

# Computer Vision

Aditya Firman Ihsan





# Some Numbers

**1.72 trillion**

photos are taken worldwide every year, which equals 54,400 per second, or 4.7 billion per day.

By 2030, around 2.3 trillion photos will be taken every year.

**12.4 trillion**

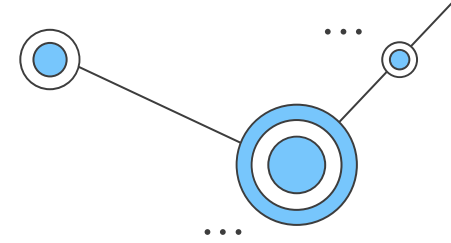
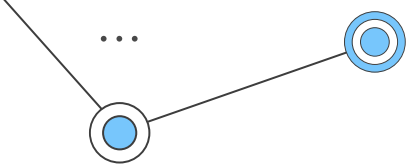
photos have been taken throughout history. By 2030, this number will increase to 28.6 trillion.

**750 billion**

images are on the internet, which is only 6% of the total photos that were ever taken since most of the photos we take are never shared.

**136 billion**

images are on Google Images. By 2030, there will be 382 billion images on Google Images.



Just like to HEAR is not the same as  
to LISTEN, to TAKE PICTURES is not  
the same as to SEE

**Fei Fei Li - Professor of Computer Science, Stanford  
University**

# Apa yang dilihat oleh komputer?

0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,45	0,18	0,00	0,00	0,00	0,34	0,29	0,00	0,00	0,00	0,01	0,00	0,00	0,00
0,00	0,00	0,01	0,00	0,08	0,71	0,75	0,78	0,97	0,81	0,41	0,99	0,87	0,80	0,79	0,35	0,00	0,00	0,01	0,00	0,00
0,00	0,01	0,00	0,00	0,69	0,81	0,75	0,75	0,74	0,76	0,85	0,78	0,75	0,75	0,78	0,85	0,39	0,00	0,02	0,00	0,00
0,00	0,02	0,00	0,16	0,80	0,75	0,75	0,75	0,75	0,73	0,73	0,74	0,77	0,77	0,73	0,79	0,72	0,00	0,00	0,00	0,00
0,00	0,01	0,00	0,37	0,81	0,76	0,71	0,73	0,75	0,73	0,74	0,75	0,76	0,75	0,69	0,78	0,73	0,00	0,00	0,01	0,00
0,00	0,00	0,00	0,56	0,81	0,80	0,72	0,71	0,75	0,73	0,75	0,76	0,76	0,72	0,72	0,77	0,80	0,05	0,00	0,02	0,00
0,00	0,00	0,00	0,72	0,78	0,85	0,82	0,71	0,74	0,73	0,75	0,74	0,73	0,70	0,76	0,78	0,84	0,26	0,00	0,01	0,00
0,00	0,00	0,00	0,82	0,80	0,77	0,76	0,76	0,77	0,75	0,75	0,76	0,80	0,73	0,64	0,78	0,86	0,43	0,00	0,01	0,00
0,00	0,00	0,00	0,74	0,83	0,65	0,64	0,80	0,76	0,76	0,76	0,78	0,78	0,75	0,53	0,77	0,85	0,61	0,00	0,00	0,00
0,00	0,00	0,01	0,78	0,83	0,66	0,67	0,82	0,76	0,77	0,79	0,79	0,77	0,76	0,53	0,77	0,84	0,73	0,00	0,00	0,00
0,00	0,00	0,05	0,79	0,83	0,51	0,69	0,83	0,77	0,76	0,79	0,80	0,77	0,77	0,53	0,75	0,85	0,69	0,00	0,00	0,00
0,00	0,00	0,12	0,82	0,82	0,55	0,75	0,83	0,80	0,75	0,79	0,82	0,77	0,76	0,58	0,75	0,85	0,74	0,00	0,00	0,00
0,00	0,00	0,18	0,84	0,81	0,53	0,79	0,80	0,80	0,76	0,79	0,83	0,77	0,75	0,64	0,73	0,85	0,77	0,07	0,00	0,00
0,00	0,00	0,25	0,85	0,81	0,56	0,82	0,80	0,81	0,76	0,79	0,84	0,78	0,75	0,69	0,70	0,86	0,78	0,09	0,00	0,00
0,00	0,00	0,32	0,87	0,80	0,68	0,82	0,80	0,82	0,76	0,80	0,84	0,78	0,75	0,73	0,65	0,86	0,80	0,09	0,00	0,00
0,00	0,00	0,34	0,86	0,78	0,67	0,82	0,80	0,82	0,75	0,80	0,86	0,79	0,73	0,76	0,62	0,87	0,80	0,12	0,00	0,00
0,00	0,00	0,28	0,86	0,78	0,71	0,80	0,80	0,83	0,75	0,81	0,87	0,80	0,71	0,81	0,60	0,86	0,82	0,07	0,00	0,00
0,00	0,00	0,36	0,85	0,69	0,65	0,84	0,81	0,85	0,73	0,81	0,87	0,80	0,70	0,85	0,49	0,82	0,86	0,32	0,00	0,00
0,00	0,00	0,89	0,83	0,63	0,63	0,84	0,80	0,85	0,73	0,80	0,88	0,80	0,69	0,89	0,44	0,80	0,85	1,00	0,00	0,00
0,00	0,36	0,86	0,79	0,83	0,76	0,80	0,83	0,86	0,73	0,80	0,90	0,81	0,72	0,82	0,71	0,84	0,80	0,87	0,16	0,00
0,00	0,30	0,85	0,81	0,88	0,84	0,77	0,86	0,88	0,74	0,80	0,91	0,81	0,72	0,80	0,81	0,85	0,79	0,86	0,33	0,00
0,00	0,58	0,80	0,95	0,96	0,85	0,80	0,87	0,89	0,74	0,80	0,92	0,82	0,72	0,83	0,79	1,00	0,83	0,67	0,59	0,00
0,00	0,24	0,00	0,00	0,00	0,70	0,87	0,87	0,89	0,74	0,81	0,93	0,83	0,68	0,97	0,29	0,00	0,09	0,00	0,13	0,00
0,00	0,00	0,00	0,00	0,00	0,81	0,85	0,88	0,89	0,74	0,82	0,95	0,83	0,69	0,99	0,40	0,00	0,00	0,00	0,00	0,00
0,00	0,02	0,01	0,00	0,00	0,85	0,85	0,89	0,89	0,74	0,82	0,95	0,84	0,74	0,85	0,42	0,00	0,02	0,00	0,02	0,00
0,00	0,00	0,00	0,00	0,00	0,74	0,85	0,90	0,85	0,73	0,85	0,94	0,82	0,71	0,85	0,43	0,00	0,01	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,79	0,91	0,94	0,91	0,96	0,85	0,98	0,89	0,77	0,90	0,57	0,00	0,03	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,18	0,12	0,01	0,33	0,48	0,00	0,00	0,53	0,44	0,40	0,22	0,00	0,02	0,00	0,00	0,00

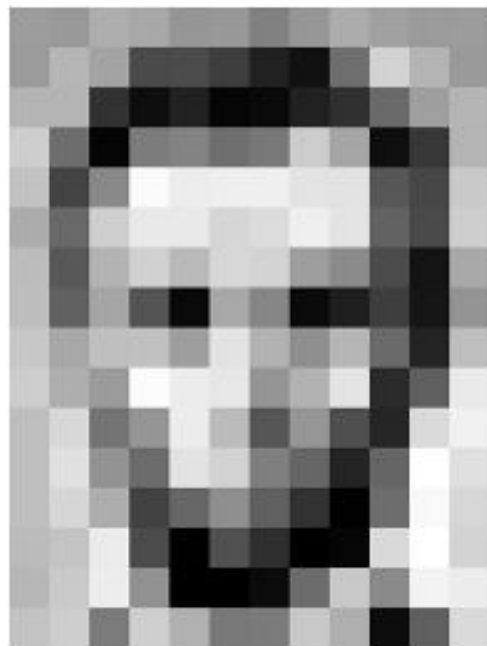
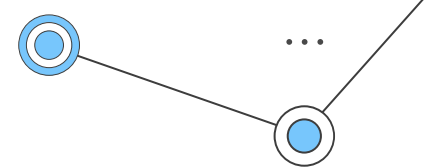
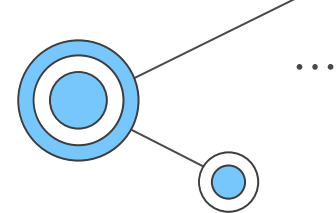
# Apa yang dilihat oleh komputer?

0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,45	0,18	0,00	0,00	0,00	0,34	0,29	0,00	0,00	0,00	0,01	0,00	0,00	0,00
0,00	0,00	0,01	0,00	0,08	0,71	0,75	0,78	0,97	0,81	0,41	0,99	0,87	0,80	0,79	0,35	0,00	0,00	0,01	0,00	0,00
0,00	0,01	0,00	0,00	0,69	0,81	0,75	0,75	0,74	0,76	0,85	0,78	0,75	0,75	0,78	0,85	0,39	0,00	0,02	0,00	0,00
0,00	0,02	0,00	0,16	0,80	0,75	0,75	0,75	0,75	0,73	0,73	0,74	0,77	0,77	0,73	0,79	0,72	0,00	0,00	0,00	0,00
0,00	0,01	0,00	0,37	0,81	0,76	0,71	0,73	0,75	0,73	0,74	0,75	0,76	0,75	0,69	0,78	0,73	0,00	0,00	0,01	0,00
0,00	0,00	0,00	0,56	0,81	0,80	0,72	0,71	0,75	0,73	0,75	0,76	0,76	0,72	0,72	0,77	0,80	0,05	0,00	0,02	0,00
0,00	0,00	0,00	0,72	0,78	0,85	0,82	0,71	0,74	0,73	0,75	0,74	0,73	0,70	0,76	0,78	0,84	0,26	0,00	0,01	0,00
0,00	0,00	0,00	0,82	0,80	0,77	0,76	0,76	0,77	0,75	0,75	0,76	0,80	0,73	0,64	0,78	0,86	0,43	0,00	0,01	0,00
0,00	0,00	0,00	0,74	0,83	0,65	0,64	0,80	0,76	0,76	0,76	0,78	0,78	0,75	0,53	0,77	0,85	0,61	0,00	0,00	0,00
0,00	0,00	0,01	0,78	0,83	0,66	0,67	0,82	0,76	0,77	0,79	0,79	0,77	0,76	0,53	0,77	0,84	0,73	0,00	0,00	0,00
0,00	0,00	0,05	0,79	0,83	0,51	0,69	0,83	0,77	0,76	0,79	0,80	0,77	0,77	0,53	0,75	0,85	0,69	0,00	0,00	0,00
0,00	0,00	0,12	0,82	0,82	0,55	0,75	0,83	0,80	0,75	0,79	0,82	0,77	0,76	0,58	0,75	0,85	0,74	0,00	0,00	0,00
0,00	0,00	0,18	0,84	0,81	0,53	0,79	0,80	0,80	0,76	0,79	0,83	0,77	0,75	0,64	0,73	0,85	0,77	0,07	0,00	0,00
0,00	0,00	0,25	0,85	0,81	0,56	0,82	0,80	0,81	0,76	0,79	0,84	0,78	0,75	0,69	0,70	0,86	0,78	0,09	0,00	0,00
0,00	0,00	0,32	0,87	0,80	0,68	0,82	0,80	0,82	0,76	0,80	0,84	0,78	0,75	0,73	0,65	0,86	0,80	0,09	0,00	0,00
0,00	0,00	0,34	0,86	0,78	0,67	0,82	0,80	0,82	0,75	0,80	0,86	0,79	0,73	0,76	0,62	0,87	0,80	0,12	0,00	0,00
0,00	0,00	0,28	0,86	0,78	0,71	0,80	0,80	0,83	0,75	0,81	0,87	0,80	0,71	0,81	0,60	0,86	0,82	0,07	0,00	0,00
0,00	0,00	0,36	0,85	0,69	0,65	0,84	0,81	0,85	0,73	0,81	0,87	0,80	0,70	0,85	0,49	0,82	0,86	0,32	0,00	0,00
0,00	0,00	0,89	0,83	0,63	0,63	0,84	0,80	0,85	0,73	0,80	0,88	0,80	0,69	0,89	0,44	0,80	0,85	1,00	0,00	0,00
0,00	0,36	0,86	0,79	0,83	0,76	0,80	0,83	0,86	0,73	0,80	0,90	0,81	0,72	0,82	0,71	0,84	0,80	0,87	0,16	0,00
0,00	0,30	0,85	0,81	0,88	0,84	0,77	0,86	0,88	0,74	0,80	0,91	0,81	0,72	0,80	0,81	0,85	0,79	0,86	0,33	0,00
0,00	0,58	0,80	0,95	0,96	0,85	0,80	0,87	0,89	0,74	0,80	0,92	0,82	0,72	0,83	0,79	1,00	0,83	0,67	0,59	0,00
0,00	0,24	0,00	0,00	0,70	0,87	0,87	0,89	0,74	0,81	0,93	0,83	0,68	0,97	0,29	0,00	0,09	0,00	0,13	0,00	0,00
0,00	0,00	0,00	0,00	0,81	0,85	0,88	0,89	0,74	0,82	0,95	0,83	0,69	0,99	0,40	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,02	0,01	0,00	0,00	0,85	0,85	0,89	0,89	0,74	0,82	0,95	0,84	0,74	0,85	0,42	0,00	0,02	0,00	0,02	0,00
0,00	0,00	0,00	0,00	0,00	0,74	0,85	0,90	0,85	0,73	0,85	0,94	0,82	0,71	0,85	0,43	0,00	0,01	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,79	0,91	0,94	0,91	0,96	0,85	0,98	0,89	0,77	0,90	0,57	0,00	0,03	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,18	0,12	0,01	0,33	0,48	0,00	0,00	0,53	0,44	0,40	0,22	0,00	0,02	0,00	0,00	0,00

# Apa yang dilihat oleh manusia?



0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,45	0,34	0,29	0,00	0,00	0,00	0,01	0,00	0,00	0,00			
0,00	0,00	0,01	0,00	0,08	0,71	0,75	0,75	0,80	0,79	0,35	0,00	0,00	0,01	0,00	0,00	0,00			
0,00	0,01	0,00	0,00	0,69	0,81	0,81	0,81	0,78	0,85	0,39	0,00	0,02	0,00	0,00	0,00	0,00			
0,00	0,02	0,00	0,16	0,89	0,89	0,89	0,89	0,72	0,72	0,72	0,00	0,00	0,00	0,00	0,00	0,00			
0,00	0,01	0,00	0,37	0,73	0,73	0,73	0,73	0,73	0,73	0,73	0,00	0,00	0,01	0,00	0,00	0,00			
0,00	0,00	0,00	0,56	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,05	0,00	0,02	0,00	0,00	0,00			
0,00	0,00	0,00	0,72	0,84	0,84	0,84	0,84	0,84	0,84	0,84	0,26	0,00	0,01	0,00	0,00	0,00			
0,00	0,00	0,00	0,82	0,43	0,43	0,43	0,43	0,43	0,43	0,43	0,43	0,00	0,01	0,00	0,00	0,00			
0,00	0,00	0,00	0,74	0,61	0,61	0,61	0,61	0,61	0,61	0,61	0,61	0,00	0,00	0,00	0,00	0,00			
0,00	0,00	0,01	0,71	0,73	0,73	0,73	0,73	0,73	0,73	0,73	0,73	0,00	0,00	0,00	0,00	0,00			
0,00	0,00	0,05	0,71	0,69	0,69	0,69	0,69	0,69	0,69	0,69	0,69	0,00	0,00	0,00	0,00	0,00			
0,00	0,00	0,12	0,81	0,74	0,74	0,74	0,74	0,74	0,74	0,74	0,74	0,00	0,00	0,00	0,00	0,00			
0,00	0,00	0,18	0,81	0,77	0,77	0,77	0,77	0,77	0,77	0,77	0,77	0,07	0,00	0,00	0,00	0,00			
0,00	0,00	0,25	0,81	0,69	0,69	0,69	0,69	0,69	0,69	0,69	0,69	0,09	0,00	0,00	0,00	0,00			
0,00	0,00	0,32	0,81	0,73	0,73	0,73	0,73	0,73	0,73	0,73	0,73	0,09	0,00	0,00	0,00	0,00			
0,00	0,00	0,34	0,81	0,76	0,76	0,76	0,76	0,76	0,76	0,76	0,76	0,12	0,00	0,00	0,00	0,00			
0,00	0,00	0,28	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,07	0,00	0,00	0,00	0,00			
0,00	0,00	0,36	0,81	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,32	0,00	0,00	0,00	0,00			
0,00	0,00	0,80	0,81	0,83	0,84	0,84	0,84	0,84	0,84	0,84	0,84	1,00	0,00	0,00	0,00	0,00			
0,00	0,36	0,81	0,81	0,76	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,87	0,16	0,00	0,00	0,00			
0,00	0,30	0,85	0,81	0,84	0,77	0,77	0,77	0,77	0,77	0,77	0,77	0,86	0,33	0,00	0,00	0,00			
0,00	0,58	0,80	0,95	0,96	0,85	0,80	0,80	0,80	0,80	0,80	0,80	0,83	0,79	1,00	0,83	0,67	0,59	0,00	
0,00	0,24	0,00	0,00	0,00	0,70	0,70	0,70	0,70	0,70	0,70	0,70	0,97	0,29	0,00	0,09	0,00	0,13	0,00	
0,00	0,00	0,00	0,00	0,00	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,99	0,40	0,00	0,00	0,00	0,00	0,00	
0,00	0,02	0,01	0,00	0,00	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,42	0,00	0,02	0,00	0,02	0,00	
0,00	0,00	0,00	0,00	0,00	0,74	0,74	0,74	0,74	0,74	0,74	0,74	0,43	0,00	0,01	0,00	0,00	0,00	0,00	
0,00	0,00	0,00	0,00	0,00	0,71	0,71	0,71	0,71	0,71	0,71	0,71	0,57	0,00	0,03	0,00	0,00	0,00	0,00	
0,00	0,00	0,00	0,00	0,18	0,12	0,01	0,55	0,48	0,00	0,00	0,55	0,44	0,40	0,22	0,00	0,02	0,00	0,00	0,00



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	54	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	218	211	168	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	165	252	236	231	149	178	228	43	95	234
190	215	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	127	102	36	181	258	224
190	214	173	66	103	143	95	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	54	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	168	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	165	252	236	231	149	178	228	43	95	234
190	215	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	127	102	36	181	258	224
190	214	173	66	103	143	95	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



...

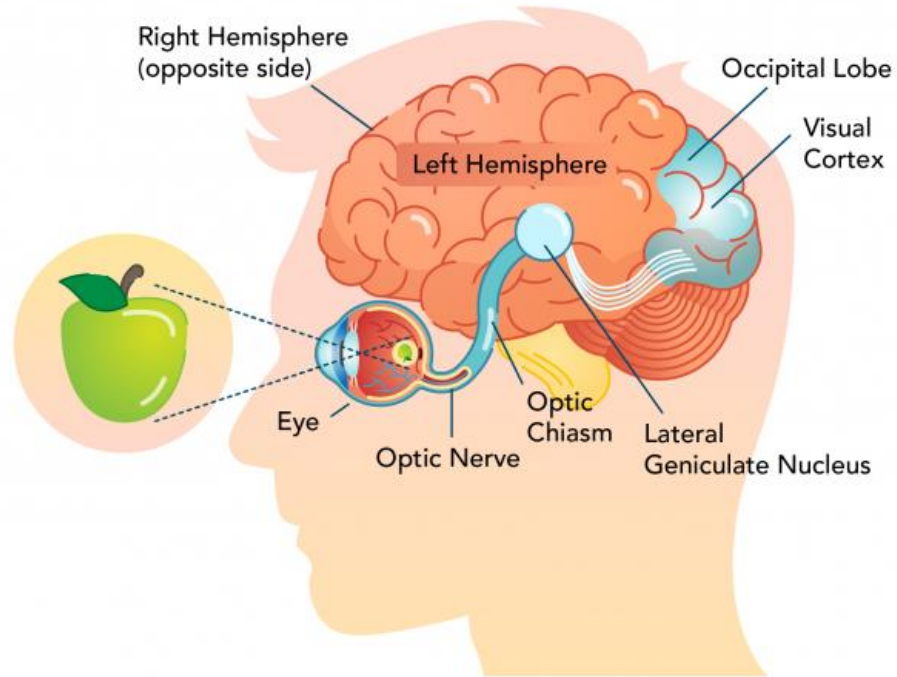
# Apa yang dilihat oleh manusia?

...

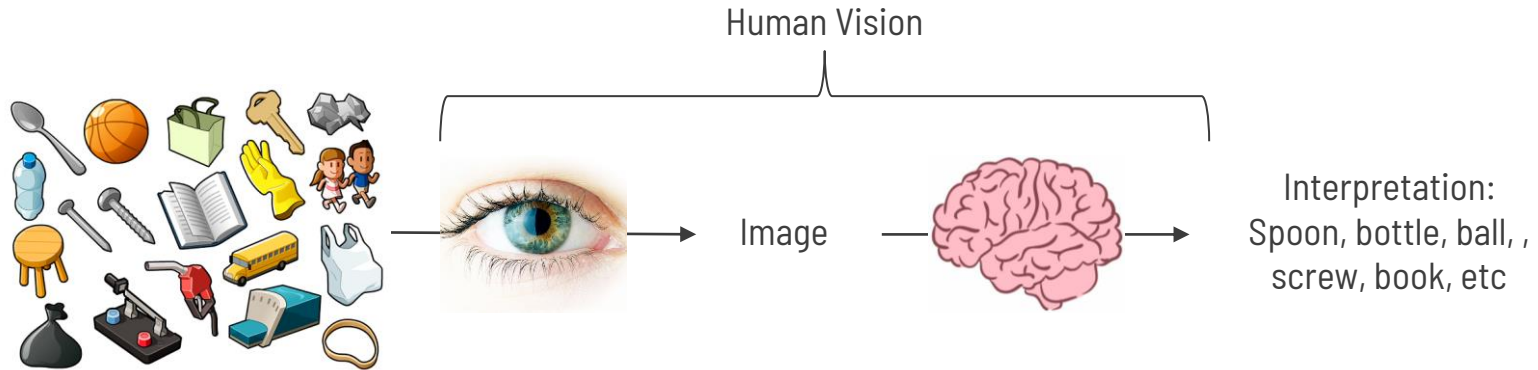




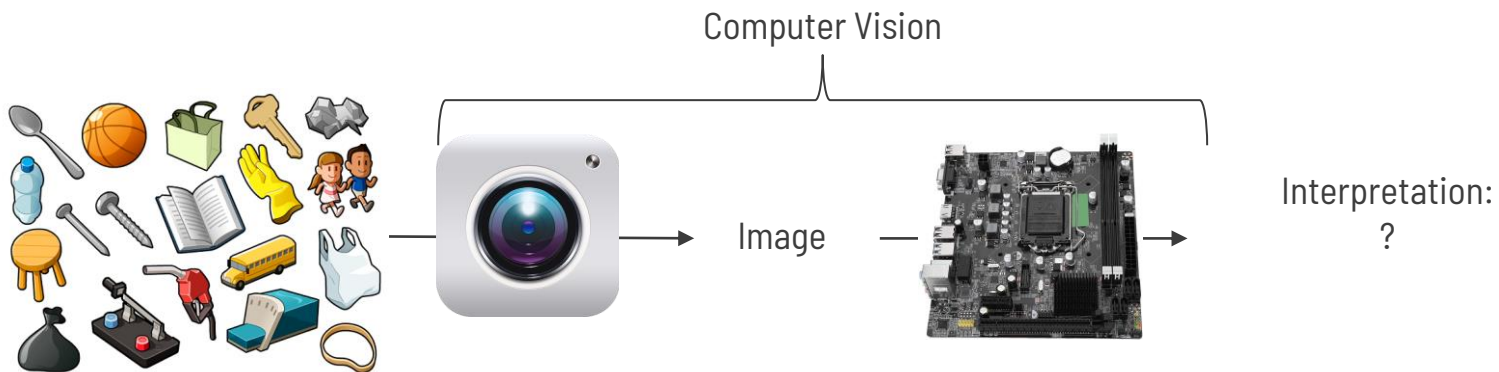
# Bagaimana manusia melihat?



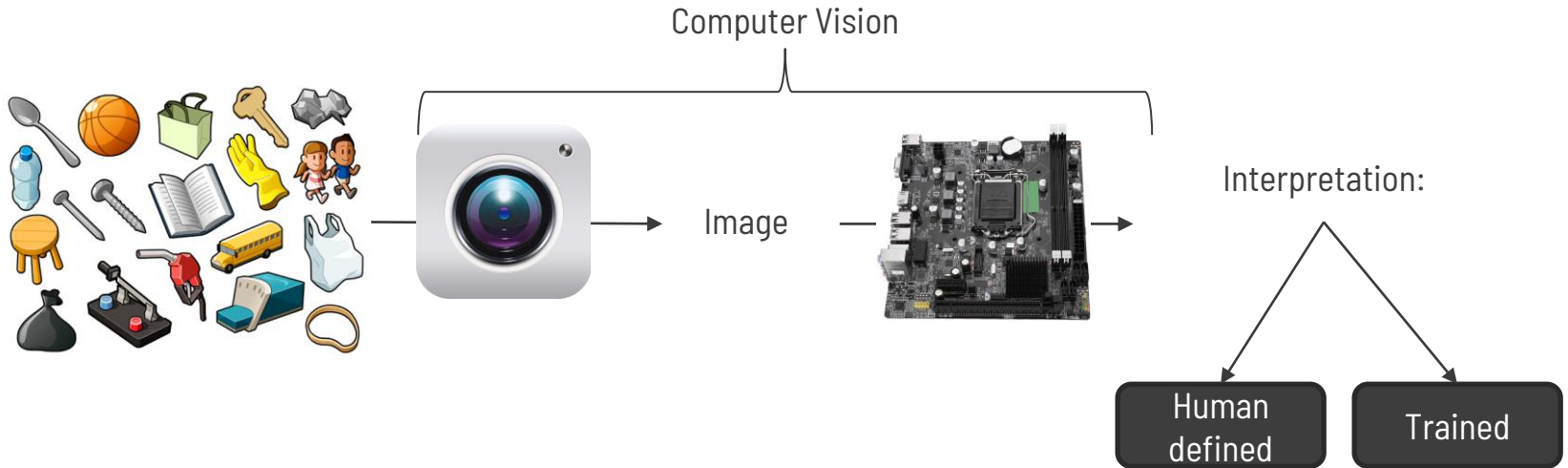
# Bagaimana manusia melihat?



# Bagaimana manusia melihat?

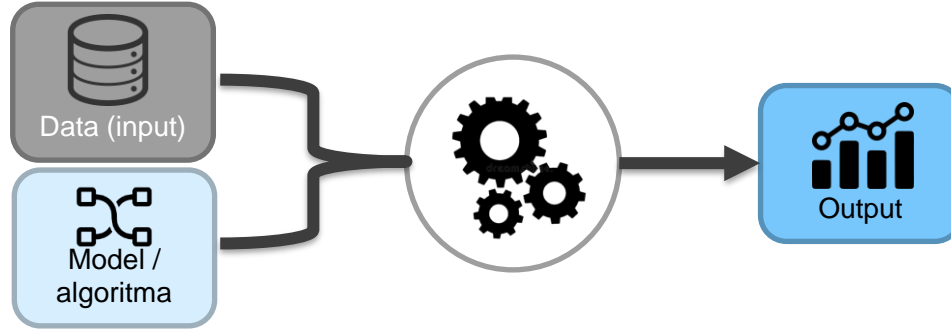


# Bagaimana manusia melihat?

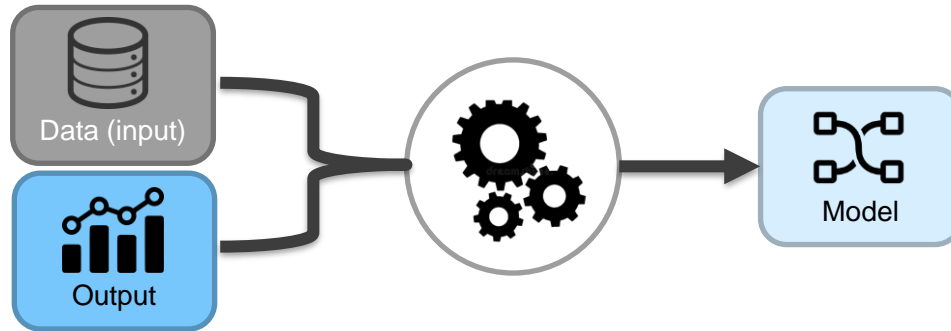


# Recall Machine Learning

Human Designed Program

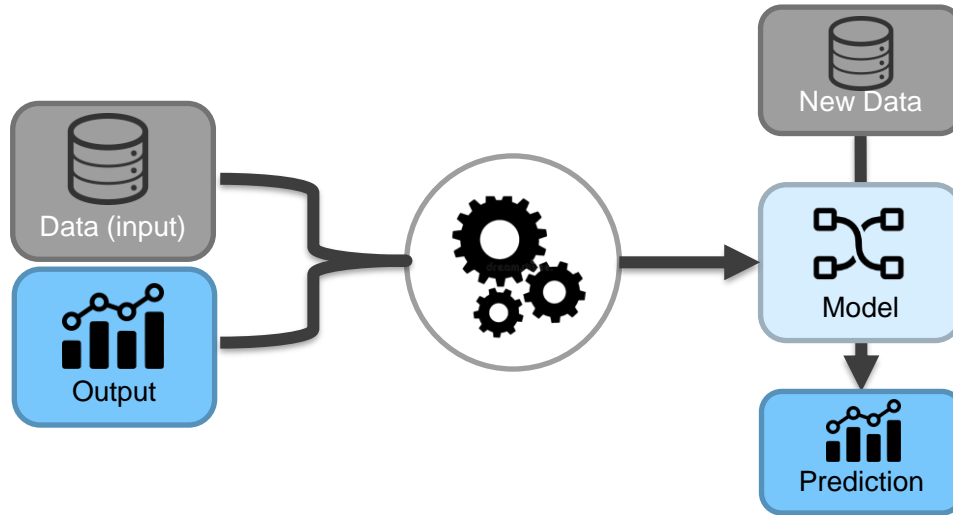


Trained Program

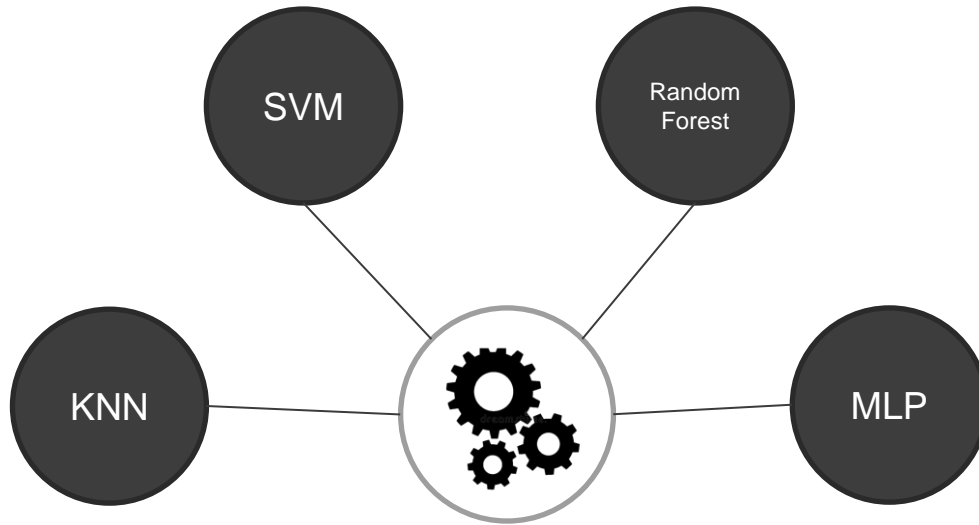


# Recall Machine Learning

Trained  
Program



# Recall Machine Learning

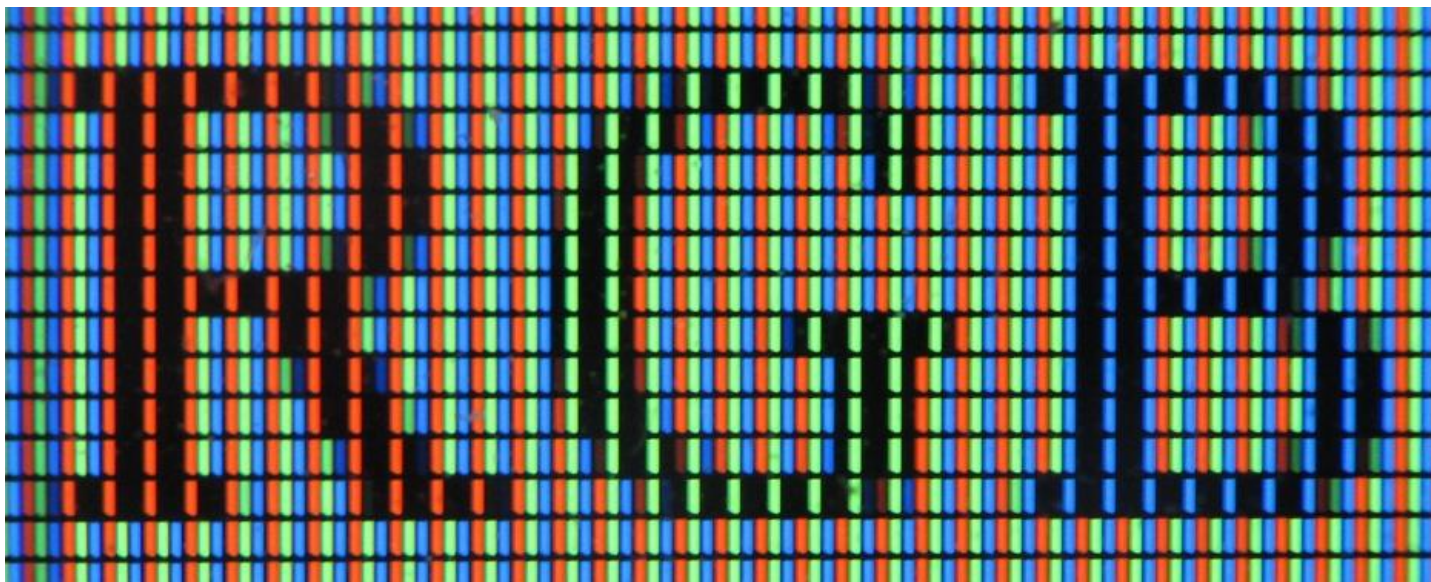






How to obtain features in **image data**?

Computer only captures pixels



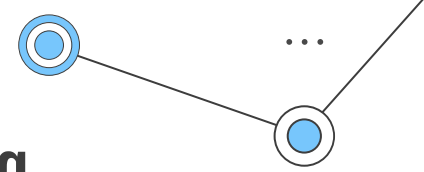
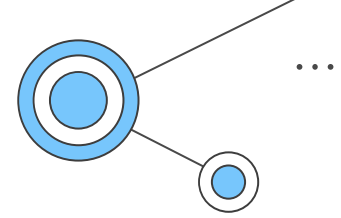
# Simple example: color marker tracking

**How to detect the ball  
in this image?**

Just find any pixel that  
equals or is close to this  
color:

 **R:234, G:31, B:63**



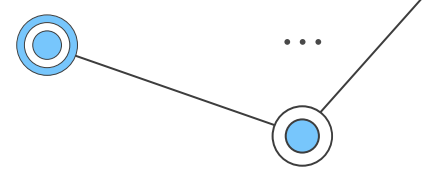
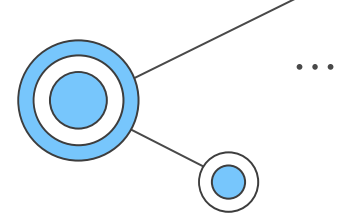


# Simple example: color marker tracking

## Weakness:

- Depends on individual pixel
- Possibility of other objects with same color
- Lighting may affect color

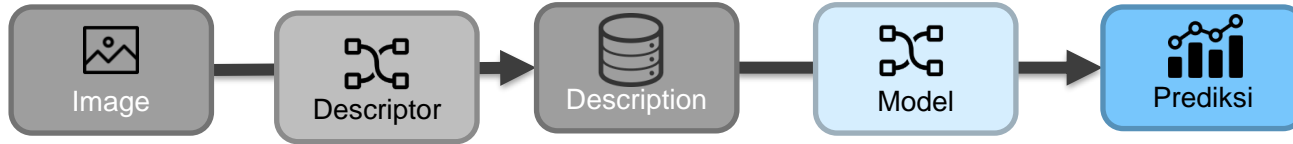




Features in images depends on  
**combination of pixels**

**Deep features:  
gradient, texture, shape, etc**

# Visual Descriptor



Pixel-Level Descriptor

Patch-level (Local Descriptor)

Region-Level

Global Descriptor

...

# Visual Descriptor

## Color

Dominant color descriptor (DCD)

Scalable color descriptor (SCD)

Color structure descriptor (CSD)

Color layout descriptor (CLD)

Group of frame (GoF) or group-of-pictures (GoP)

## Texture

Homogeneous texture descriptor (HTD)

Texture browsing descriptor (TBD)

Edge histogram descriptor (EHD)

## Shape

Region-based shape descriptor (RSD)

Contour-based shape descriptor (CSD)

3-D shape descriptor (3-D SD)

...

## Motion

Motion activity descriptor (MAD)

Camera motion descriptor (CMD)

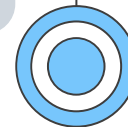
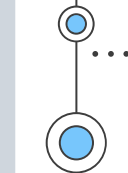
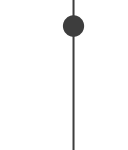
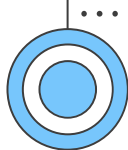
Motion trajectory descriptor (MTD)

Warping and parametric motion descriptor (WMD and PMD)

## Location

Region locator descriptor (RLD)

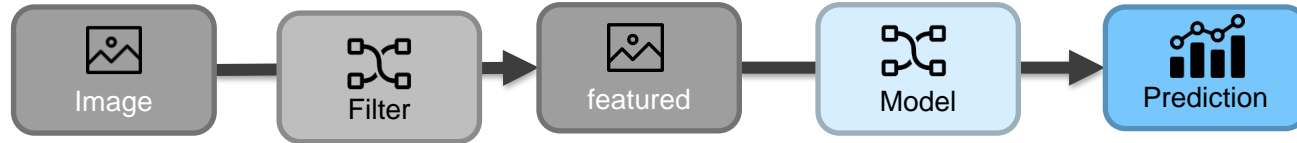
Spatio temporal locator descriptor (STLD)



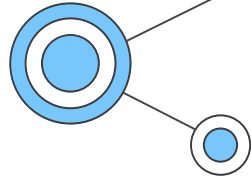


# Alternative: Using filter

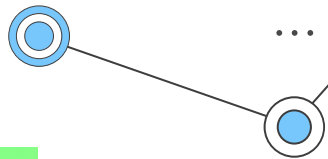
**Filtering** is one of image processing techniques



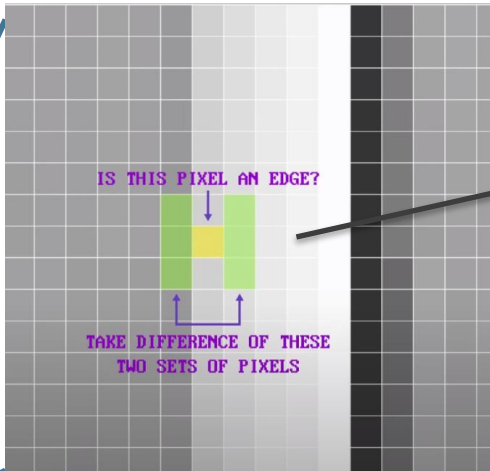
...



...



...

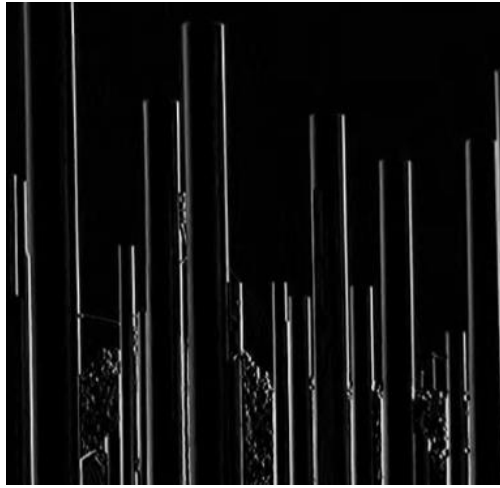


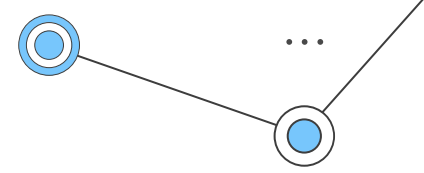
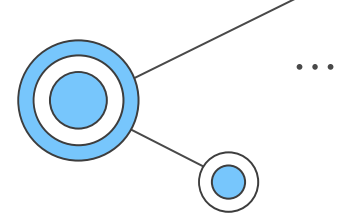
-1	0	1
-1	0	1
-1	0	1

**Prewitt Operator  
(Edge Detection)**



**Convolution /  
Filtering**





Input image

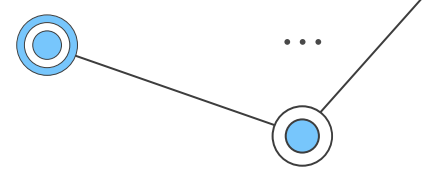
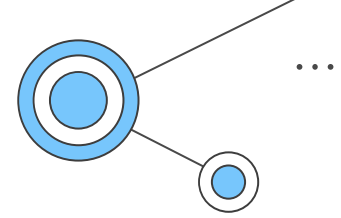


Convolution  
Kernel

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

Feature map





**How do we choose the right  
kernel to obtain useful  
information?**

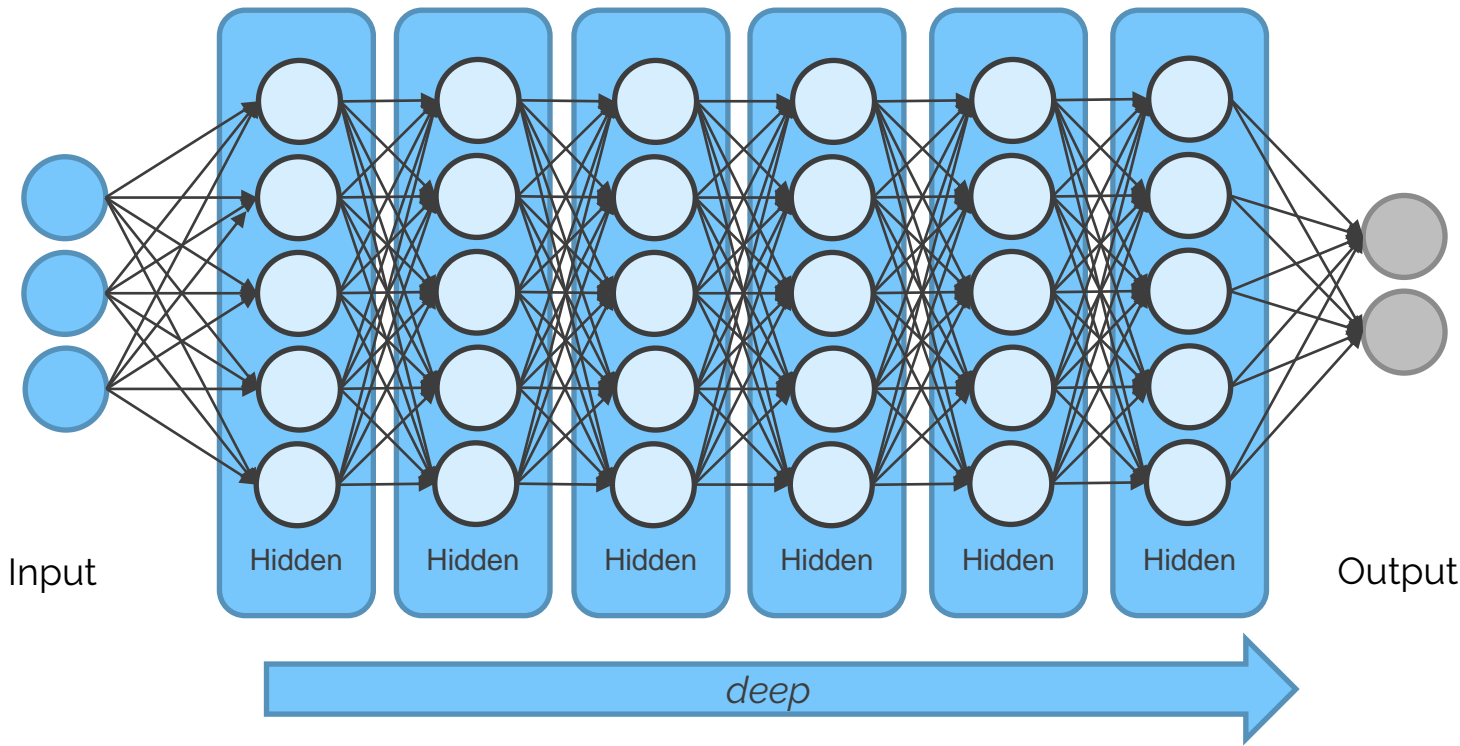
# Viola-Jones Framework (2001)

**First Facial  
Recognition**

**Scan through  
images using  
combination of  
kernels**

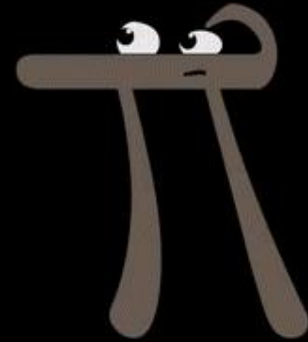
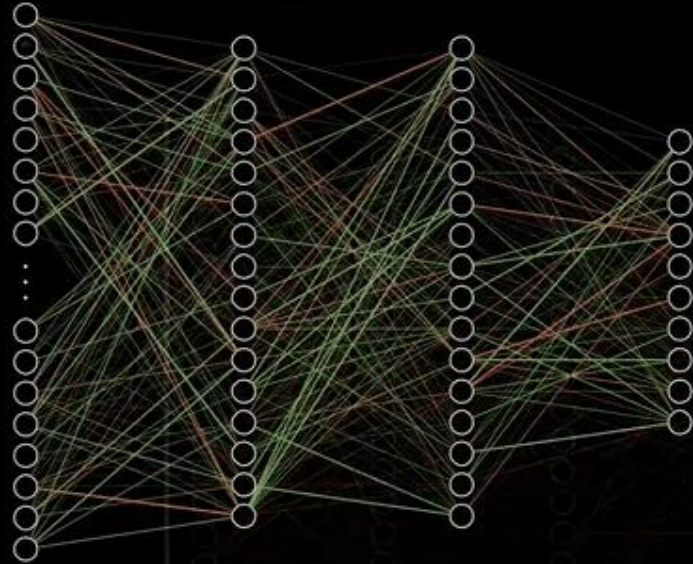


# Deep Neural Network



Training in progress.

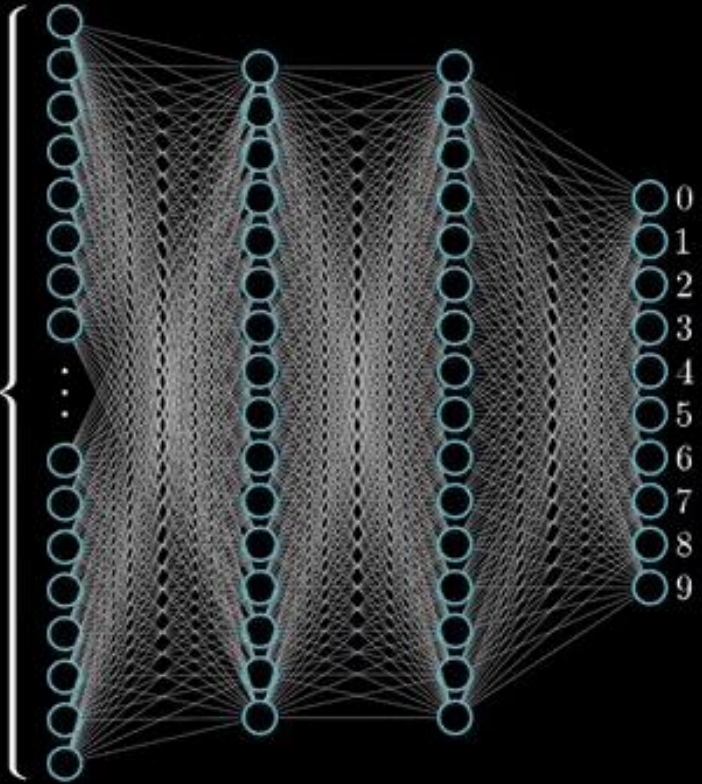
1 → 1







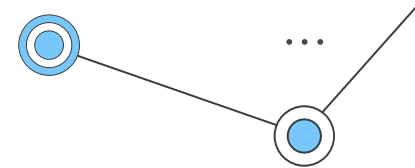
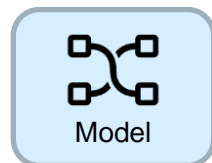
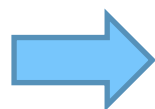
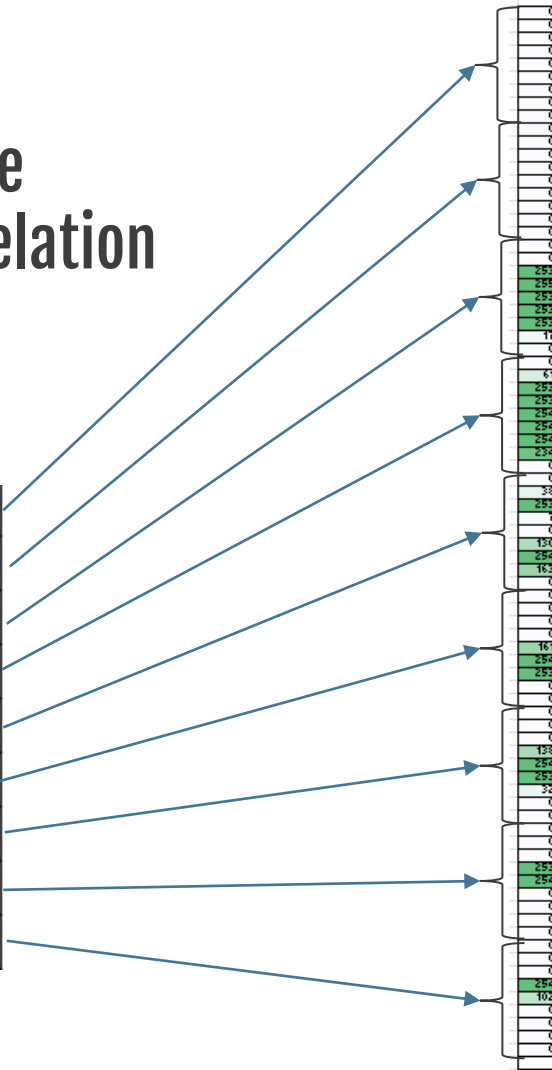
784



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

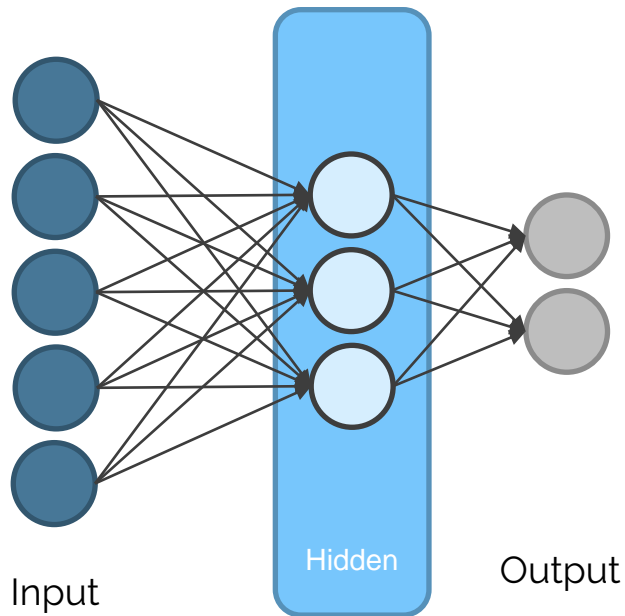
# Direct Flattening lose information about relation between rows

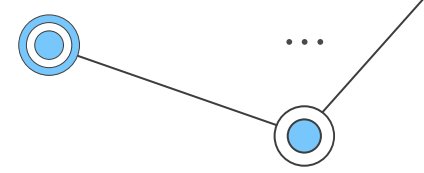
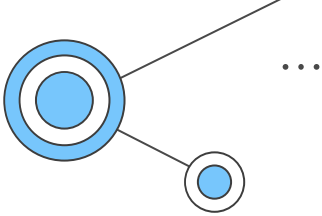
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	253	255	253	253	253	17	0	0
0	61	253	253	254	254	254	234	0	0
0	38	253	7	0	130	254	163	0	0
0	0	0	0	161	254	253	0	0	0
0	0	0	138	254	253	32	0	0	0
0	0	0	253	254	0	0	0	0	0
0	0	0	254	102	0	0	0	0	0



# Understanding Convolution

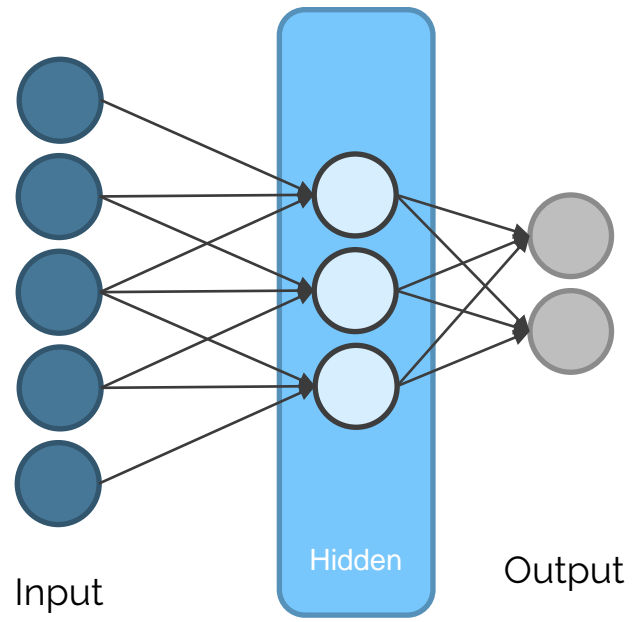
**Fully connected:**  
To detect every possible combination of features

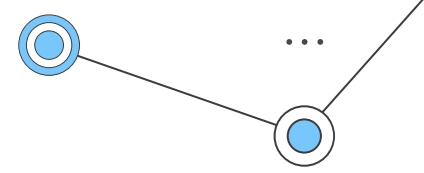
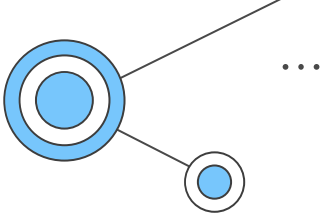




**Why not just like this?**

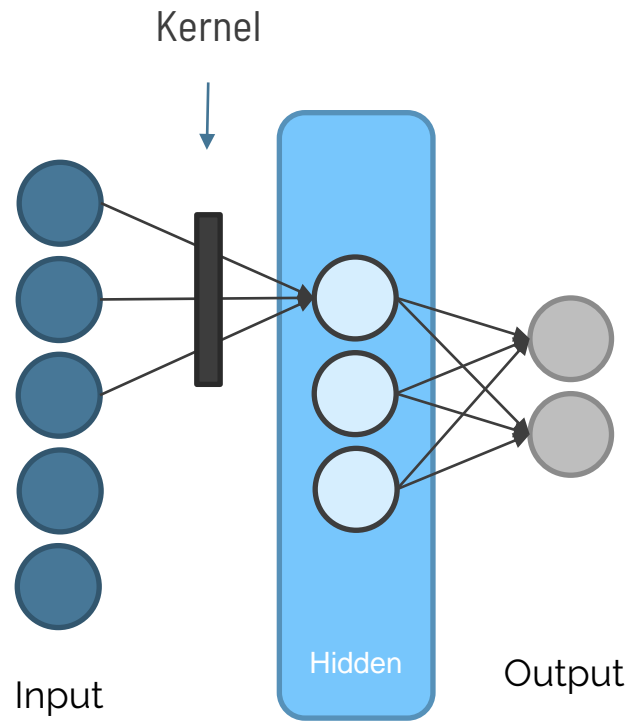
Not every feature is related

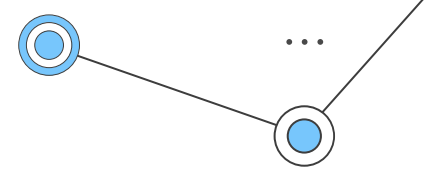
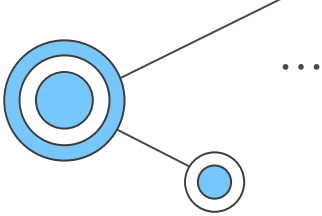




**If we assume that only adjacent features are related,**

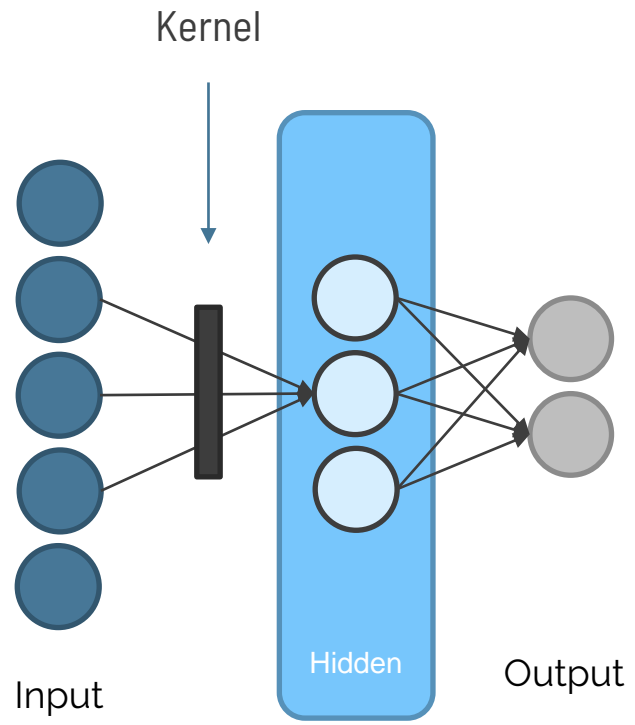
**We can imagine a sliding kernel to compute it**

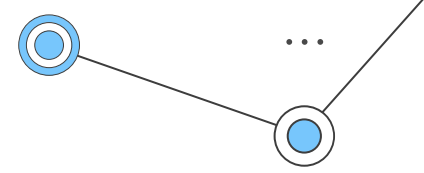
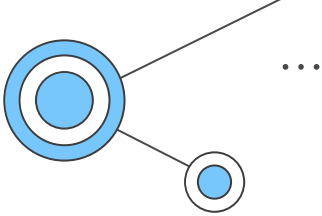




**If we assume that only adjacent features are related,**

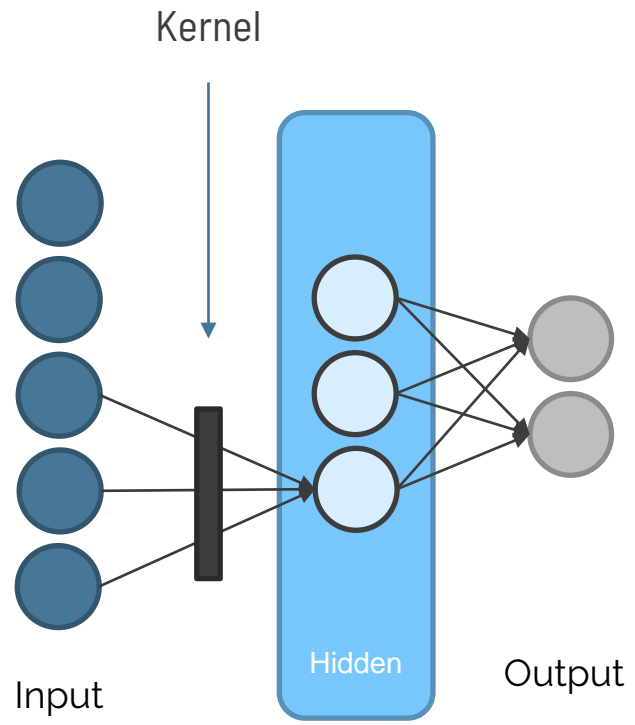
**We can imagine a sliding kernel to compute it**



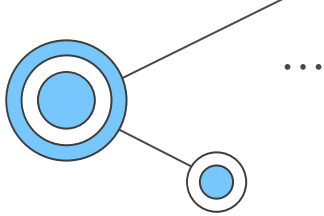


**If we assume that only adjacent features are related,**

**We can imagine a sliding kernel to compute it**

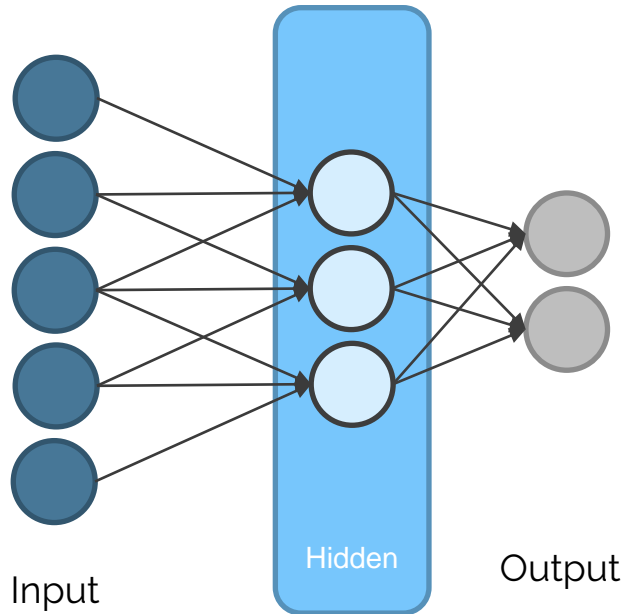




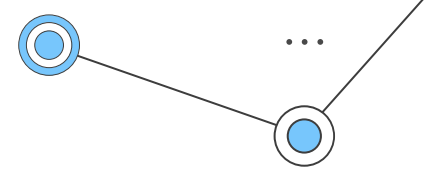


This process is called (1-D) Convolution

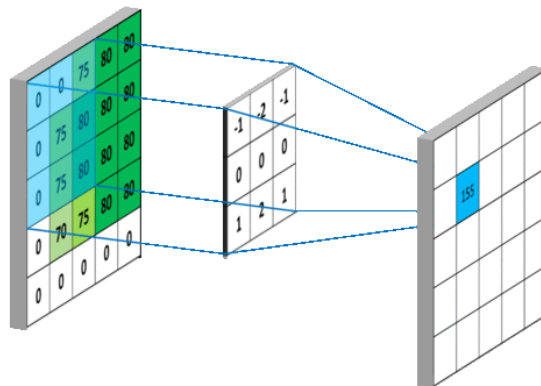
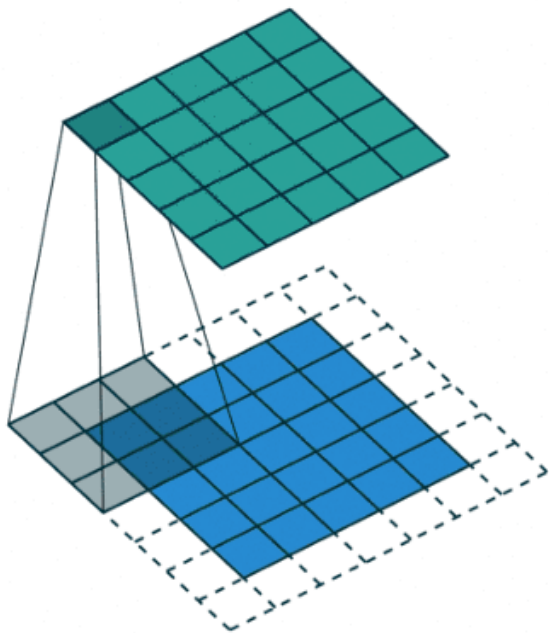
And the model is called (1-D) Convolutional Neural Network (CNN)



In neural network, the kernel contains weights of the layer, so it is trained, rather than predefined



# The same is true for 2D Convolution



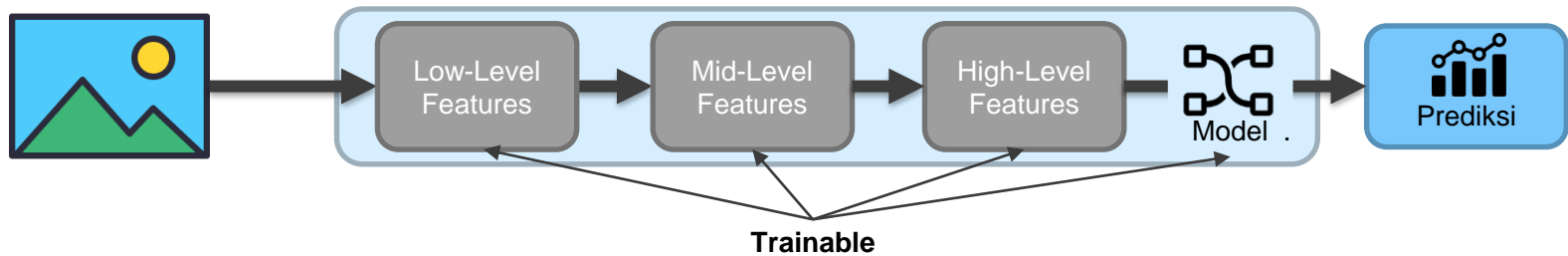
Information in a patch can be obtained

Also, rather than predefine the filter/kernel, it can be trained

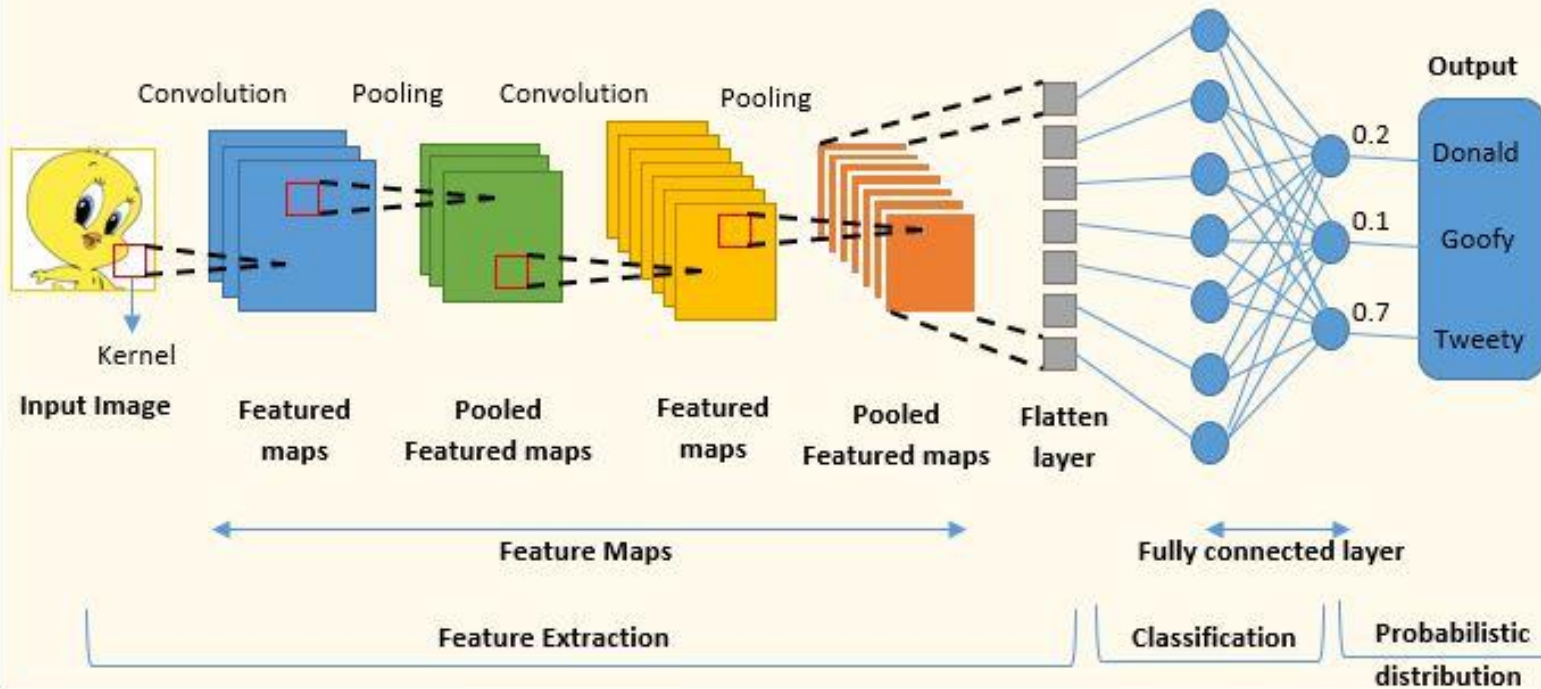
# Traditional Machine Learning

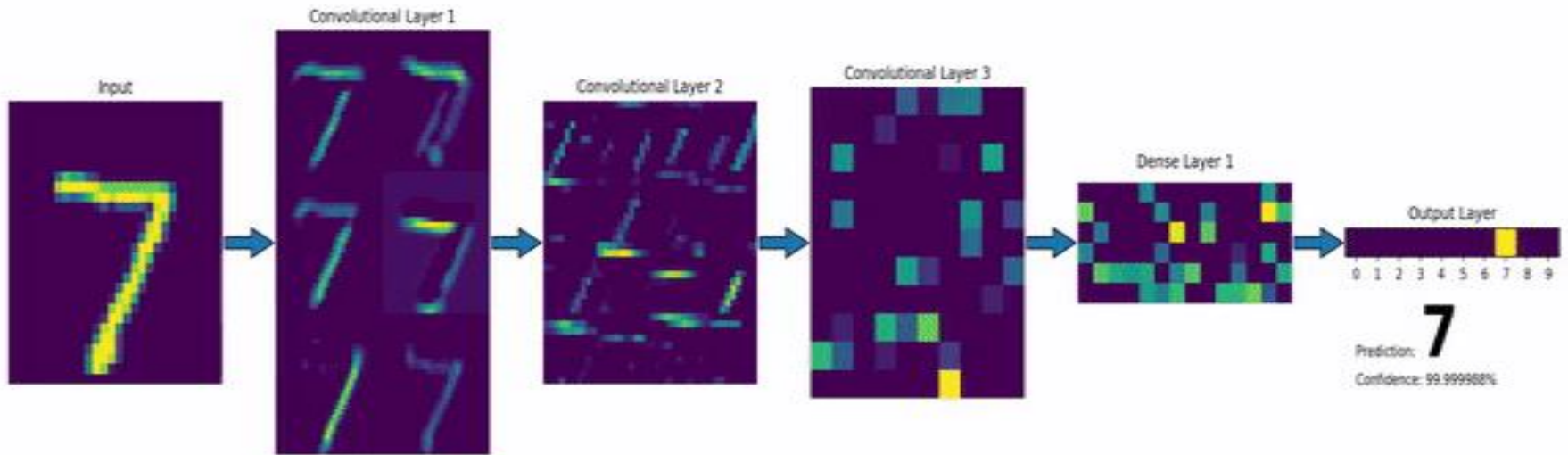
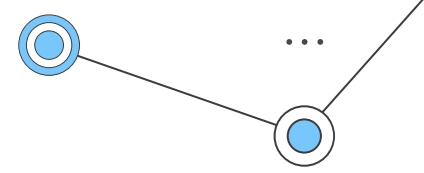
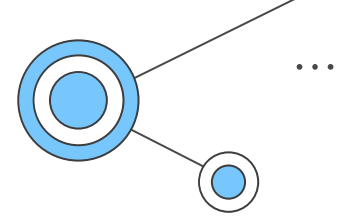


# Deep Learning

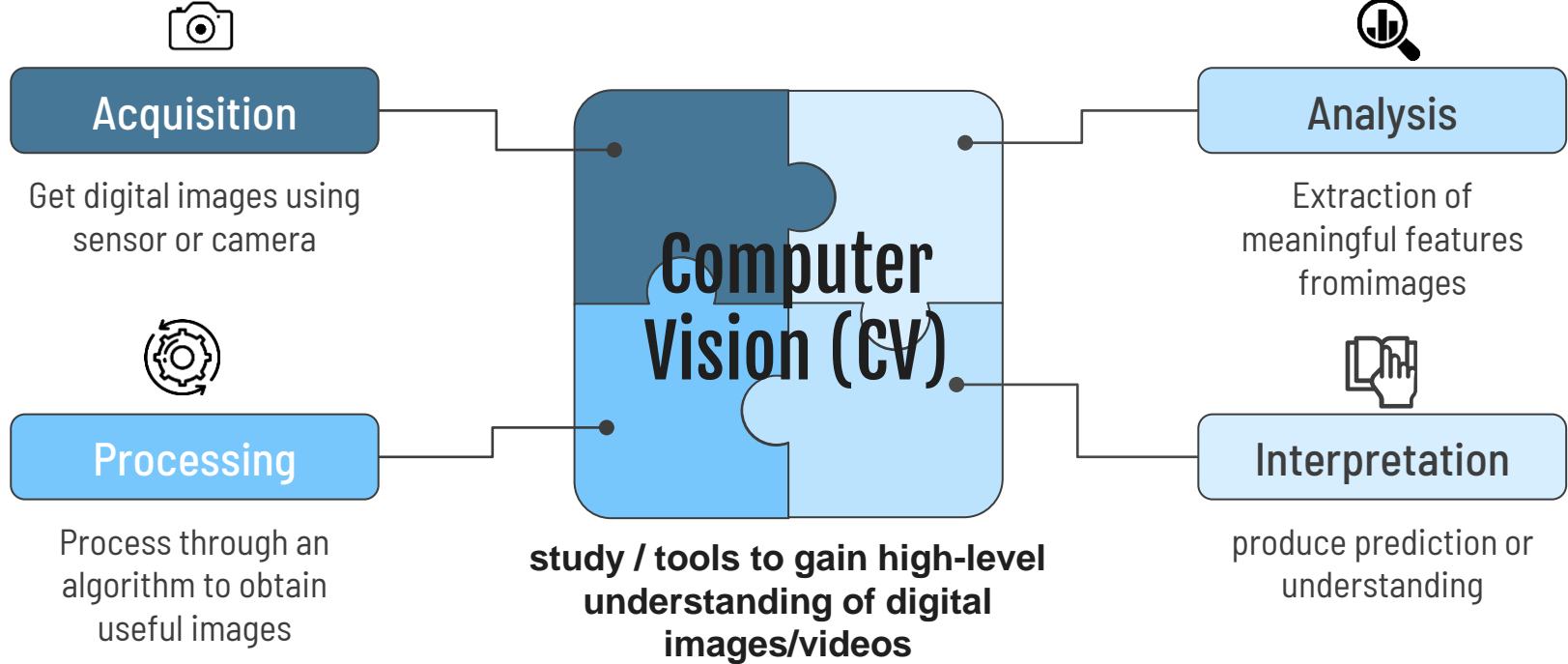


## A Typical Convolutional Neural Network (CNN)

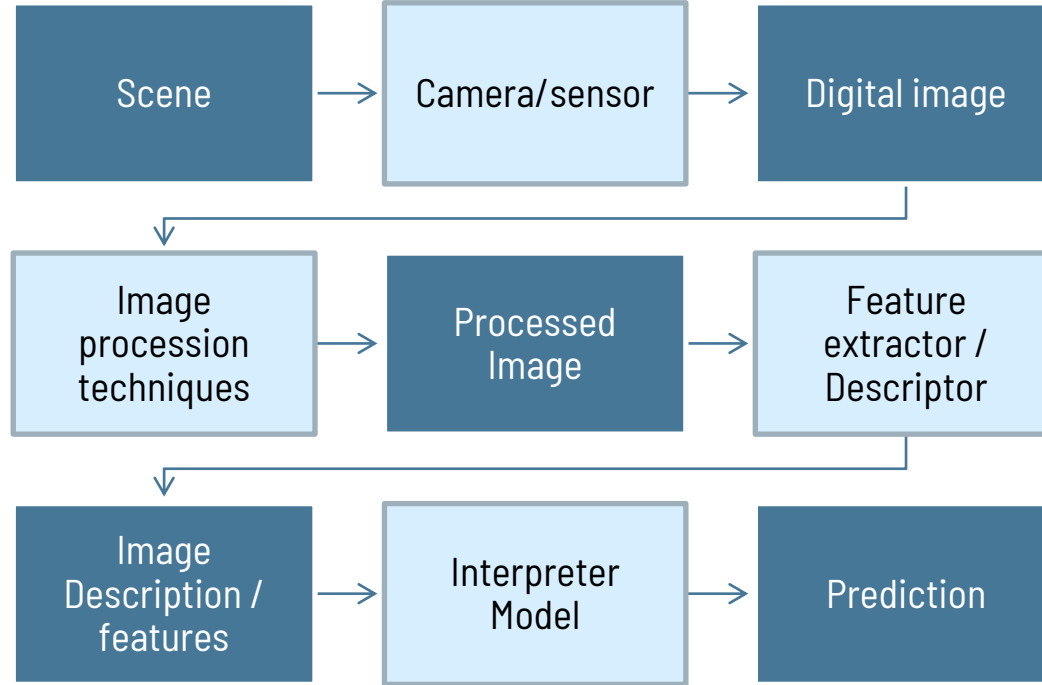




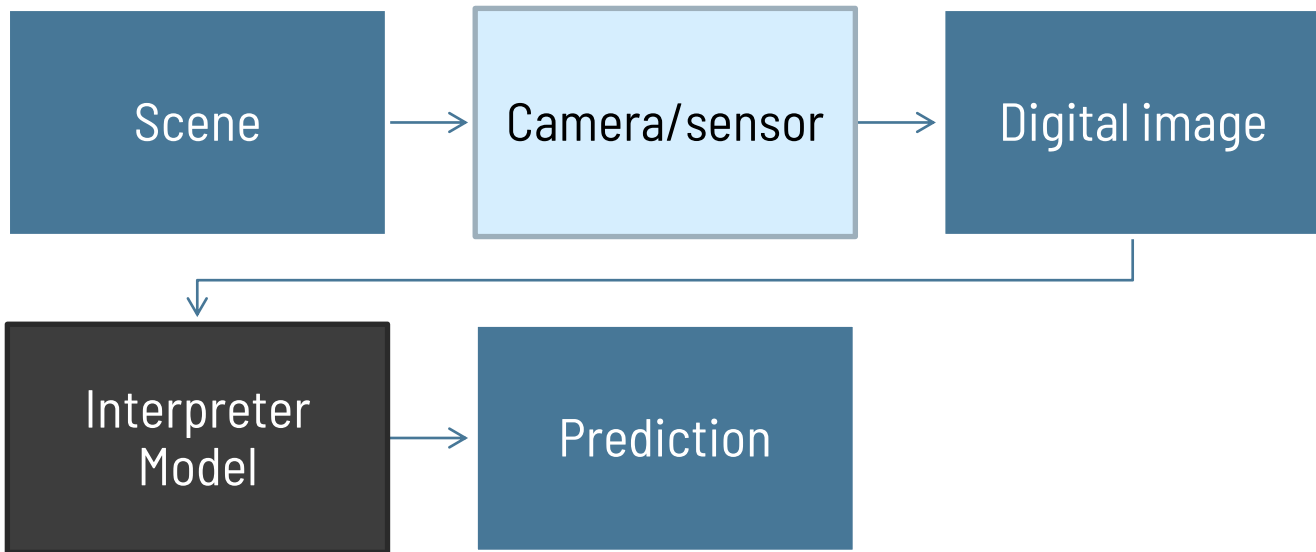
# So, what is CV?



# Full Process of CV

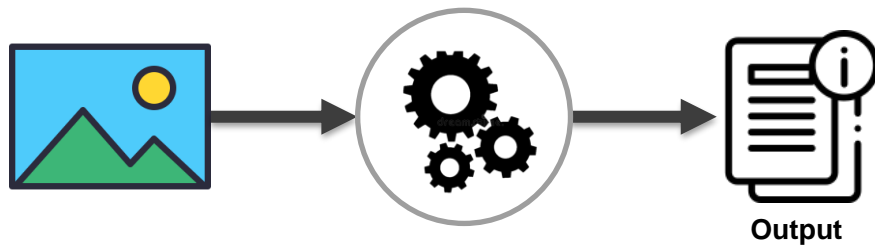


# CV in Deep Learning Paradigm

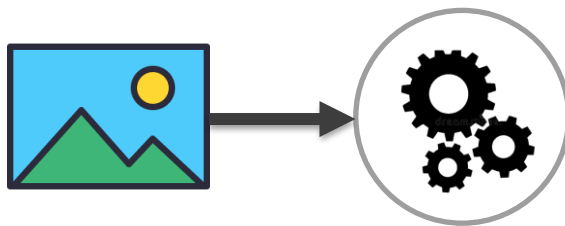


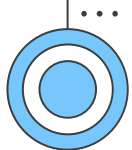


# Deep Learning CV



However, recently any model that is **trained** by images is CV model





...



# Computer Vision

## Recognition Model

Object Detection

Image Classification

Image Segmentation

Pose Estimation

## Generative Model

Image Generation

Style Transfer

Image Enhancement

Image to Image translation

Drawing

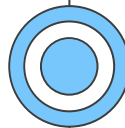
## Video

Anomaly Detection

Activity Recognition

Object Tracking

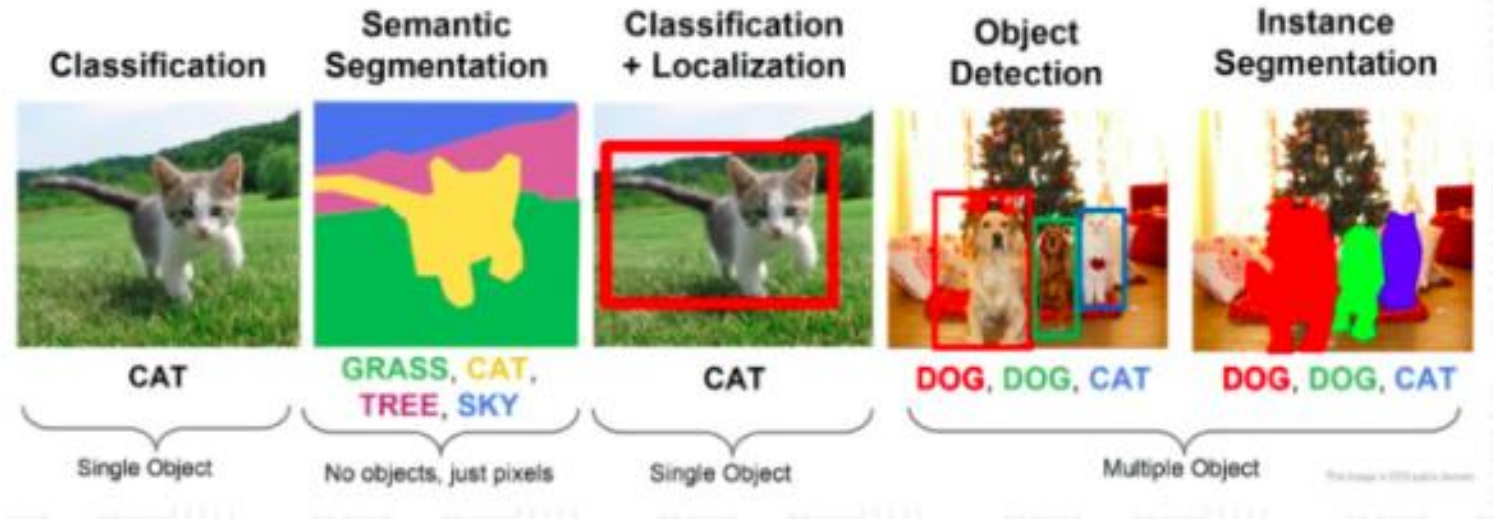
...



...



# Recognition Model

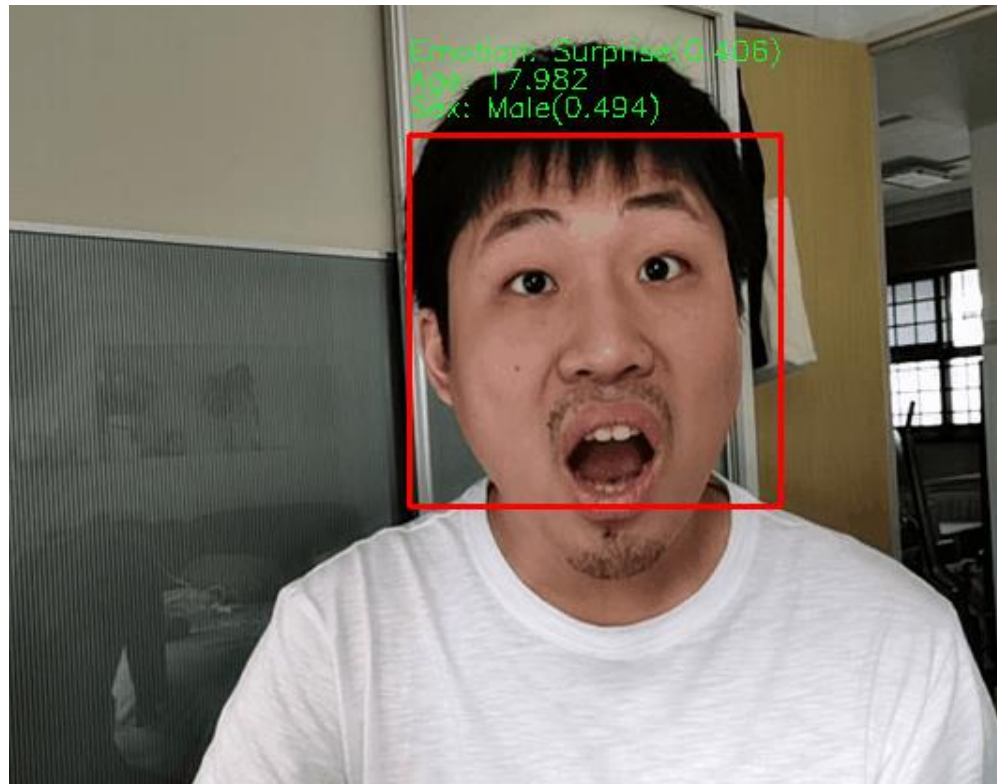


...

# Character Detection



# Emotion Detection



# Instance Segmentation

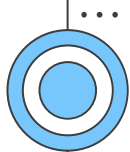
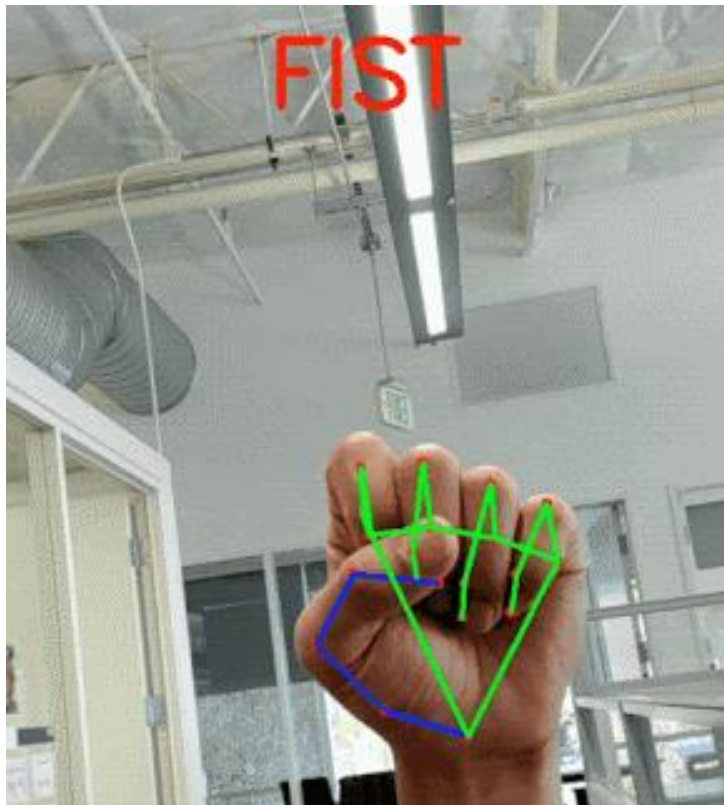




# Pose Estimation



# Sign Language Recognition







Avatar movie



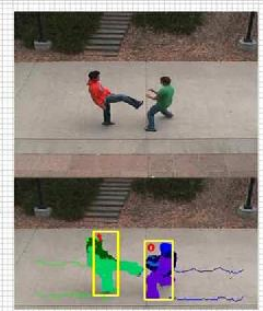
The Digital Dance Lesson  
MOVIES, 3DTV, &  
ANIMATIONS



Microsoft Kinect  
GESTURE-BASED  
INTERACTIVE GAMES



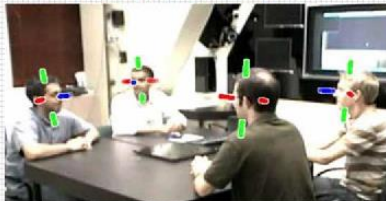
Interactive Balloon Game



SURVEILLANCE



PHYSICAL THERAPY



SMART  
ENVIRONMENTS

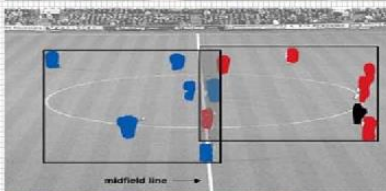
INTELLIGENT  
DRIVER  
ASSISTANCE



"Keeping hands on the  
wheel & eyes on the road"



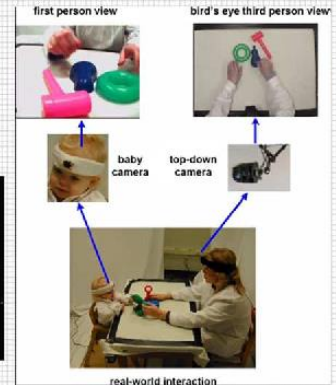
Intelligent brake assistance



VIDEO  
ANNOTATION

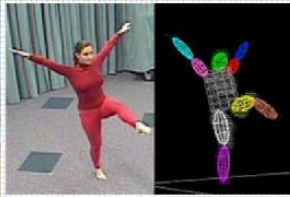
ADVANCE HCI

AUTONOMOUS  
MENTAL  
DEVELOPMENT

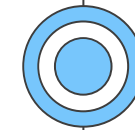
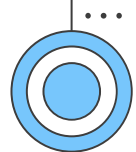


SPORT MOTION  
ANALYSIS

ASSISTED LIVING



Fall detection







# Generative Models

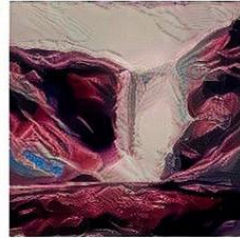
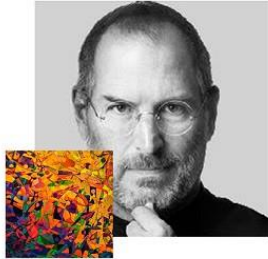
Source A: gender, age, hair length, glasses, pose



Source B:  
everything  
else

Result of combining A and B

# Style Transfer



# Image Restoration/Enhancement





# Image Translation

Zebras  $\rightleftarrows$  Horses

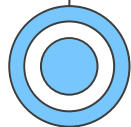
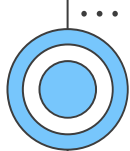
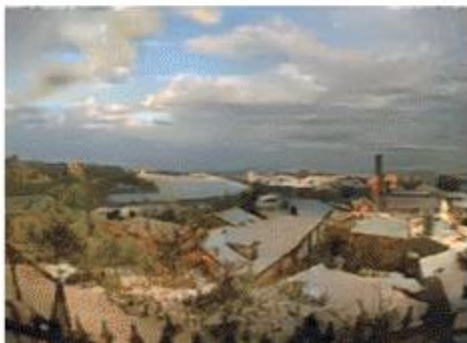
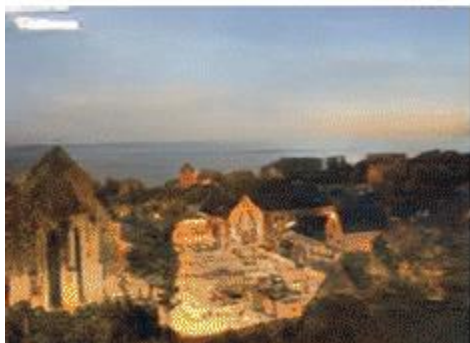


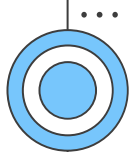
zebra  $\rightarrow$  horse



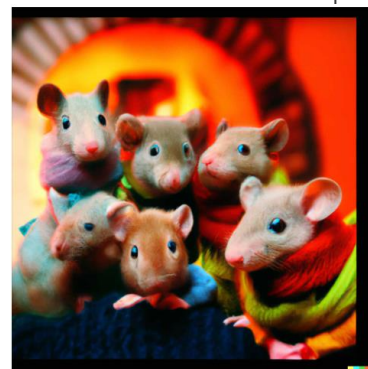
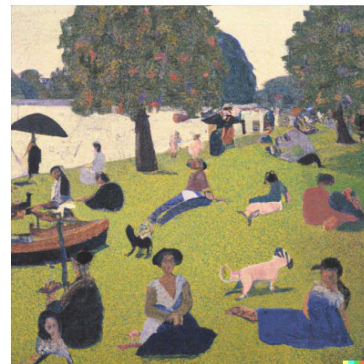
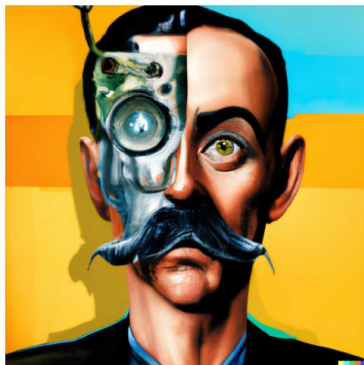
horse  $\rightarrow$  zebra

# Image Translation



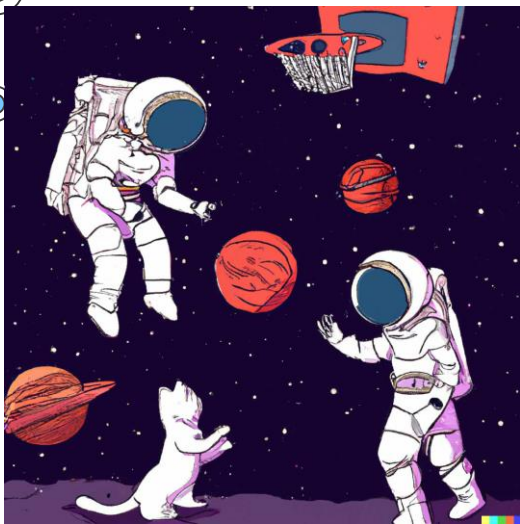


# Image Translation (Dall-E)





# Image Generation from Text (Dall-E)



An astronaut playing basketball with cats in space as a children's book illustration

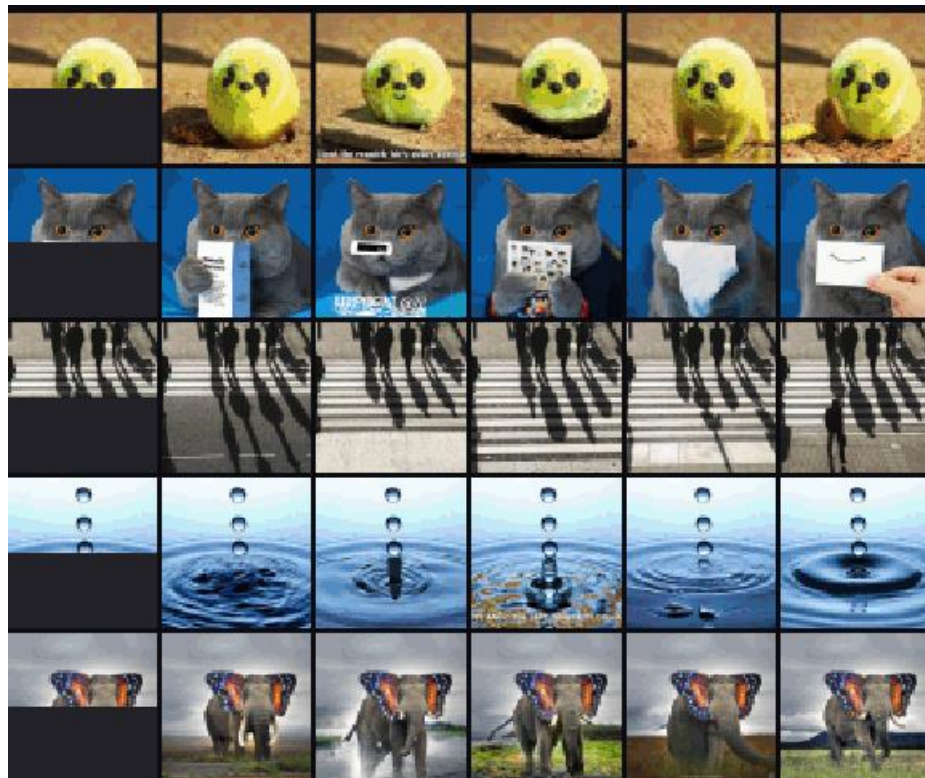


A bowl of soup that looks like a monster made out of plasticine



Teddy bears shopping for groceries in ancient Egypt

# Image Completion



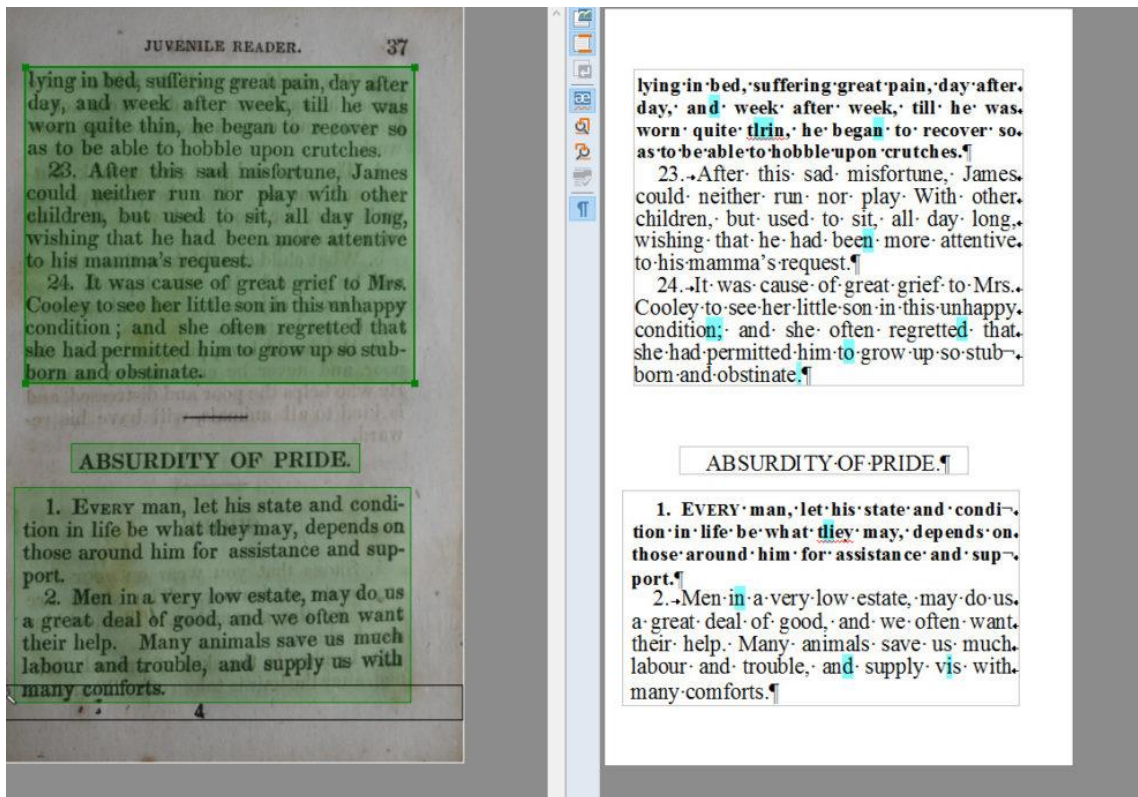


# Some applicative use case

...

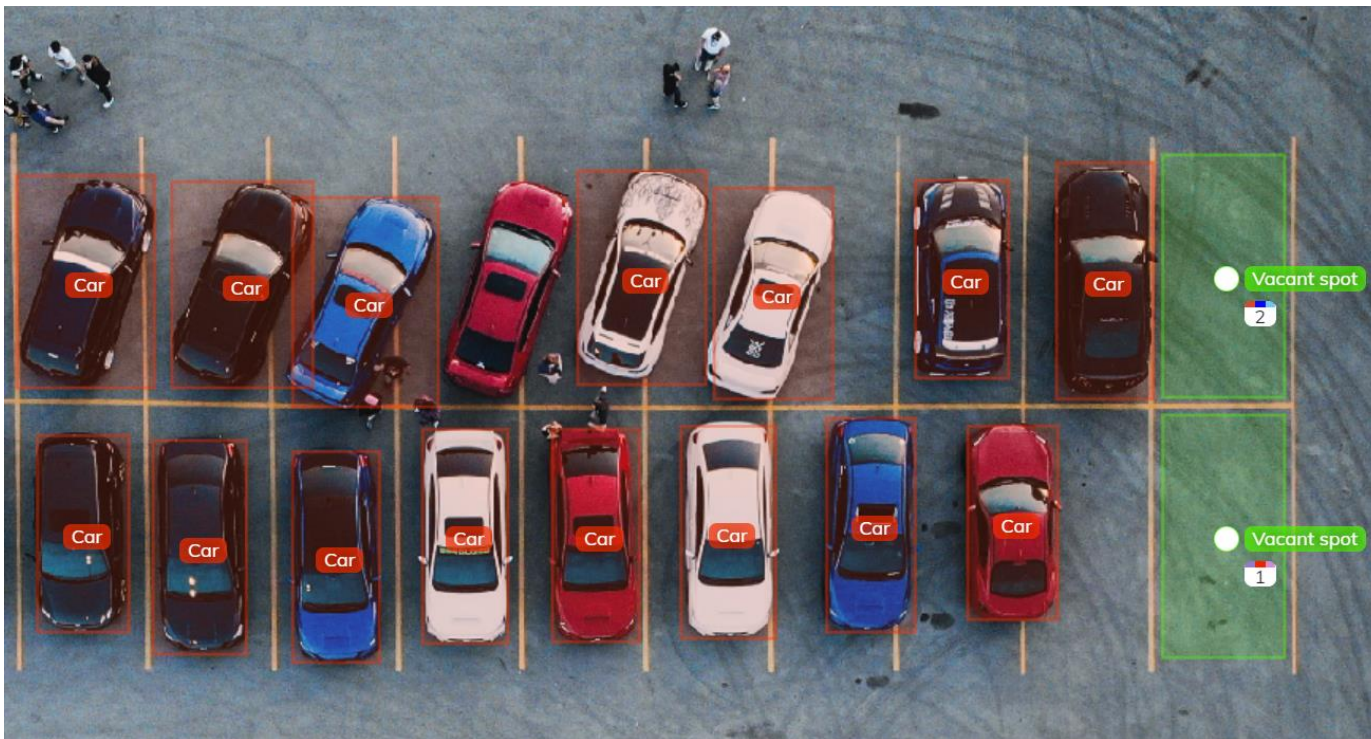


# Optical Character Recognition (OCR)

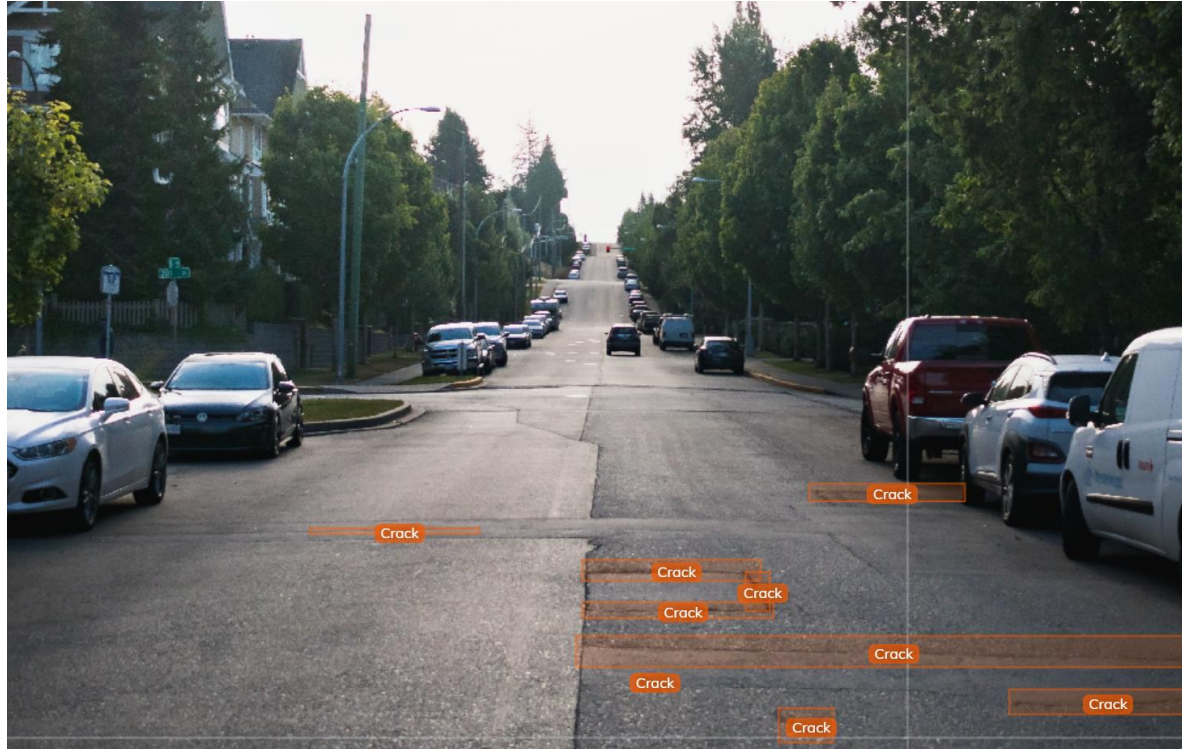




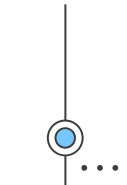
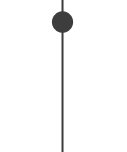
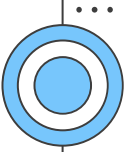
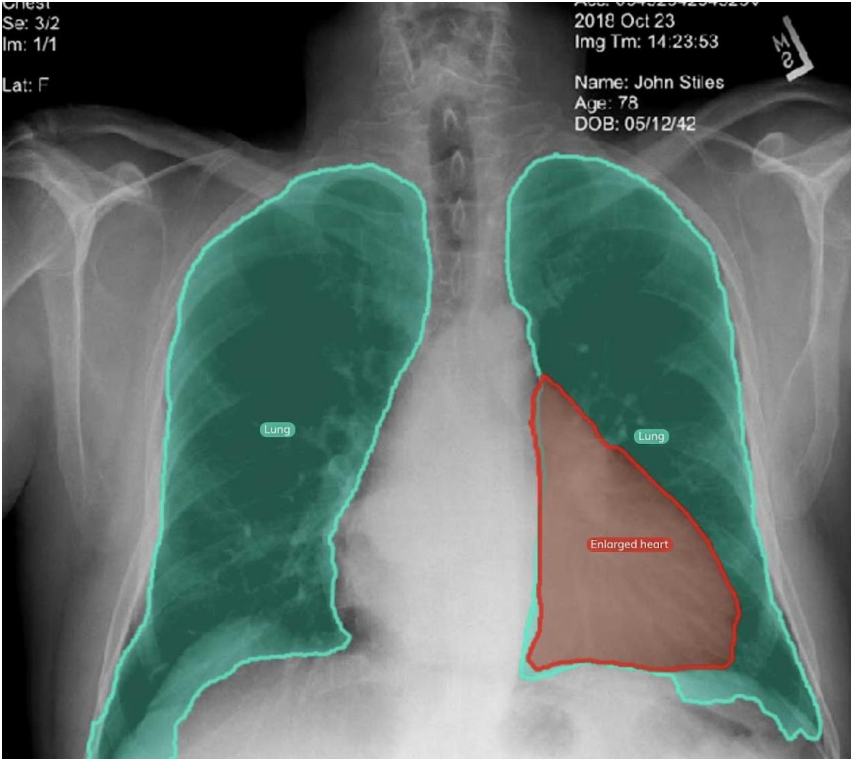
# Parking Occupancy Detection



# Road condition monitoring



# Medical Imaging





# Defect Inspection

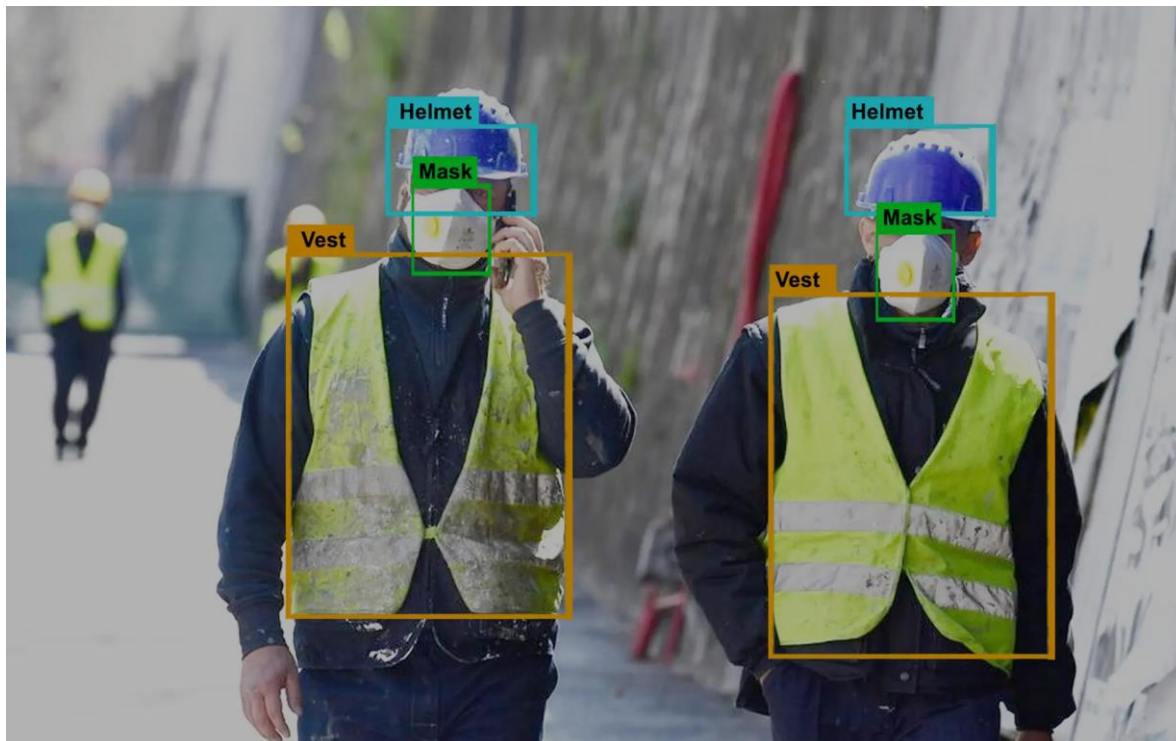




# Barcode & Label Reader



# PPE Detection

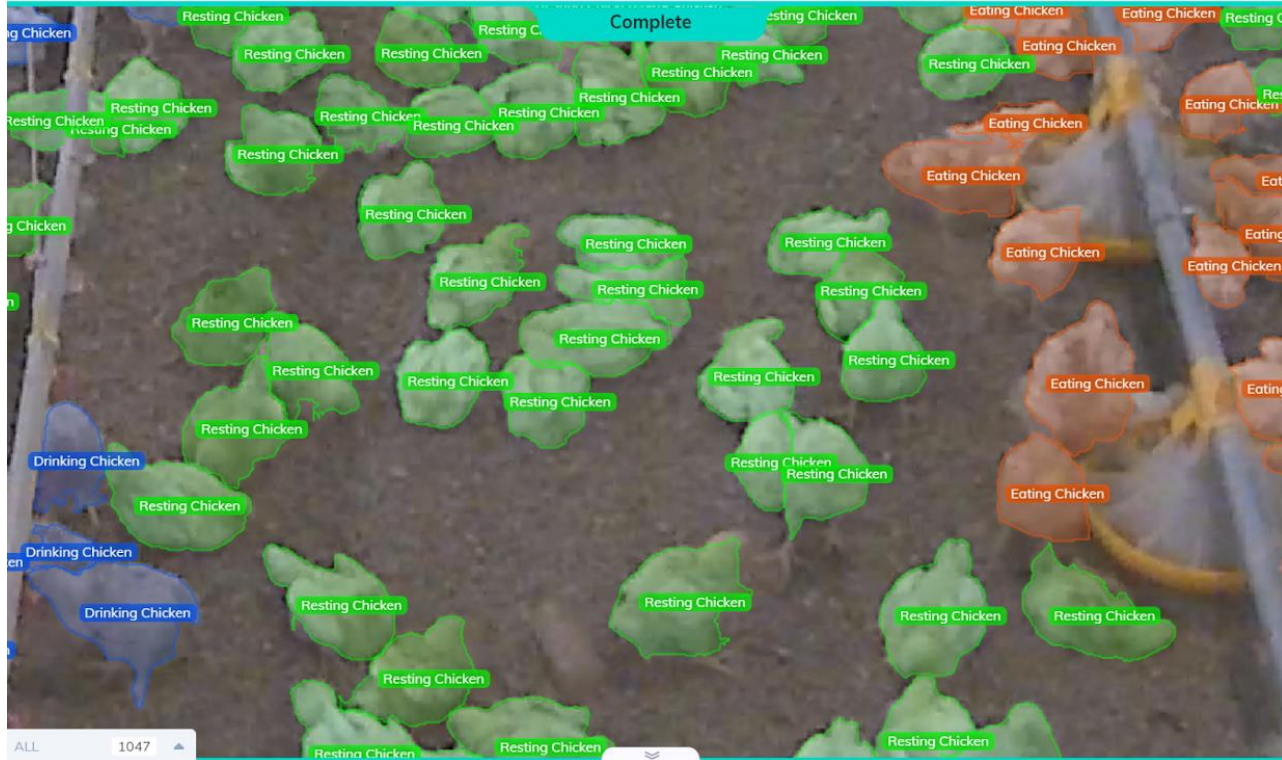


# Field Monitoring





# Livestock Monitoring



# Plant Disease Detection



# Surveillance Monitoring





# People Counting



# Automatic Replenishment



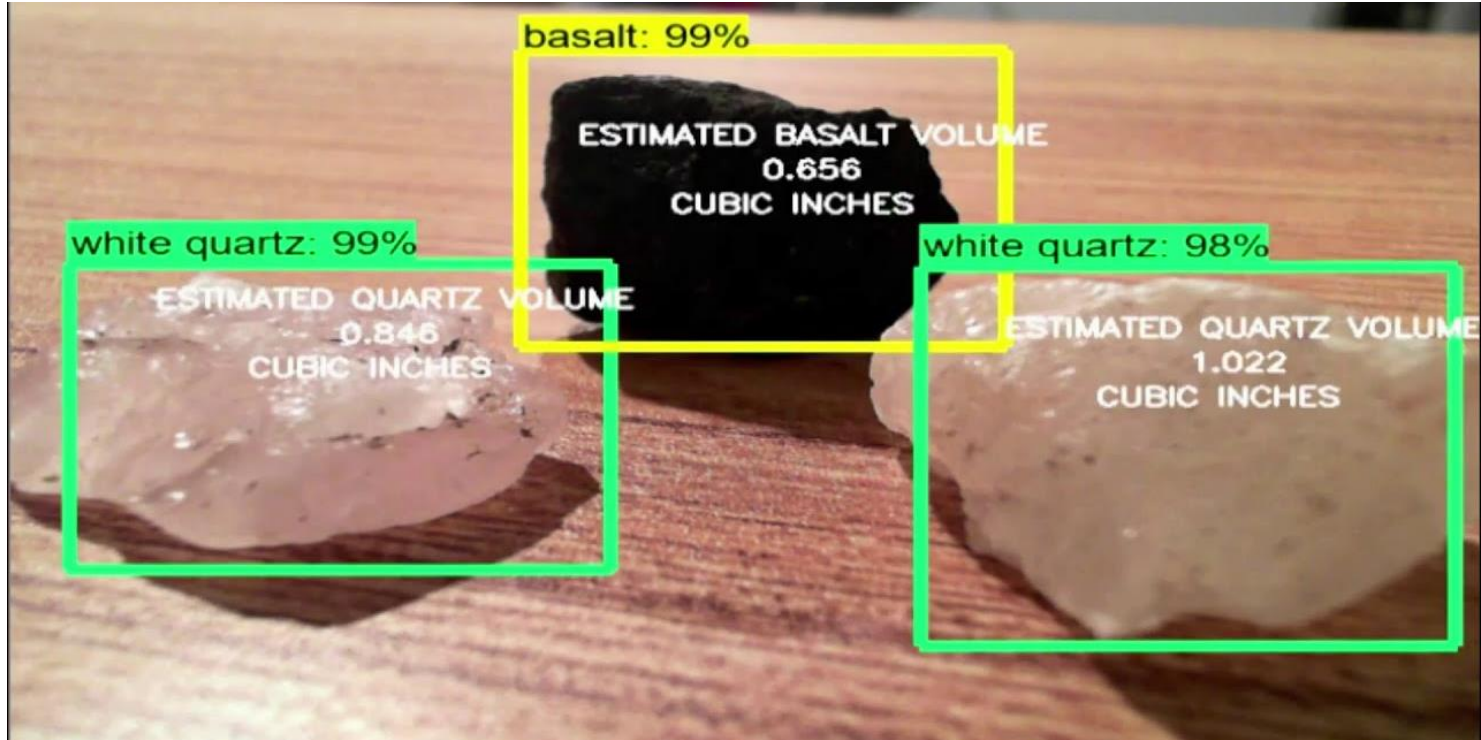


# Vehicle Tracking



...

# Mineral Identification





**Thank you**