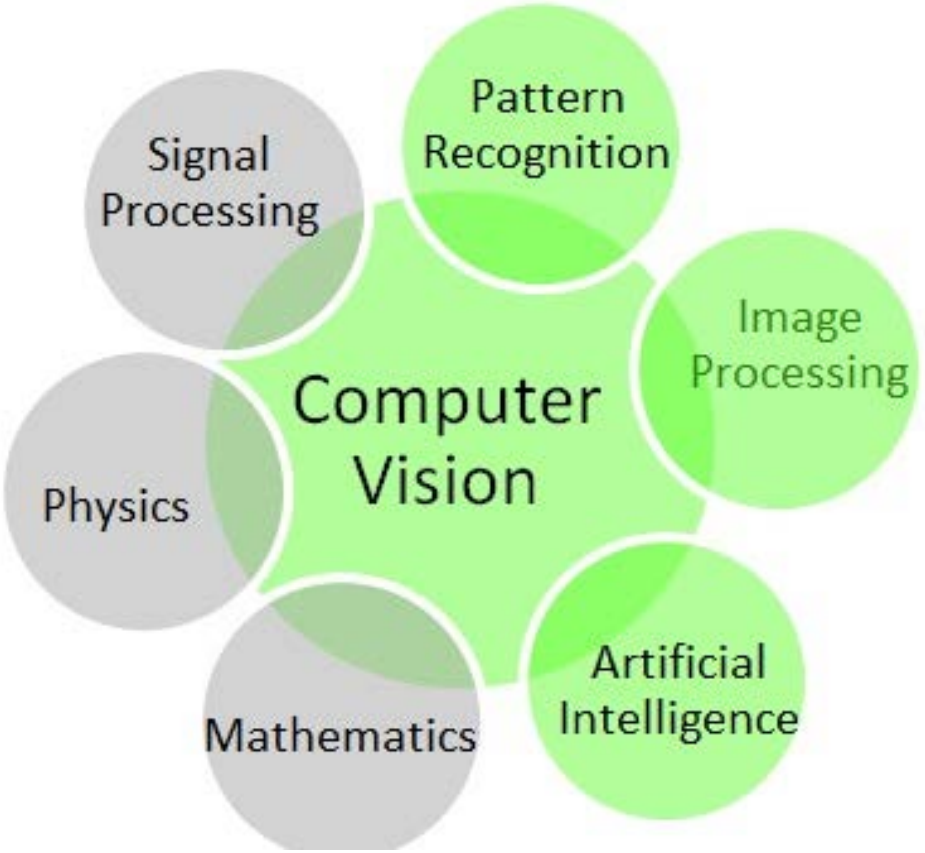
Rough Translation

Computer vision is pretty much

what it sounds like, with just a few

little extra consideration and

distinctions from other related field.



The Long and Storied History of CV

- People have been trying to give computers the gift of vision since right around the time Walt Disney.

- Suffice it to say that half a century of advances in computing, mathematics, and related fields have allowed us to make a lot of progress since then.

"First ever report proposal on Computer Vision by 1st year Undergrad Computer Science student at MIT assigned by Minsky"

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

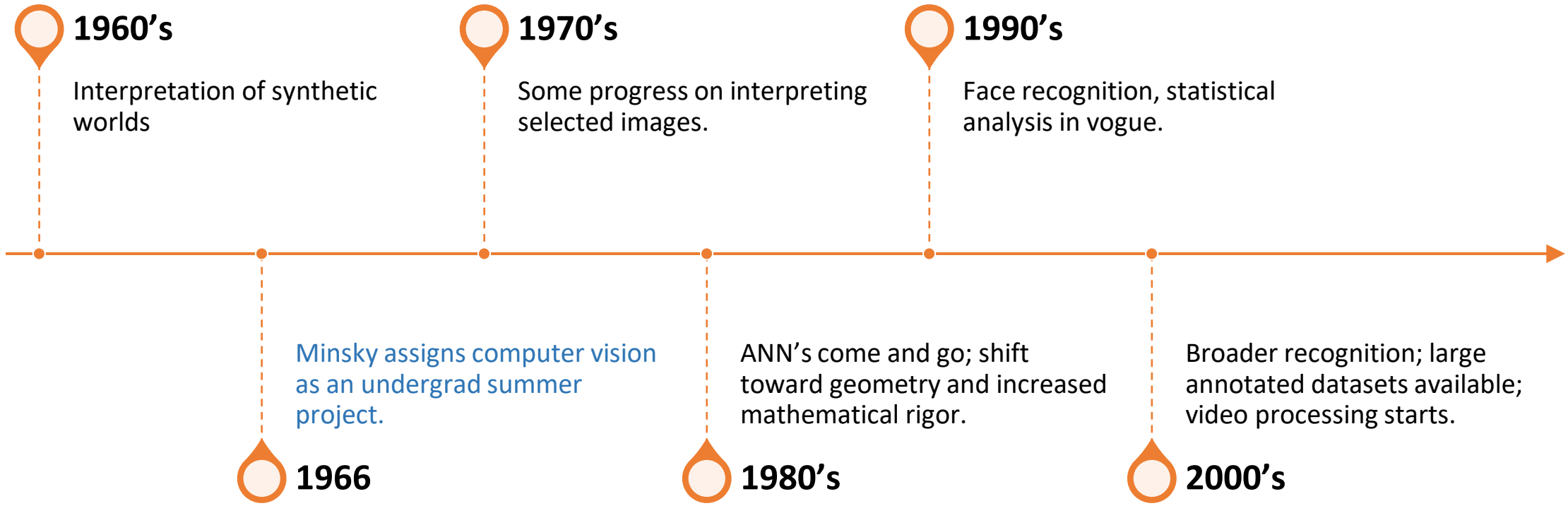
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas


1960’s
Interpretation of synthetic worlds
1970’s
Some progress on interpreting selected images.
1990’s
Face recognition, statistical analysis in vogue.
Minsky assigns computer vision as an undergrad summer project.
1966
ANN’s come and go; shift toward geometry and increased mathematical rigor.
1980’s
Broader recognition; large annotated datasets available; video processing starts.
2000’s

Goal for CV:

Computer vision has dual goals

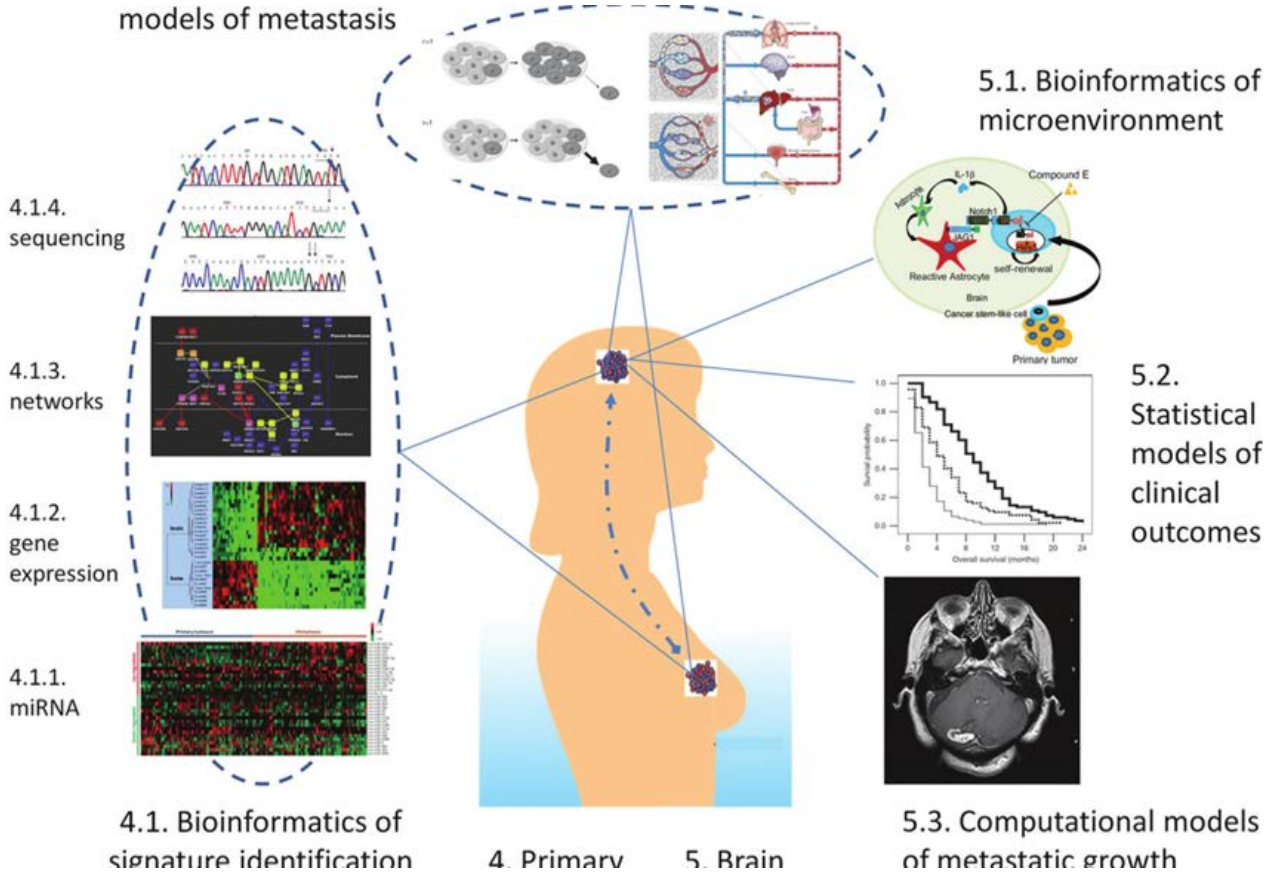
- From biological point of view.
- From engineering point of view.

Biological point of View

It aims to come

up with computational models

of the human visual system.



Engineering Point of View

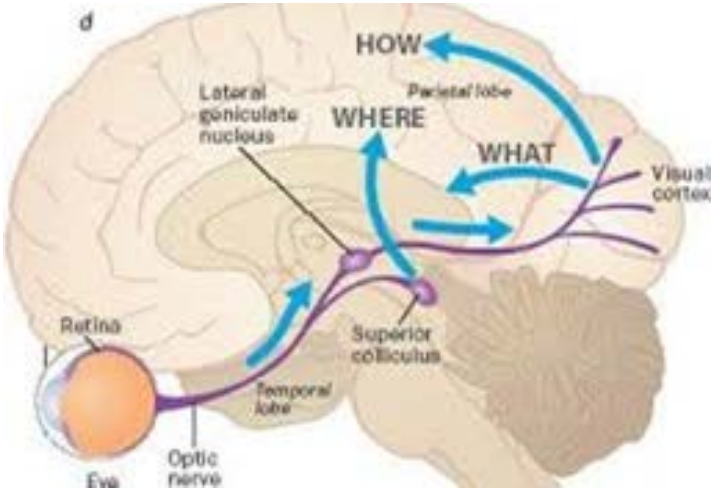
It aims to build autonomous

system which could perform

some of the task which the

human visual system can perform.

Visual System Representation



- It includes the eyes, the connecting pathways through to the visual cortex and other parts of the brain.

- The illustration shows the mammalian system.

What is Visual System?

PART OF THE CENTRAL NERVOUS SYSTEM WHICH GIVES ORGANISMS THE ABILITY TO PROCESS VISUAL DETAIL AS SIGHT, AS WELL AS ENABLING THE FORMATION OF SEVERAL NON-IMAGE PHOTO RESPONSE FUNCTIONS.

IT DETECTS AND INTERPRETS FORMATION OF SEVERAL NON-IMAGE PHOTO RESPONSE FUNCTIONS.

DETECTS AND INTERPRETS INFORMATION FROM VISIBLE LIGHT TO BUILD A REPRESENTATION OF THE SURROUNDING ENVIRONMENT.

AGENDA

Brief walk into the world of Computer Vision

Common Problems that Computer Vision Solves

Few solutions by popular companies in Computer Vision Space

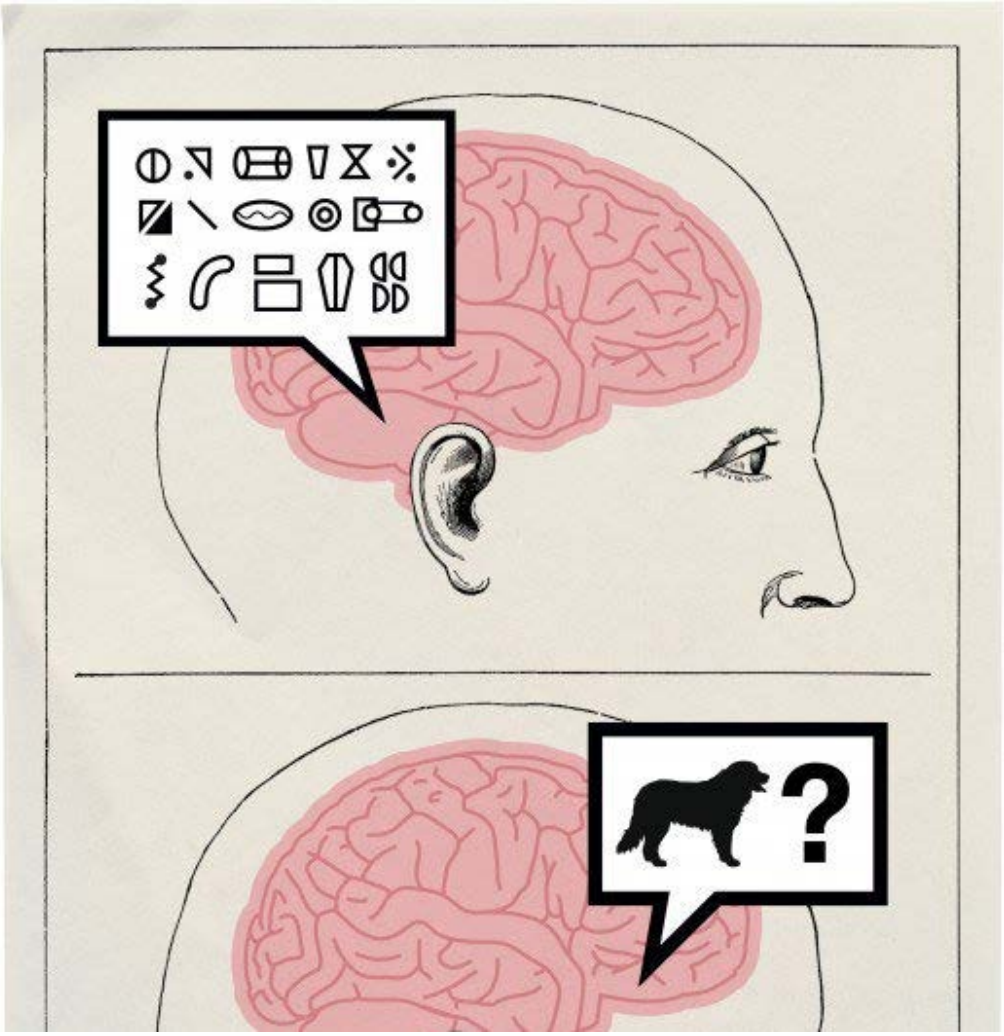
Advancement in the field of Computer Vision

How Computer Vision Works?

One of the major open questions in both Neuroscience and Machine Learning is:

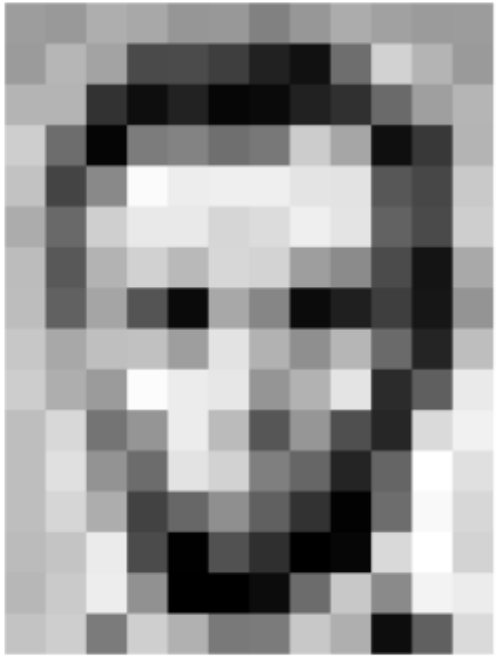
How exactly do our brains work, and how can we approximate that with our own algorithms?

Since we are not decided on how the brain and eyes process images, its difficult to say how all the algorithms used in production approximate our own internal mental processes.



Machines interpret images very simply: as a series of pixels, each with their own set of color values.

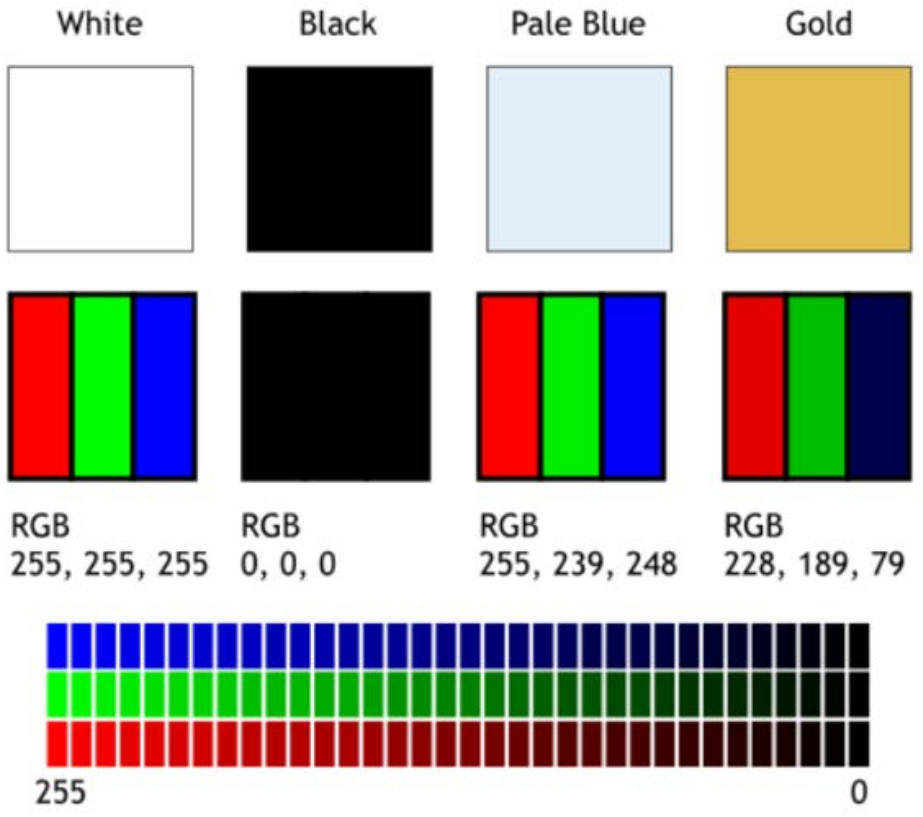
Consider the simplified image adjacent, and how greyscale values are converted into a simple array of numbers:



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

- When we start to add in color, things get more complicated.
- Computers usually read color as a series of 3 values
  - Red
  - Green
  - Blue
  - On that same 0 – 255 scales.



For some perspective on how computationally expensive this is, consider this tree:

Each color value is stored in 8 bits.

8 bits * 3 colors per pixel = 24 bits per pixel

A normal sized 1024 * 768 image * 24 bits per pixel = almost 19M bits, or about 2.36 megabytes.

Implications of Computer Vision

Broad parent name for any computations involving visual content - which includes

Images,

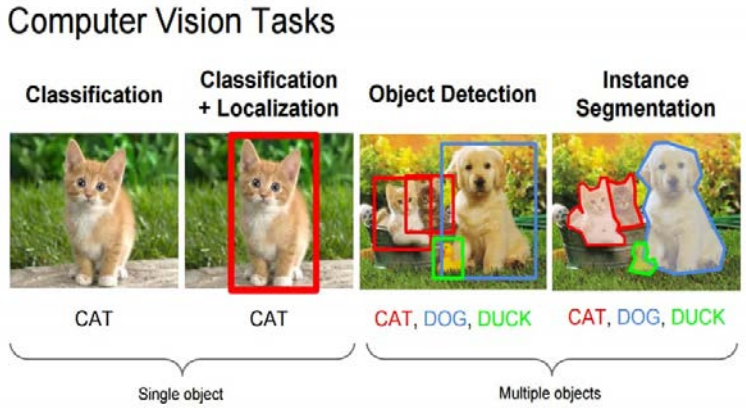
Videos,

Icons, and

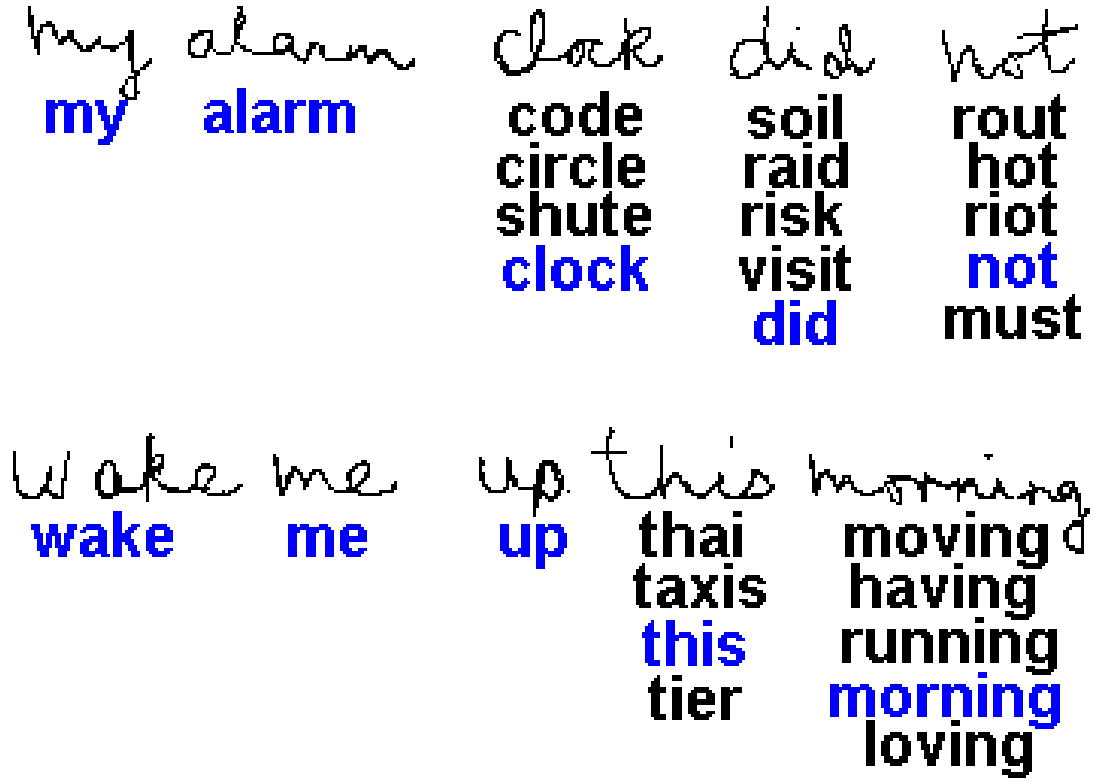
anything else with pixels involved.

But within this parent idea, there are a few specific tasks that are core building blocks:

- Object classification
- Object Identification



A classical application of computer vision is handwriting recognition for digitizing handwritten context.



Outside of just recognition, other method of analysis include:

1. Motion Analysis
2. Image restoration
3. Scene reconstruction
4. Image restoration

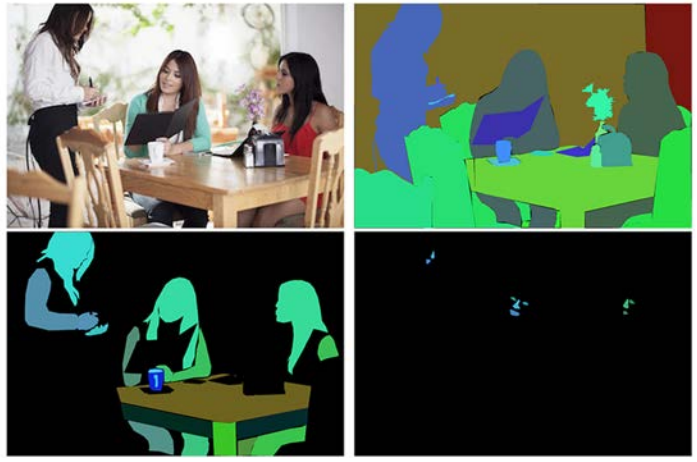
Motion Analysis

Uses Computer vision to estimate

the velocity of objects in a video,

or camera itself

Image Segmentation



Algorithms partition images into multiple set of views.

Scene Reconstruction

Creates 3D model of the scene inputted through images or videos.

Image Restoration

Before

After

Noise such as blurring is removed from photos using Machine learning filters.

Why do we need CV anyway?

CV ALLOWS COMPUTERS, AND THUS ROBOTS, OTHER COMPUTER-CONTROLLED VEHICLES, AND

EVERYTHING FROM FACTORIES AND FARM EQUIPMENT TO SEMI-AUTONOMOUS CARS AND DRONES,

TO RUN MORE EFFICIENTLY AND INTELLIGENTLY AND EVEN SAFELY.

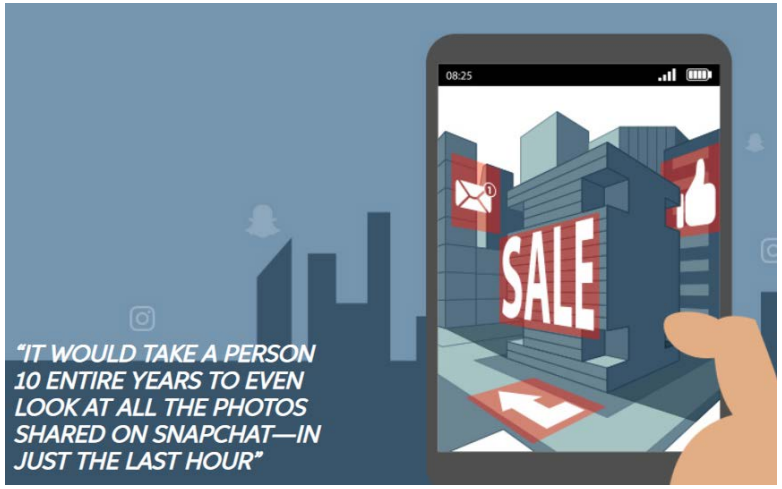
Simply put, our world has become

increasingly filled with digital images

and we need computers to make

sense of it all- it's already well past

human capabilities to keep up



AGENDA

Brief walk into the world of Computer Vision

Common Problems that Computer Vision Solves

Few solutions by popular companies in Computer Vision Space

Advancement in the field of Computer Vision

Market Leaders Using CV

- Today, top technology companies such as Amazon, Google, Microsoft and Facebook are investing billions of dollars in computer vision research and product development.

Computer Vision, an AI technology that allows computers to understand and label images, is now used in:

Convenience stores

Driverless car testing

Daily medical diagnostics

Monitoring the health of crops and livestock.

Use Cases of Computer Vision

Retail and Retail Security

Automotive

Healthcare

Agriculture

Banking

Industrial

Retail and Retail Security

AMAZON

- Amazon recently opened to the public the Amazon Go store where shoppers need not wait in line at the checkout counter to pay for their purchases.
- https://youtu.be/uoKsY9HDk6o

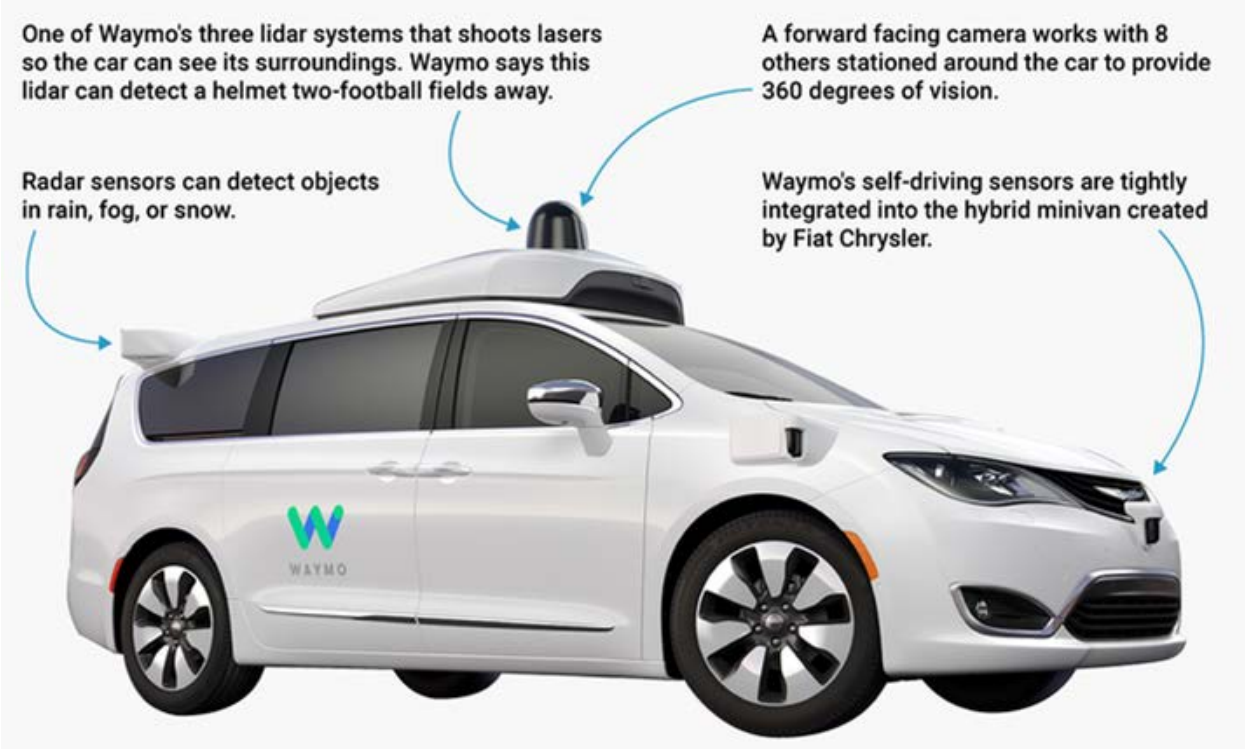
In Retail Fashion

- Amazon has applied for a patent for a virtual mirror.
- The Virtual mirror technology, is described as a blended-reality display that puts a shopper's image into an augmented scene and puts the individual in a virtual dress.
- https://youtu.be/Mr71jrkzWq8

AUTOMOTIVE

WAYMO

- Formerly knowns as Google self-driving car.
- Equipped with sensors and software that can detect 360 degrees of movements of pedestrians, cyclists, vehicles, road work and other objects from up to three football fields away.
- https://youtu.be/TsaES--OTzM



TESLA

- Claims all 3 Autopilot car models are equipped for full self-driving capability.
- Fitted with 8 cameras for 360-degree visibility around the car with a viewing distance of 260 meters around.
- 12 ultrasonic sensors enable the car to detect both hard and soft objects.
- They claim that a forward-facing radar enables the car to see through heavy rain, fog, dust and even the car ahead.
- https://youtu.be/UgNhYGAgmZo

HEALTHCARE

GAUSS SURGICAL

- Developed blood monitoring solutions.
- The solution, the website reports, maximizes transfusions and recognizes hemorrhage better than the human eye.
- Monitoring solution include Triton OR which uses an iPad-based app to capture images of blood on surgical sponges and suction canisters.
- https://youtu.be/cBzJ43zU4FY

DeepLens and DermLens

- Developed by Amazon Web Services(AWS).
- DermLens aims to assist patients to monitor and manage a skin condition called psoriasis.
- DermLens app is intended as a continuous care service where the reported data is available for the physician and care team.
- https://youtu.be/-JDLpSsO45A

AGRICULTURE

SLANTRANGE:

- Claims to offer computer vision – equipped drones that are connected to what the company calls intelligence system.
- Consists of sensors, processors, storage devices, networks, an artificial intelligence software.
- At 120 meters above ground level.
- Camera has a resolution od 4.8 cm/pixel.
- https://youtu.be/EbdwUg_MU-U

CAINTHUS

- It offers animal facial recognition.
- Uses predictive imaging analysis to monitor the health and well-being of crops and livestock.
- https://www.youtube.com/watch?v=022XsKEBMmA&t=11s

BANKING

MITEK SYSTEMS

- Offers image recognition application
- Use machine learning to classify, extract data, and authenticate documents such as passports, ID cards, driver's licenses and checks.
- The application works by having customers take a photo of an ID or a paper check using their mobile devices and sent the user's bank where CV software on the bank's side verifies authenticity.
- https://youtu.be/sGD49ybxS2Q

INDUSTRIAL

OSPREY INFORMATICS

- Used to monitor the status of critical infrastructure such as remote wells, industrial facilities, work activity and site security.
- https://youtu.be/WHhCmmZnAfg

AGENDA

Brief walk into the world of Computer Vision

Common Problems that Computer Vision Solves

Few solutions by popular companies in Computer Vision Space

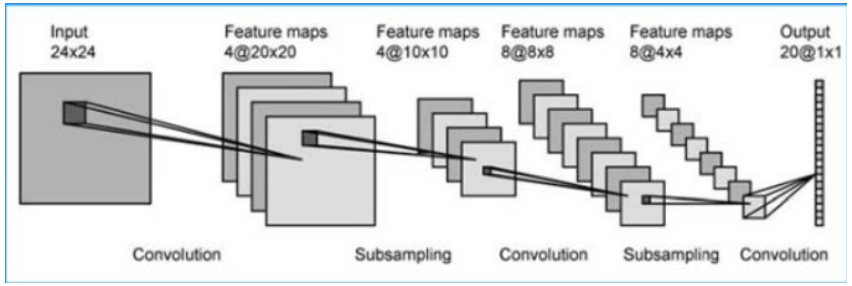
Advancement in the field of Computer Vision

Advancement in Computer Vision

Much of the progress made in computer vision accuracy over the past few years is due in part to a special type of algorithm.

Convolution Neural network are a subset of Deep Learning with a few extra added operations, and they have been shown to achieve impressive accuracy on image-associated tasks.

Convolution Neural Network



- Utilizes the same concepts of Neural Networks, but add in some steps before the normal architecture
- These steps are focused on feature extraction.

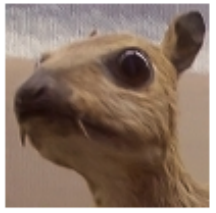
ConvNet Operations

Convolution Neural Net comprises of 3 operations:

- Convolution
- Rectified Linear Unit (ReLU)
- Pooling

Convolution

Input image



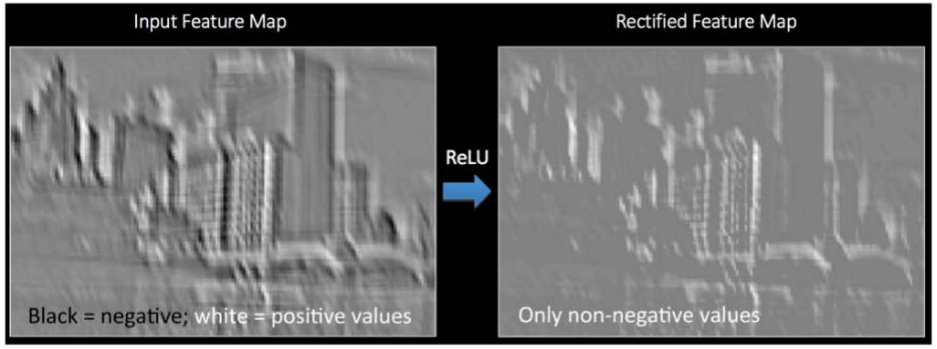
Convolution Kernel

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Feature map

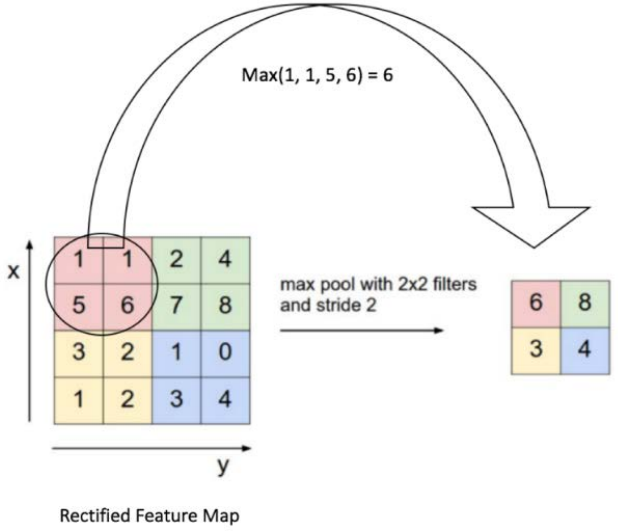
- The input image pixels are modified by a filter.
- This is just a matrix that we multiply different pieces of the input image by.
- The output often called as Feature Map – will usually be smaller than the original image, and theoretically be more informative.

ReLU



- Stands for Rectified Linear Unit.
- Function to introduce non-linearity into the feature map.
- All negative values are simply changed to zero, removing all black form the image.
- The formal function is y=max(0, x).

Pooling



- The image is scanned over by a set width of pixels, and either the max, sum, or average of those pixels is taken as a representation of that portion of the image.
- Further reduces the size of the feature maps(s) by a factor of whatever size is pooled.

Summing up the CNN

All of these operations – Convolution, ReLU, and Pooling – are often applied twice in a row before concluding the process of the feature extraction.

- The output of this whole process are then passes into a neural net for classification.
- The final architecture looks as follow: