VISUAL RECOGNITION

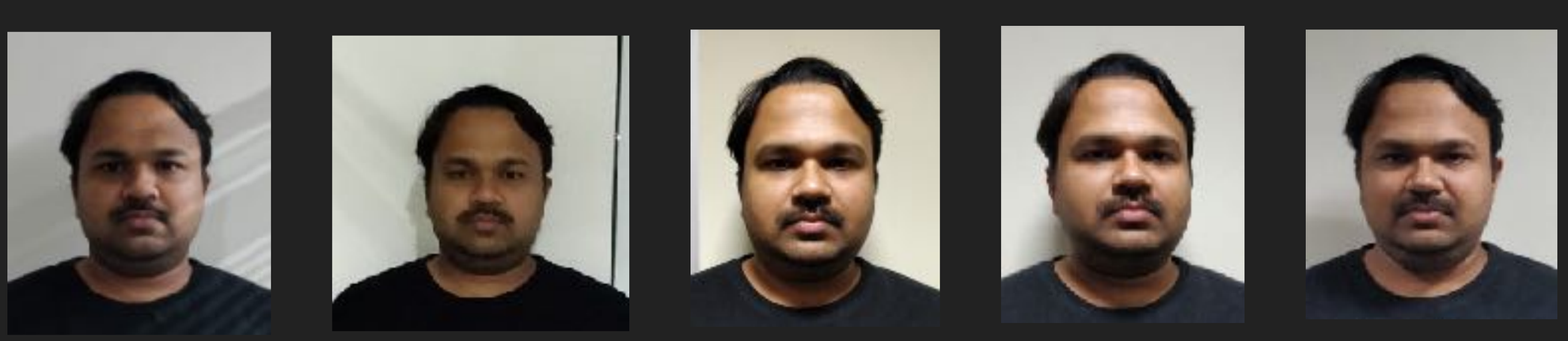
ANUSH SANKARAN
IBM RESEARCH

VISUAL RECOGNITION – WHAT IS THE PROBLEM?

Who is this?

VISUAL RECOGNITION – WHAT IS THE PROBLEM?

Variation in image capture: Illumination, pose, blur, resolution etc.



Who is this?

VISUAL RECOGNITION – WHAT IS THE PROBLEM?

Variation in image object: disguise, age, group etc.

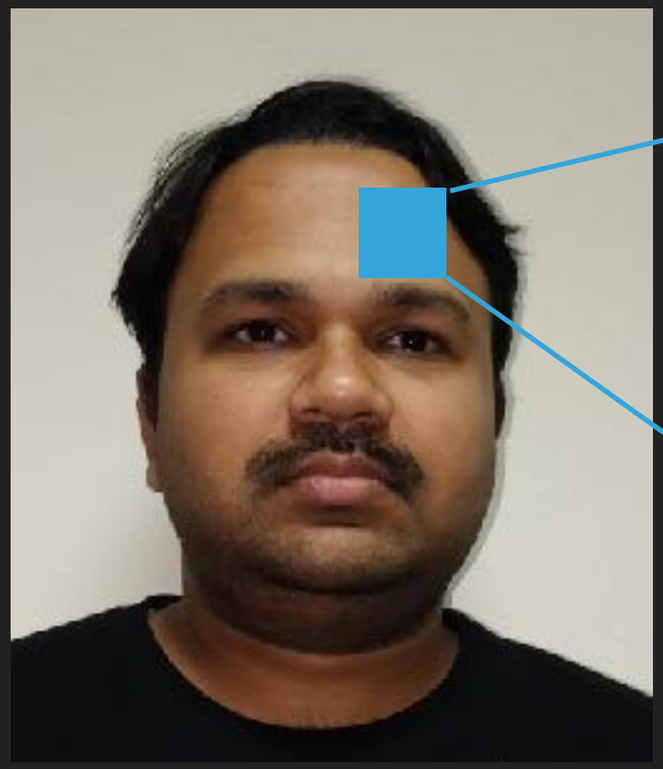
Who is this?

VISUAL RECOGNITION – WHAT IS THE PROBLEM?

- You are a genius …
- Only if a machine or an algorithm could do it …

- Make a computer understand the image …
- Understand the background & environment …
- Classify various objects in the image …

HOW DO WE MAKE A MACHINE CLASSIFY IMAGES?



What we see

195	2	4	213	132	196	81	202	124	186	199
3	164	113	209	36	88	70	221	189	147	247
191	8	88	118	181	182	30	192	50	27	129
114	191	71	249	230	129	248	121	167	61	150
9	133	217	73	251	108	157	232	177	109	102
252	171	125	200	26	38	249	240	93	242	217
101	111	22	118	203	140	27	229	215	106	128
1	192	148	206	199	97	106	64	12	76	170
232	239	254	131	9	180	35	217	22	46	163
111	33	187	135	254	46	57	5	173	48	15
6	59	235	182	153	22	235	106	65	189	15

What computer see

HOW DO WE MAKE A MACHINE CLASSIFY IMAGES?

Image Intensities

195	2	4	213	132	196	81	202	124	186	199
3	164	113	209	36	88	70	221	189	147	247
191	8	88	118	181	182	30	192	50	27	129
114	191	71	249	230	129	248	121	167	61	150
9	133	217	73	251	108	157	232	177	109	102
252	171	125	200	26	38	249	240	93	242	217
101	111	22	118	203	140	27	229	215	106	128
1	192	148	206	199	97	106	64	12	76	170
232	239	254	131	9	180	35	217	22	46	163
111	33	187	135	254	46	57	5	173	48	15
6	59	235	182	153	22	235	106	65	189	15

180	170	100	28	240	148	109	72	13	45	127
219	32	35	42	124	209	133	121	127	181	164
174	101	76	105	58	83	162	7	178	41	104
179	202	250	104	180	8	168	35	68	114	147
215	144	158	145	75	197	83	243	17	111	23
71	254	149	205	233	122	197	191	248	186	85
111	17	185	33	8	92	194	225	198	75	14
210	3	145	207	39	19	91	209	11	254	12
154	72	129	99	103	185	251	82	150	54	216
110	25	136	89	60	49	17	156	116	180	41
237	69	192	92	15	172	196	110	65	34	141

Difference/

Distance/

Similarity

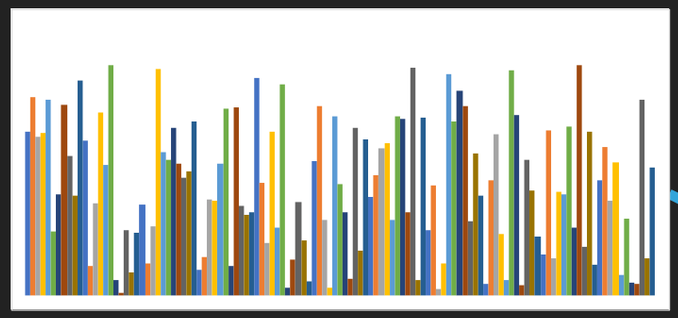
This matrix of numbers is highly sensitive to capture/ object variations

HOW DO WE MAKE A MACHINE CLASSIFY IMAGES?

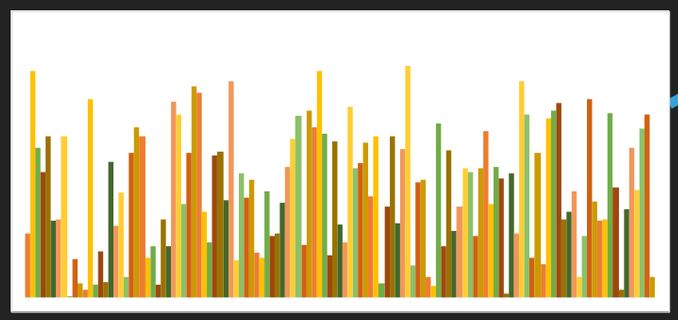
Image Intensities

195	2	4	213	132	196	81	202	124	186	199
3	164	113	209	36	88	70	221	189	147	247
191	8	88	118	181	182	30	192	50	27	129
114	191	71	249	230	129	248	121	167	61	150
9	133	217	73	251	108	157	232	177	109	102
252	171	125	200	26	38	249	240	93	242	217
101	111	22	118	203	140	27	229	215	106	128
1	192	148	206	199	97	106	64	12	76	170
232	239	254	131	9	180	35	217	22	46	163
111	33	187	135	254	46	57	5	173	48	15
6	59	235	182	153	22	235	106	65	189	15

Extract Features



180	170	100	28	240	148	109	72	13	45	127
219	32	35	42	124	209	133	121	127	181	164
174	101	76	105	58	83	162	7	178	41	104
179	202	250	104	180	8	168	35	68	114	147
215	144	158	145	75	197	83	243	17	111	23
71	254	149	205	233	122	197	191	248	186	85
111	17	185	33	8	92	194	225	198	75	14
210	3	145	207	39	19	91	209	11	254	12
154	72	129	99	103	185	251	82	150	54	216
110	25	136	89	60	49	17	156	116	180	41
237	69	192	92	15	172	196	110	65	34	141



Difference/

Distance/

Similarity

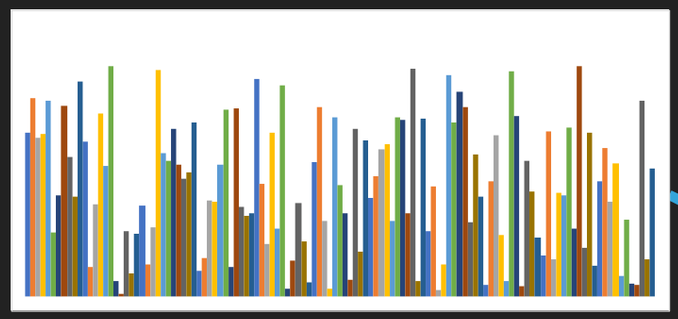
Still I have some problems with a simple Difference function

HOW DO WE MAKE A MACHINE CLASSIFY IMAGES?

Image Intensities

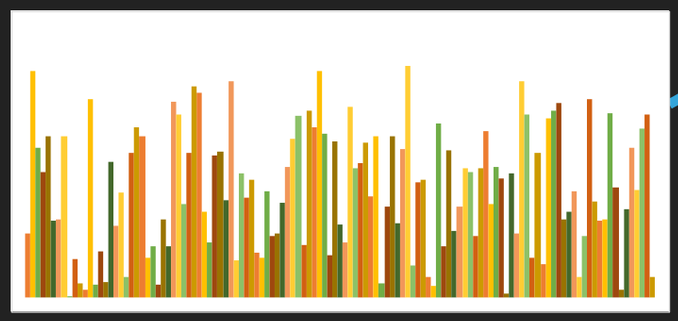
195	2	4	213	132	196	81	202	124	186	199
3	164	113	209	36	88	70	221	189	147	247
191	8	88	118	181	182	30	192	50	27	129
114	191	71	249	230	129	248	121	167	61	150
9	133	217	73	251	108	157	232	177	109	102
252	171	125	200	26	38	249	240	93	242	217
101	111	22	118	203	140	27	229	215	106	128
1	192	148	206	199	97	106	64	12	76	170
232	239	254	131	9	180	35	217	22	46	163
111	33	187	135	254	46	57	5	173	48	15
6	59	235	182	153	22	235	106	65	189	15

Extract Features



Learn how to match similarity

180	170	100	28	240	148	109	72	13	45	127
219	32	35	42	124	209	133	121	127	181	164
174	101	76	105	58	83	162	7	178	41	104
179	202	250	104	180	8	168	35	68	114	147
215	144	158	145	75	197	83	243	17	111	23
71	254	149	205	233	122	197	191	248	186	85
111	17	185	33	8	92	194	225	198	75	14
210	3	145	207	39	19	91	209	11	254	12
154	72	129	99	103	185	251	82	150	54	216
110	25	136	89	60	49	17	156	116	180	41
237	69	192	92	15	172	196	110	65	34	141



This is called the data-driven approach for learning

HOW DO WE MAKE A MACHINE CLASSIFY IMAGES?

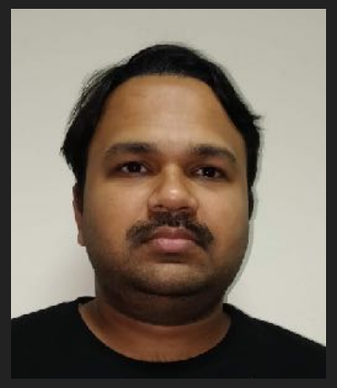
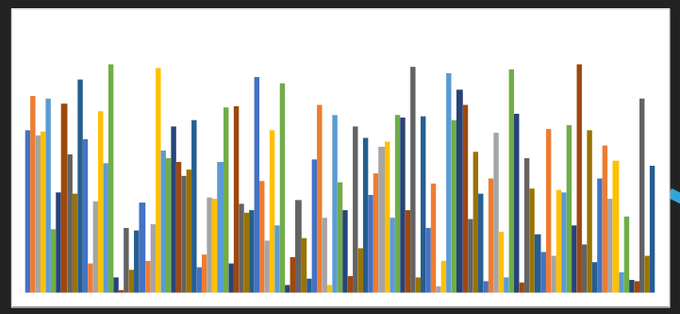


Image Intensities

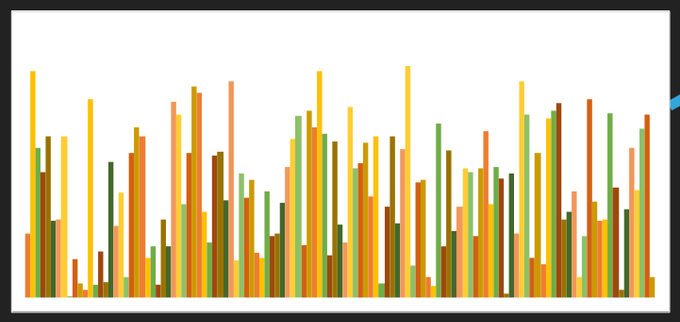
195	2	4	213	132	196	81	202	124	186	199
3	164	113	209	36	88	70	221	189	147	247
191	8	88	118	181	182	30	192	50	27	129
114	191	71	249	230	129	248	121	167	61	150
9	133	217	73	251	108	157	232	177	109	102
252	171	125	200	26	38	249	240	93	242	217
101	111	22	118	203	140	27	229	215	106	128
1	192	148	206	199	97	106	64	12	76	170
232	239	254	131	9	180	35	217	22	46	163
111	33	187	135	254	46	57	5	173	48	15
6	59	235	182	153	22	235	106	65	189	15

Extract Features

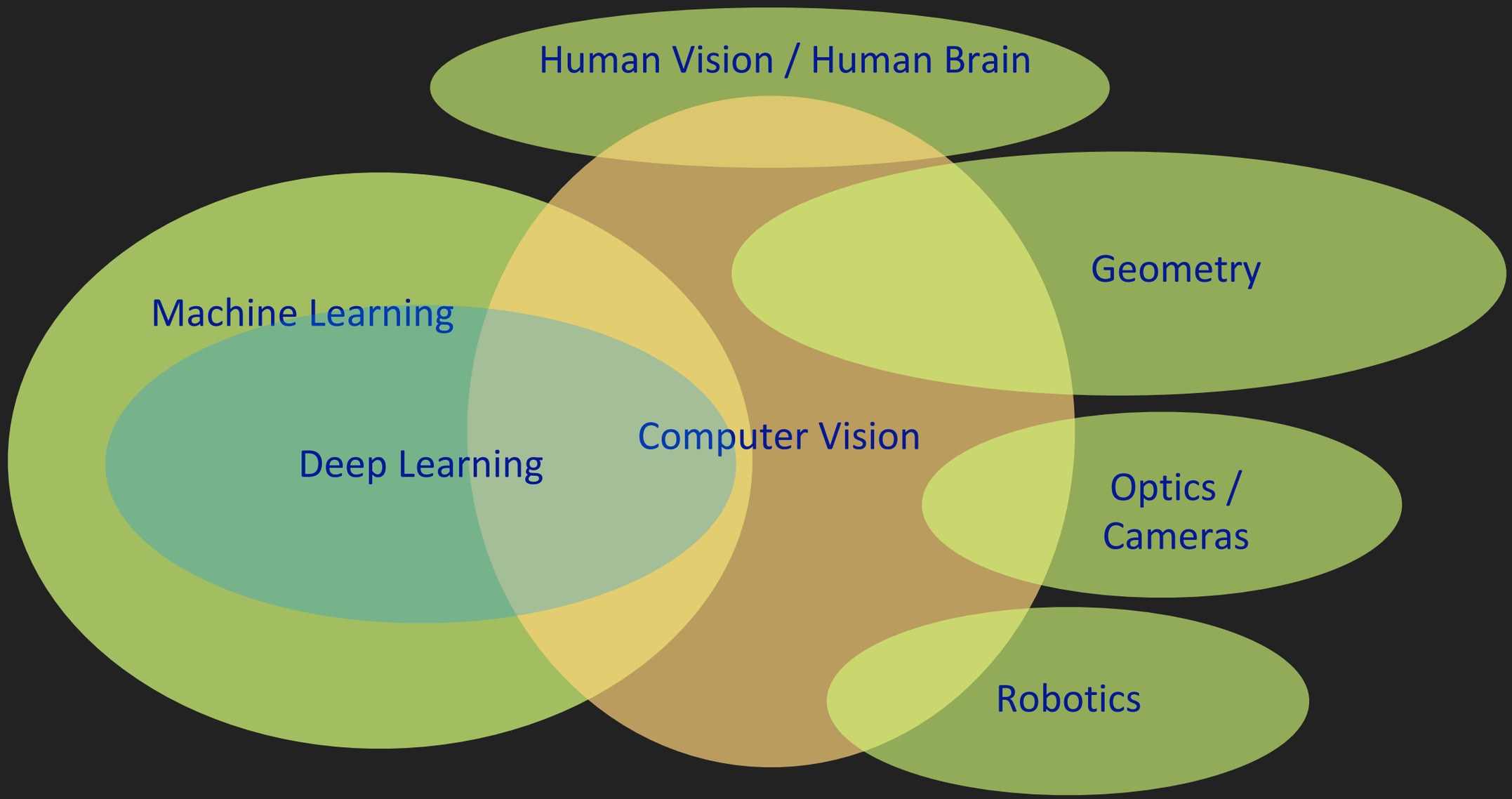


Learn how to match similarity

180	170	100	28	240	148	109	72	13	45	127
219	32	35	42	124	209	133	121	127	181	164
174	101	76	105	58	83	162	7	178	41	104
179	202	250	104	180	8	168	35	68	114	147
215	144	158	145	75	197	83	243	17	111	23
71	254	149	205	233	122	197	191	248	186	85
111	17	185	33	8	92	194	225	198	75	14
210	3	145	207	39	19	91	209	11	254	12
154	72	129	99	103	185	251	82	150	54	216
110	25	136	89	60	49	17	156	116	180	41
237	69	192	92	15	172	196	110	65	34	141

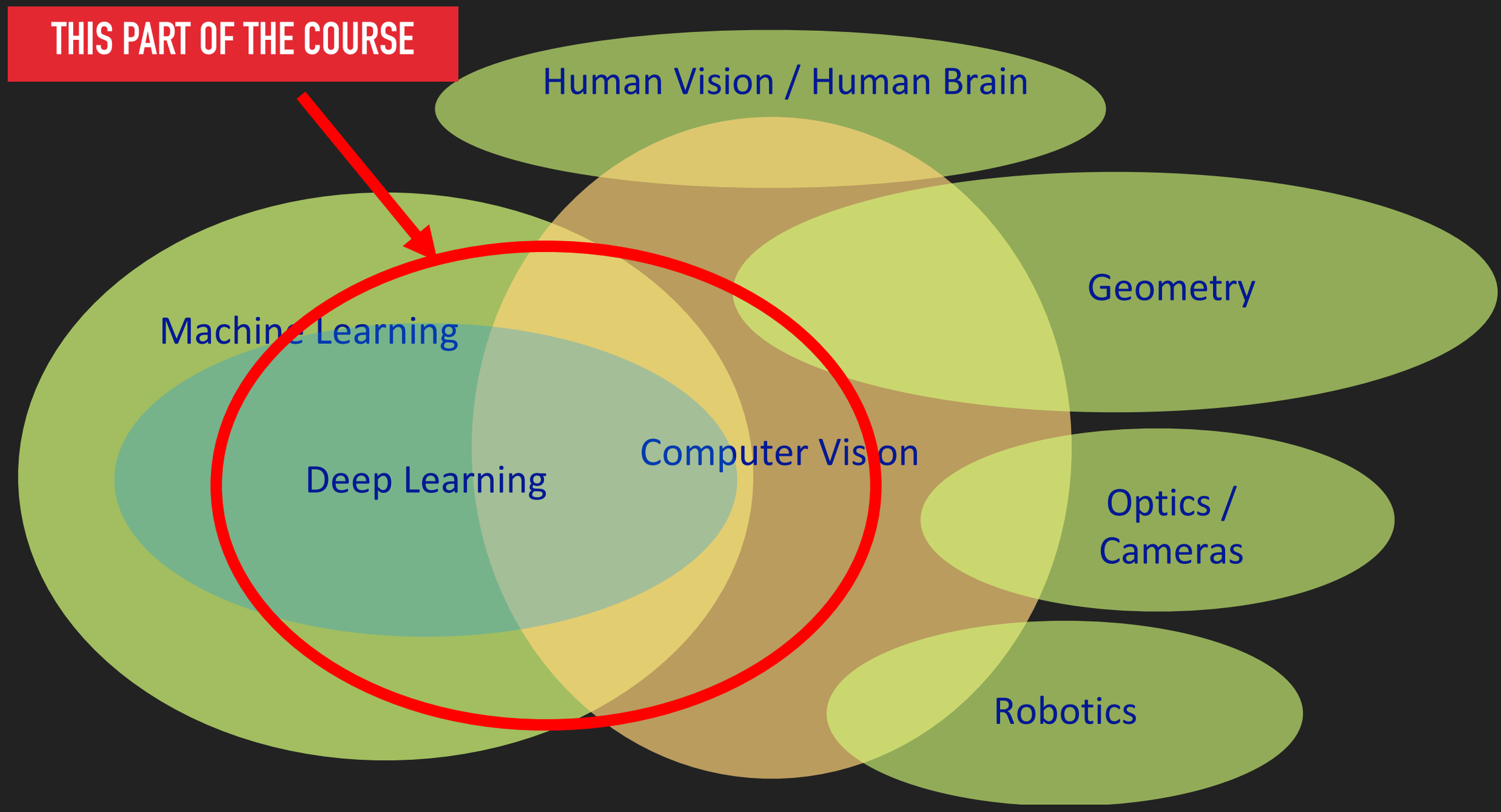


BIG PICTURE OF COMPUTER VISION



http://www.cs.virginia.edu/~vicente/deeplearning/slides/lecture01.pdf

BIG PICTURE OF COMPUTER VISION



http://www.cs.virginia.edu/~vicente/deeplearning/slides/lecture01.pdf

REMAINING STRUCTURE OF THIS COURSE

- Week 1: Introduction to Neural Networks, Loss Functions and Optimization , Intro to Image Classification
- Week 2: Convolutional Neural Networks, Training Neural Networks
- Week 3: Different CNN Architectures
- Week 4: Advanced training strategies and interesting applications
- Week 5: Generative Models , Visualizing, and Understanding
- Week 6: Face recognition, Face Detection, and generating Face images

And we will have lots of fun…

Don’t forget that I am a wannabe StandUp Comedian

REFERENCES (THEY HAVE BETTER SLIDES . . .)

- Book on “Deep Learning” (https://www.deeplearningbook.org/)
- CS231n: Convolutional Neural Networks for Visual Recognition (http://vision.stanford.edu/teaching/cs231n/index.html)
- CS 6501-004: Deep Learning for Visual Recognition (http://vicenteordonez.com/deeplearning/)
- ECE 6504 Deep Learning for Perception (https://computing.ece.vt.edu/~f15ece6504/)

A SIMPLE EXAMPLE

Classifying a person as MALE or FEMALE

Data points

1. Height
2. Weight

Feature extraction

1. Male:
   1. Weight > 75kg
   2. Height > 5’9”
2. Female:
   1. Weight < 70kg
   2. Height < 5’7”

Predict

Loss of generalisation !!!

A SIMPLE EXAMPLE

Learn a classifier: a mapping function

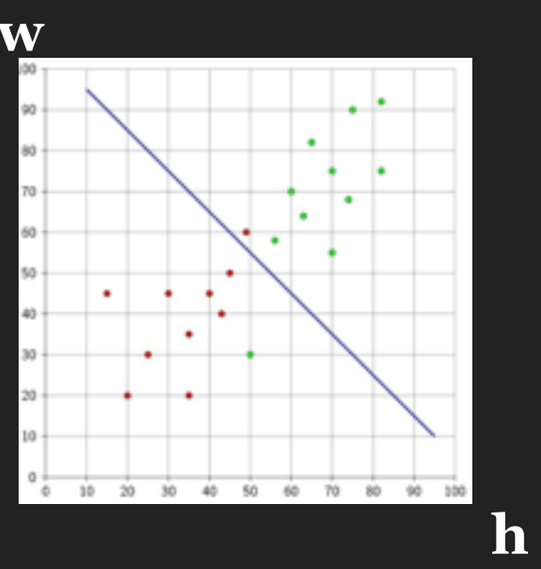
Data points

1. M: <h1, w1>
2. F: <h2, w2>
3. …

…

N. M: <hn, wn>

Feature extraction

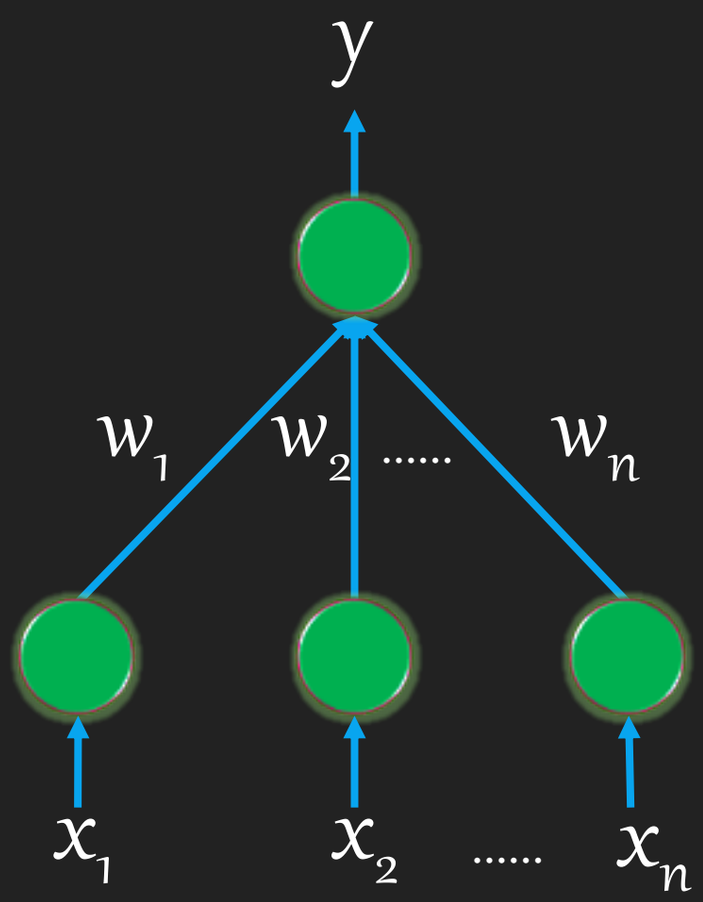


Learn & Predict

A classifier is only as good as the features are … !

HOW TO LEARN TO CLASSIFY?

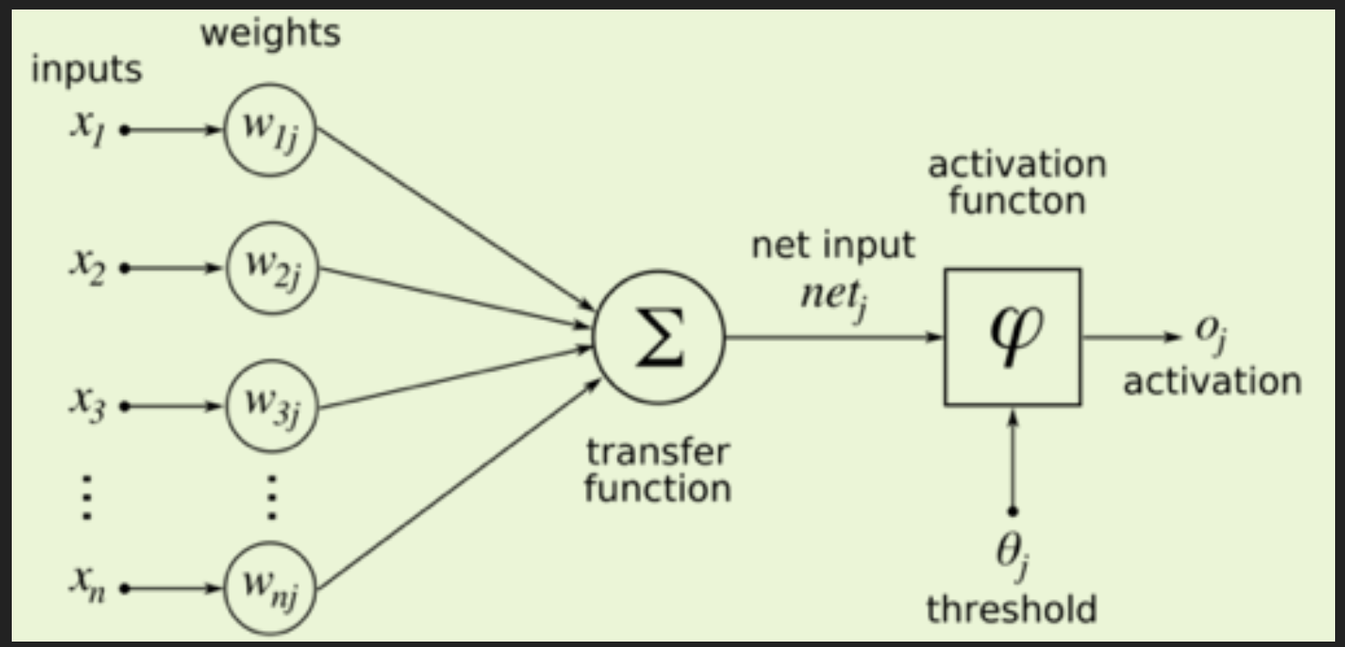
$$y = f\left(\Sigma_i w_i x_i\right)$$

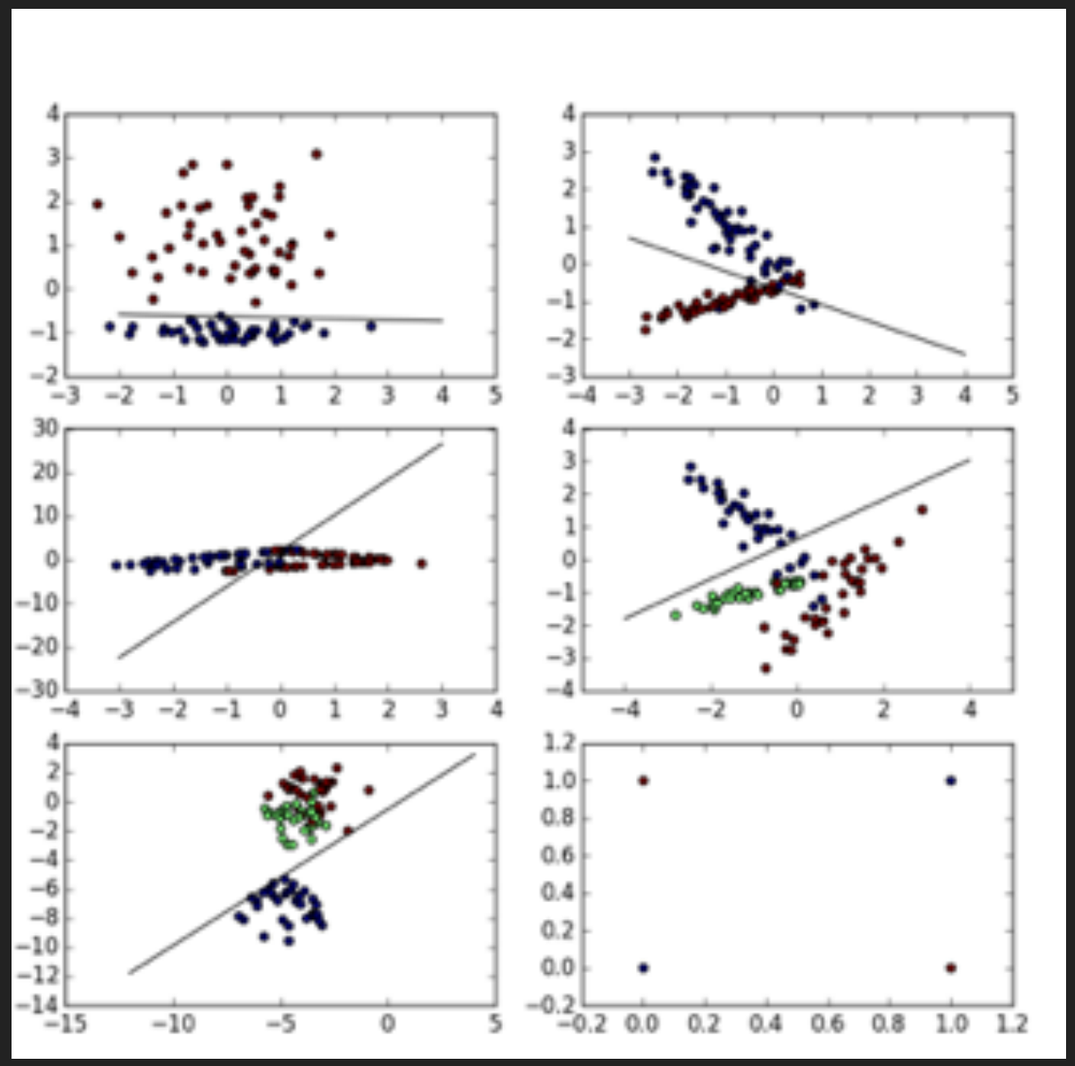


Error = $(y - y')^2$

Update **w** in such a way that the error is minimized !

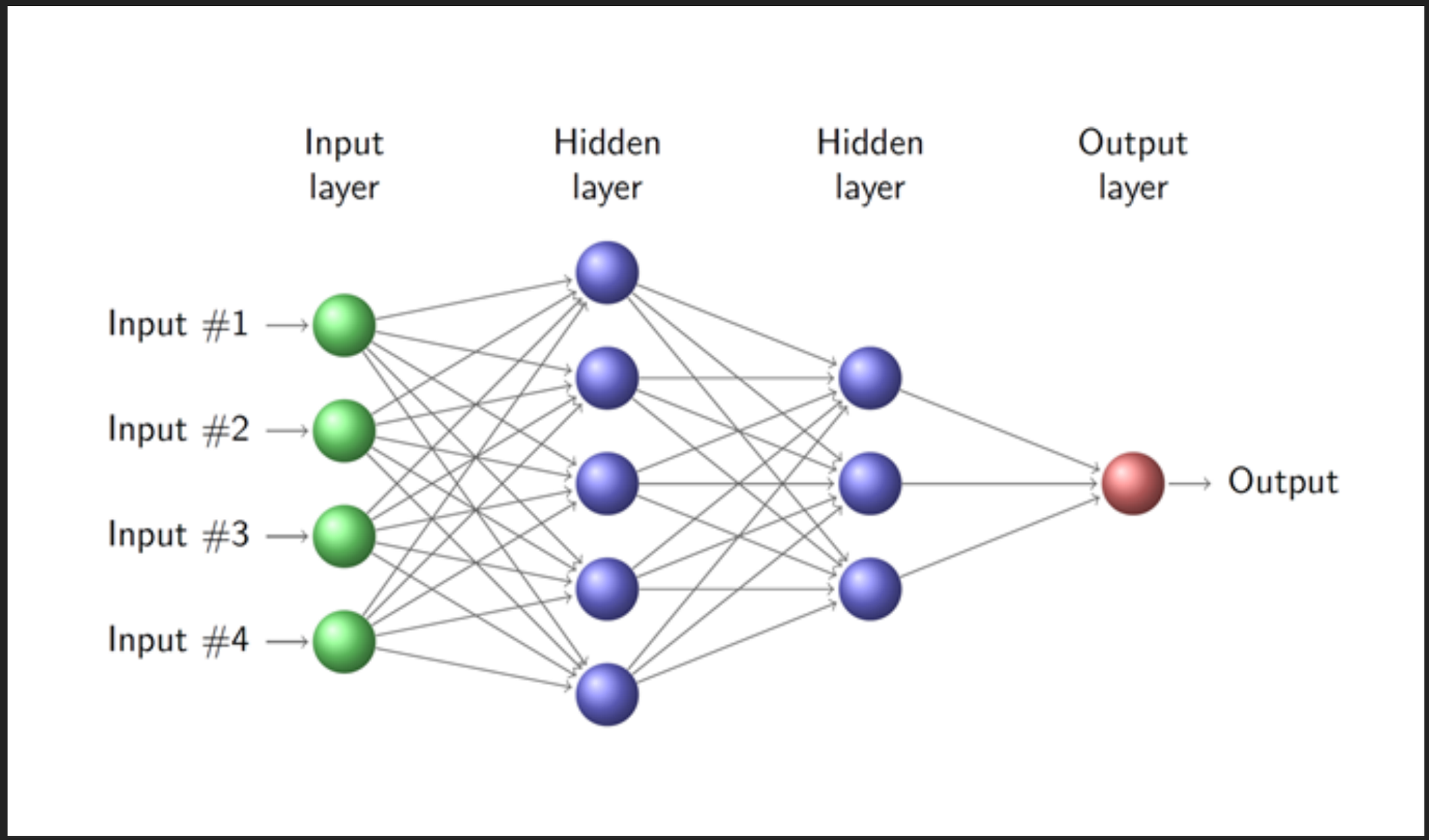
HOW TO LEARN TO CLASSIFY? PERCEPTRON





It’s linear !!!

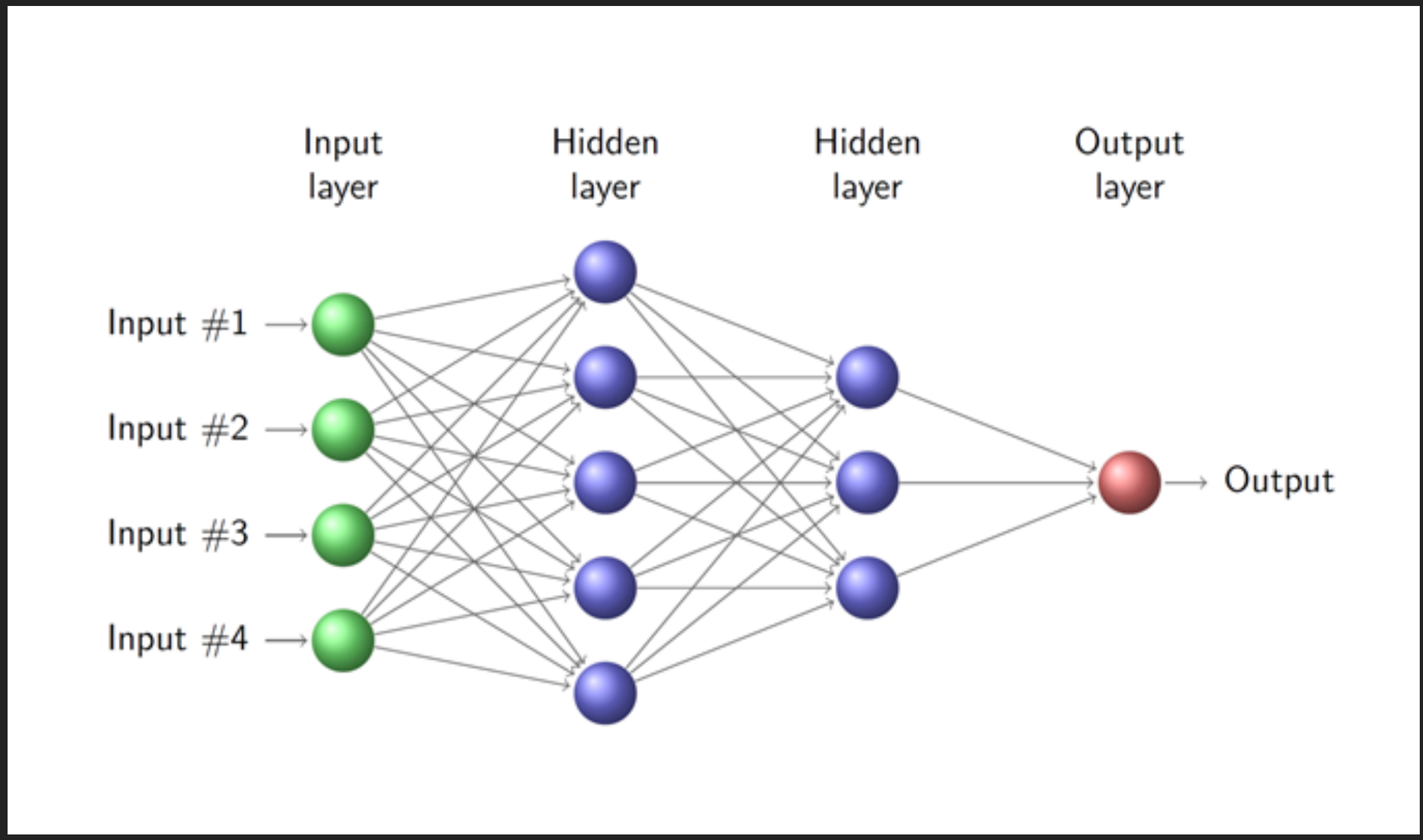
MULTI LAYER PERCEPTRON: NEURAL NETWORK



HOW TO LEARN?

- Stochastic Gradient Descent (SGD)
- Back propagation (BackProp)
- Vanishing gradient problem
- Single hidden layer NN is a universal approximate

WHAT IS THE INPUT HERE?



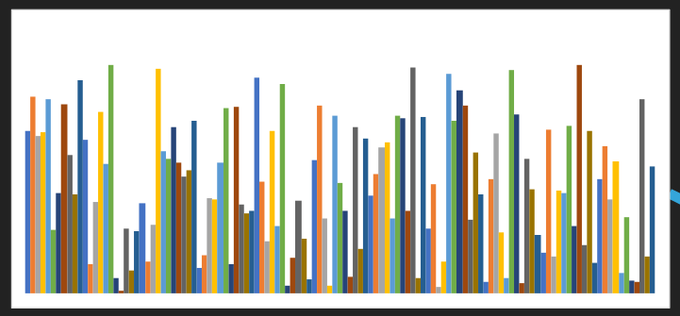
Assumption is that the input here is sufficient.. What if it is not?

WHAT ARE THE FEATURES?

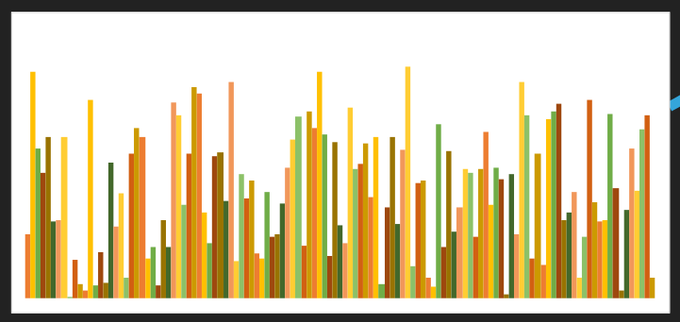
Image Intensities

195	2	4	213	132	196	81	202	124	186	199
3	164	113	209	36	88	70	221	189	147	247
191	8	88	118	181	182	30	192	50	27	129
114	191	71	249	230	129	248	121	167	61	150
9	133	217	73	251	108	157	232	177	109	102
252	171	125	200	26	38	249	240	93	242	217
101	111	22	118	203	140	27	229	215	106	128
1	192	148	206	199	97	106	64	12	76	170
232	239	254	131	9	180	35	217	22	46	163
111	33	187	135	254	46	57	5	173	48	15
6	59	235	182	153	22	235	106	65	189	15

Extract Features



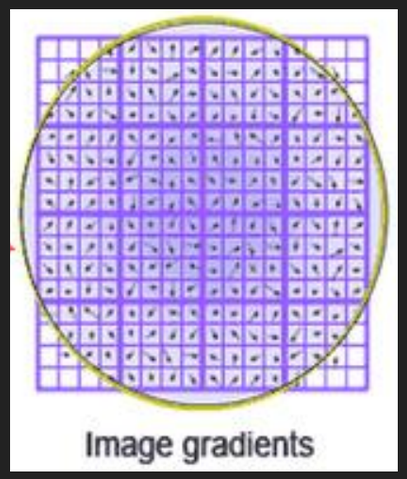
180	170	100	28	240	148	109	72	13	45	127
219	32	35	42	124	209	133	121	127	181	164
174	101	76	105	58	83	162	7	178	41	104
179	202	250	104	180	8	168	35	68	114	147
215	144	158	145	75	197	83	243	17	111	23
71	254	149	205	233	122	197	191	248	186	85
111	17	185	33	8	92	194	225	198	75	14
210	3	145	207	39	19	91	209	11	254	12
154	72	129	99	103	185	251	82	150	54	216
110	25	136	89	60	49	17	156	116	180	41
237	69	192	92	15	172	196	110	65	34	141



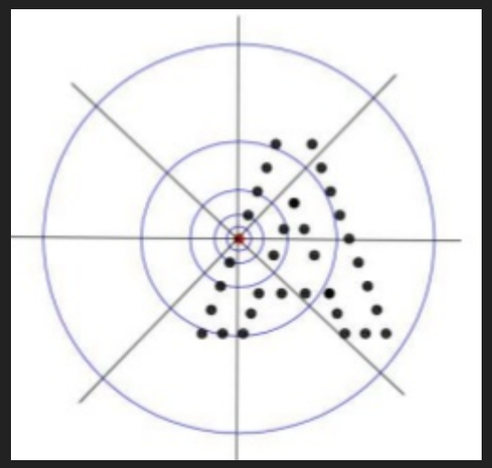
Learn how to match similarity

CLASSICAL COMPUTER VISION FEATURES

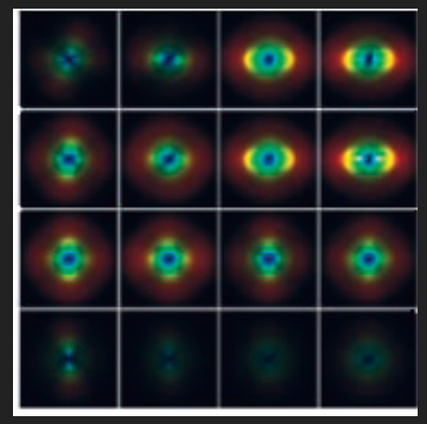
SIFT



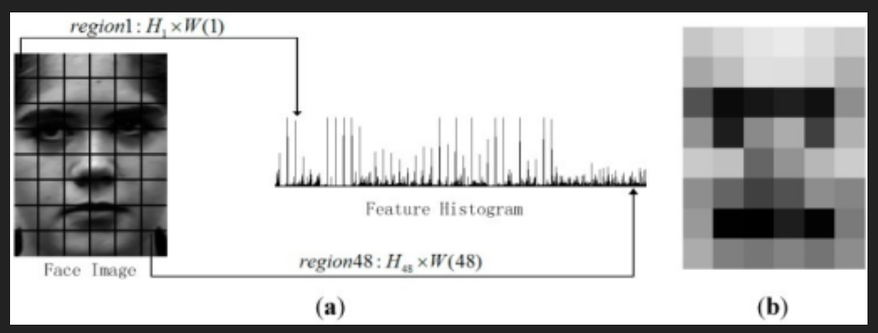
Shape Vector



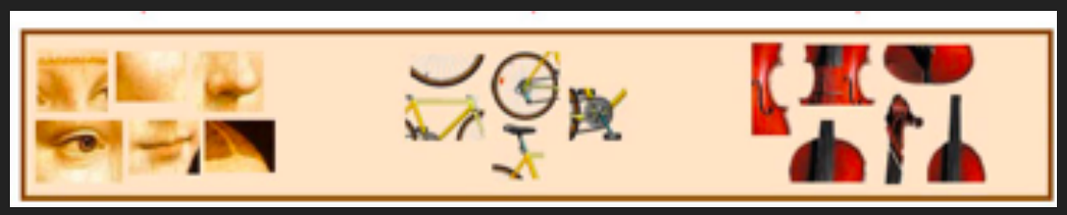
GIST



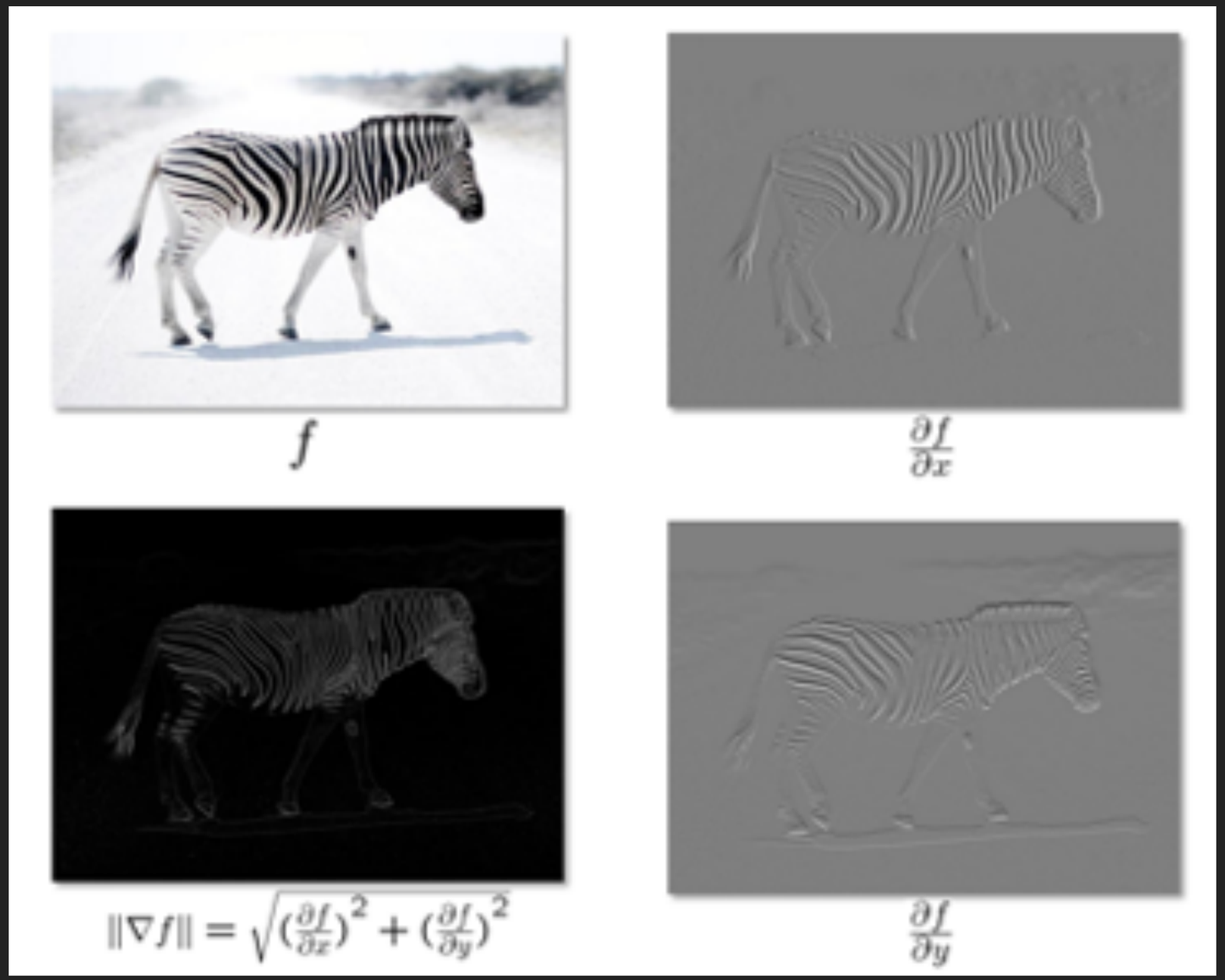
LBP



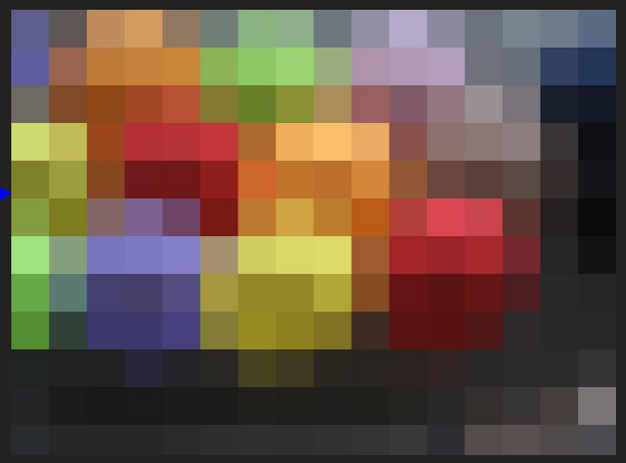
Bag of Words

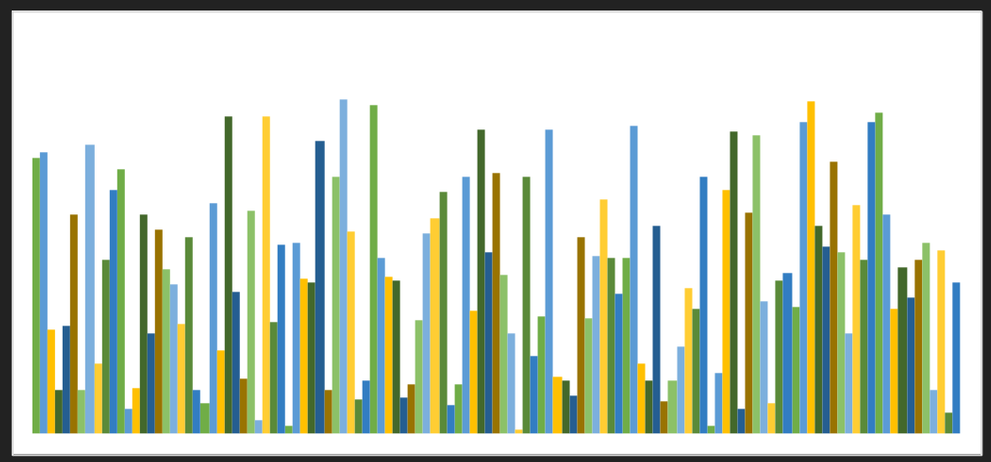


FEATURES: IMAGE GRADIENT

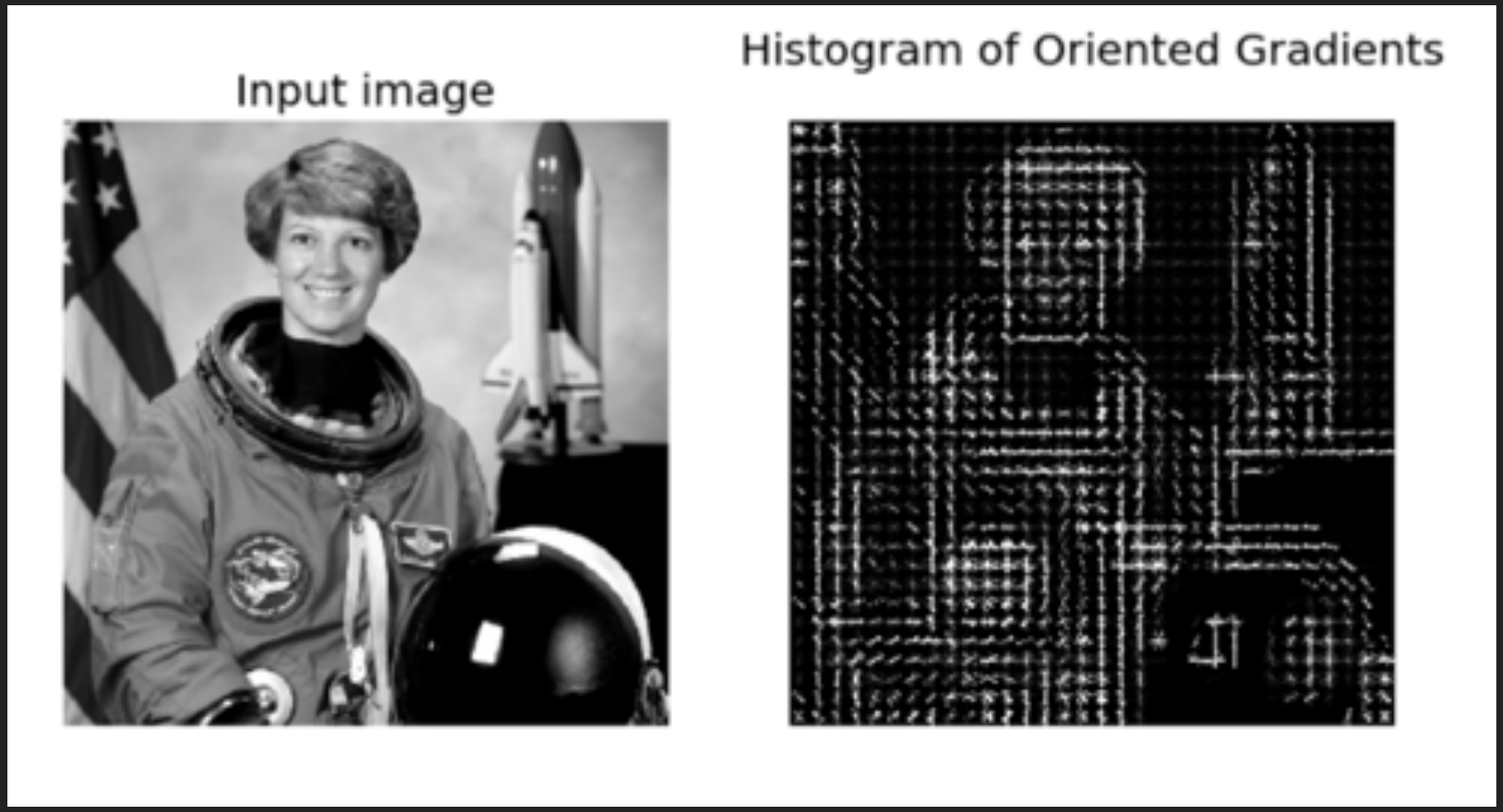


FEATURES: COLOR HISTOGRAMS

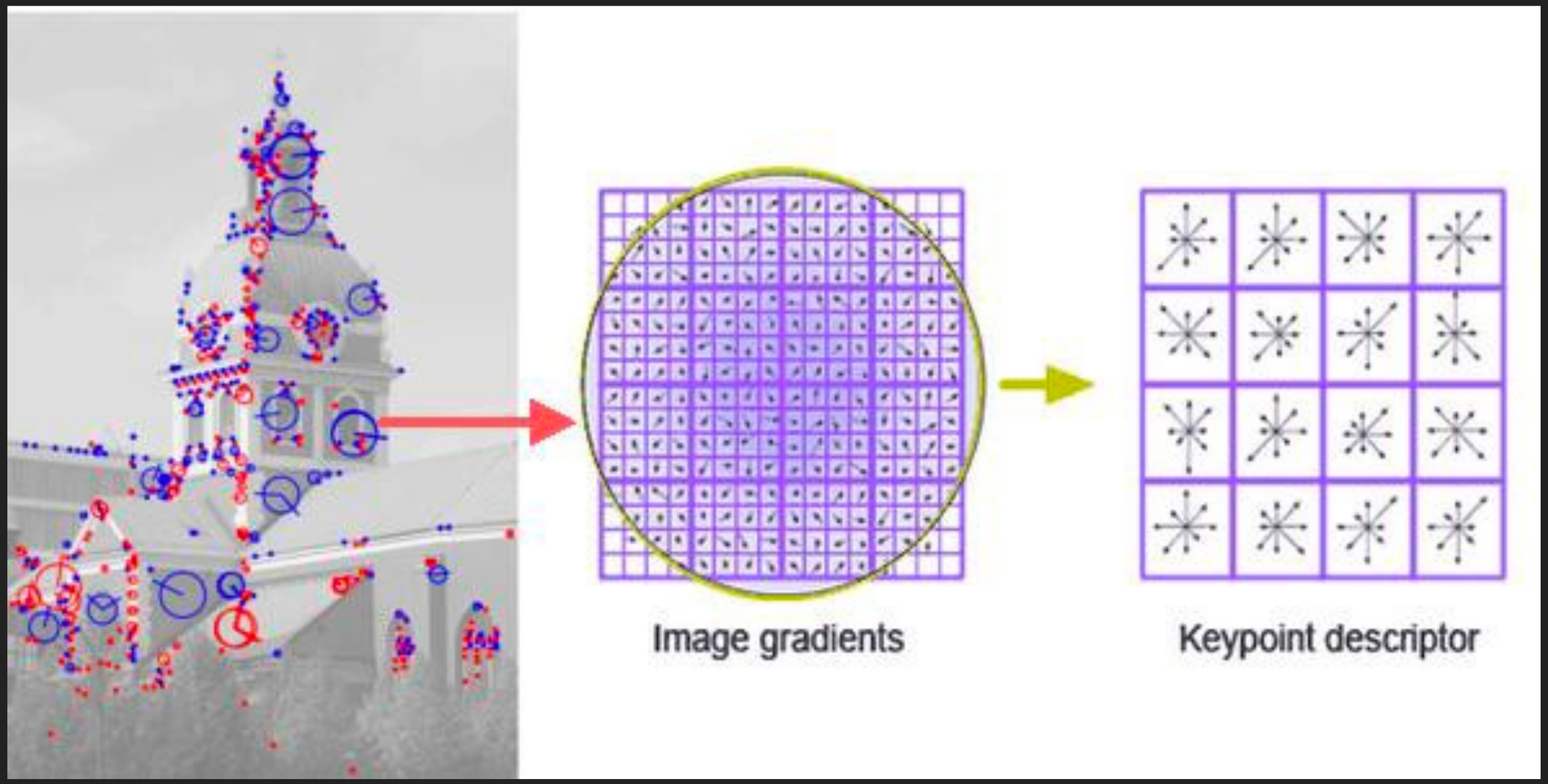




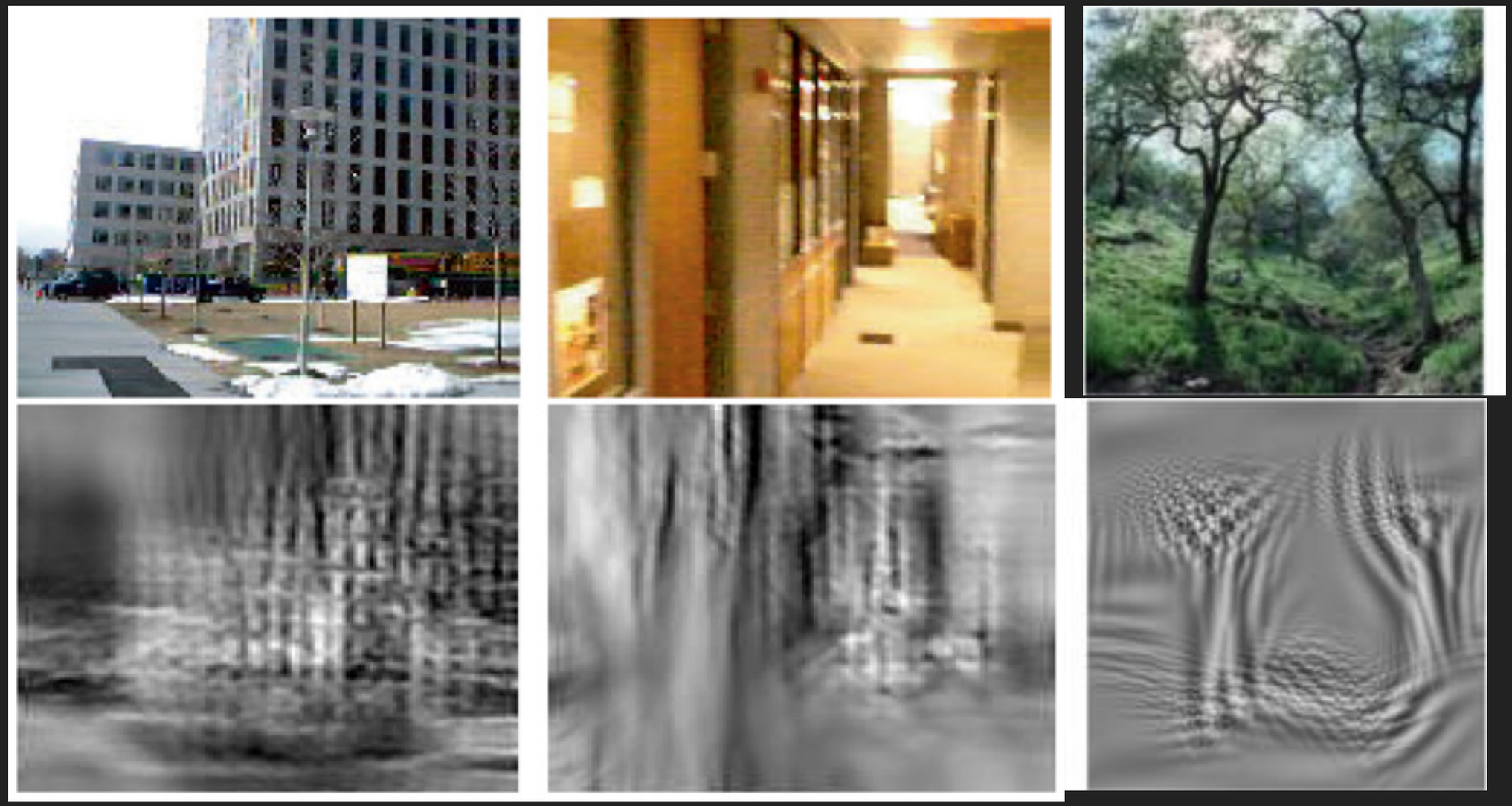
FEATURES: HISTOGRAM OF ORIENTED GRADIENTS (HOG)



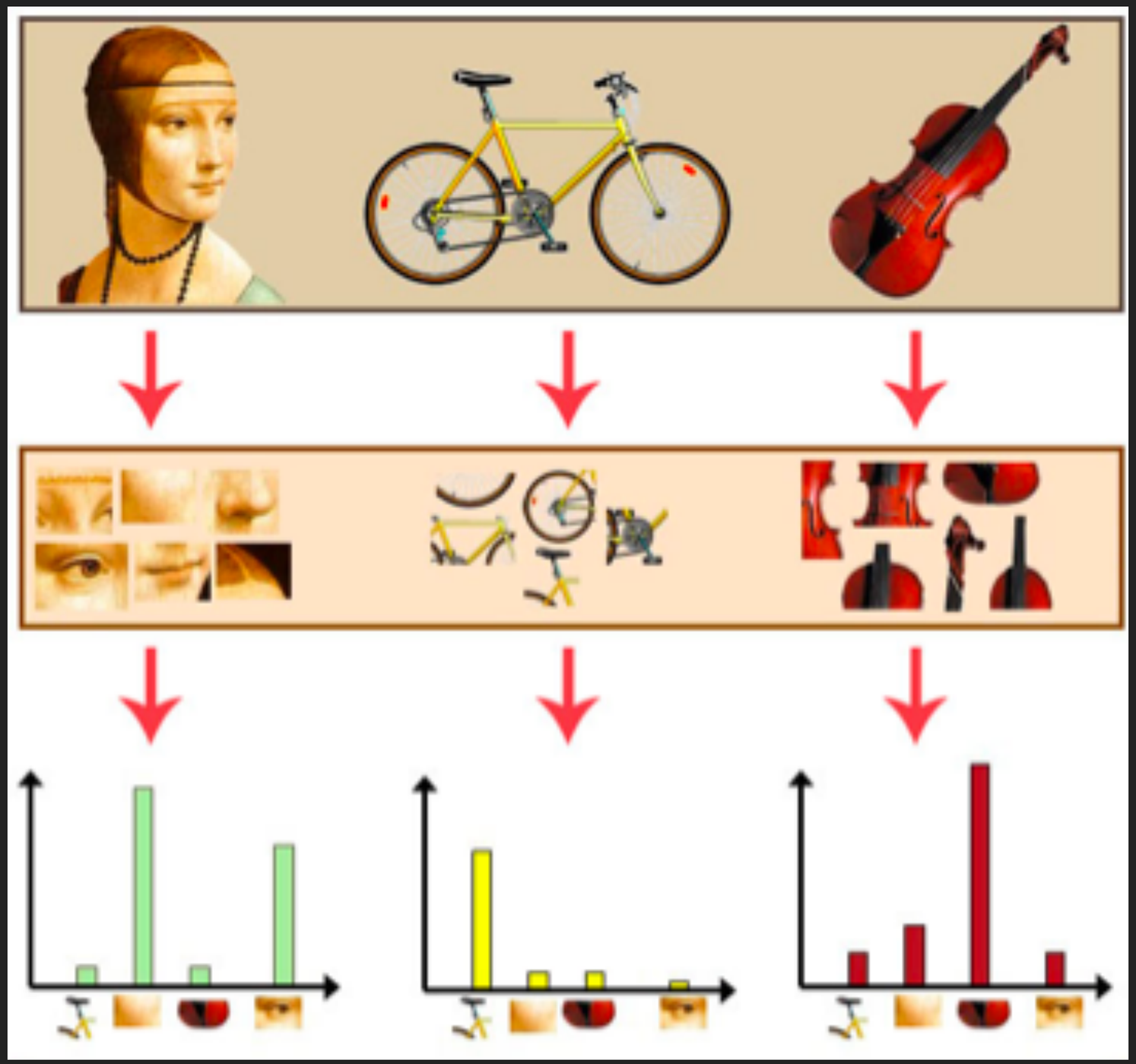
FEATURES: SIFT KEYPOINT DETECTOR



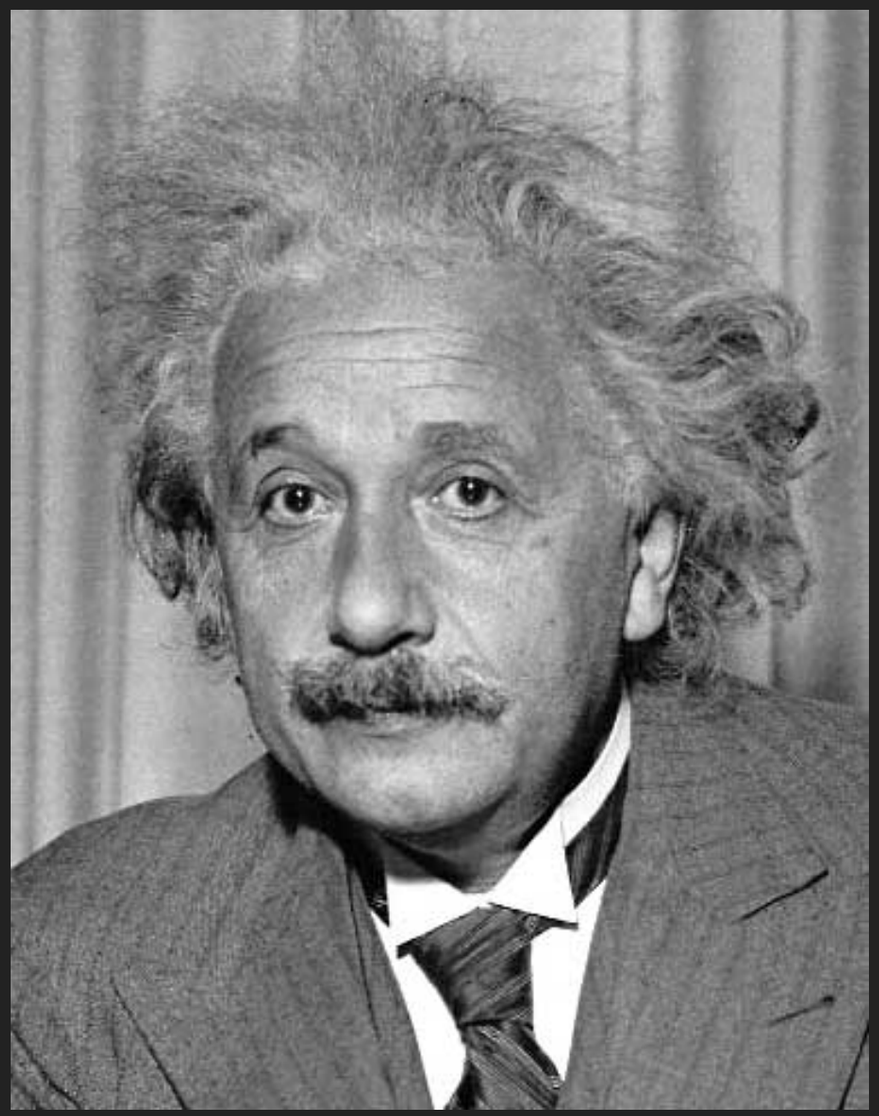
FEATURES: GIST



FEATURES: BAG-OF-WORDS

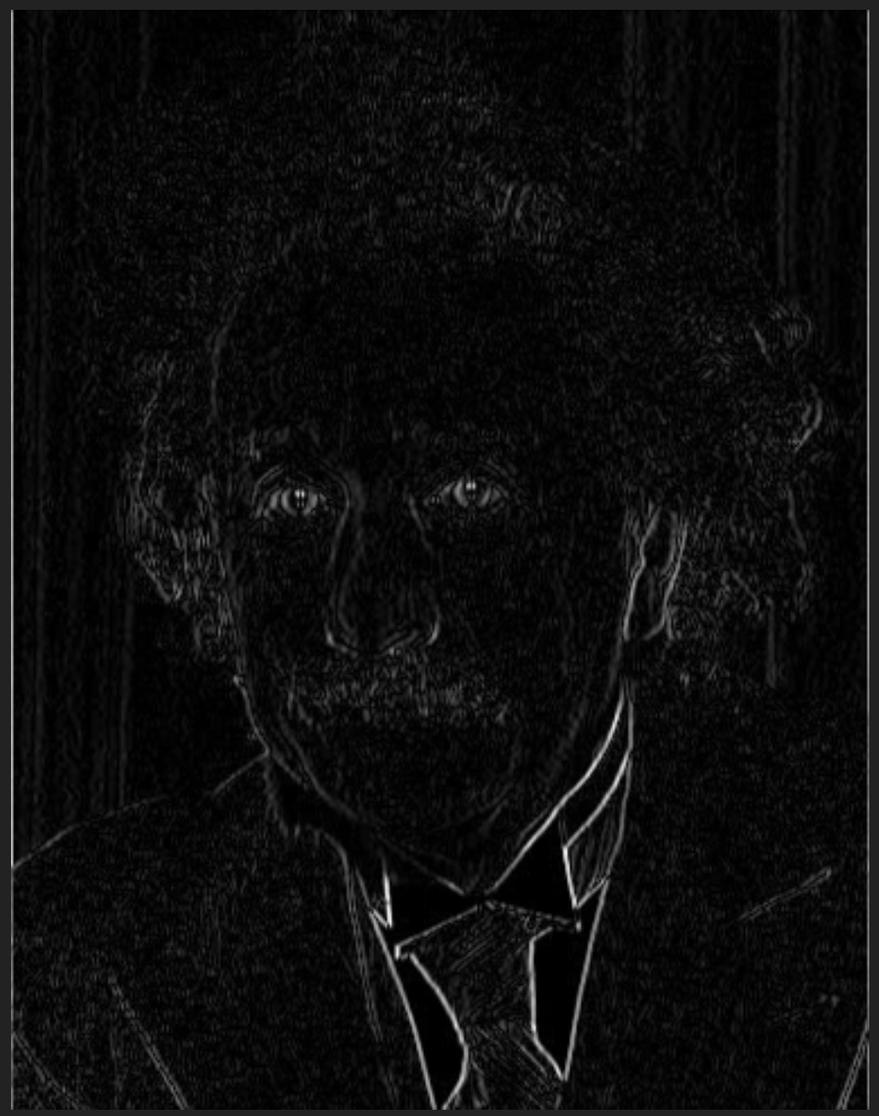


FEATURES: VERTICAL SOBEL OPERATOR

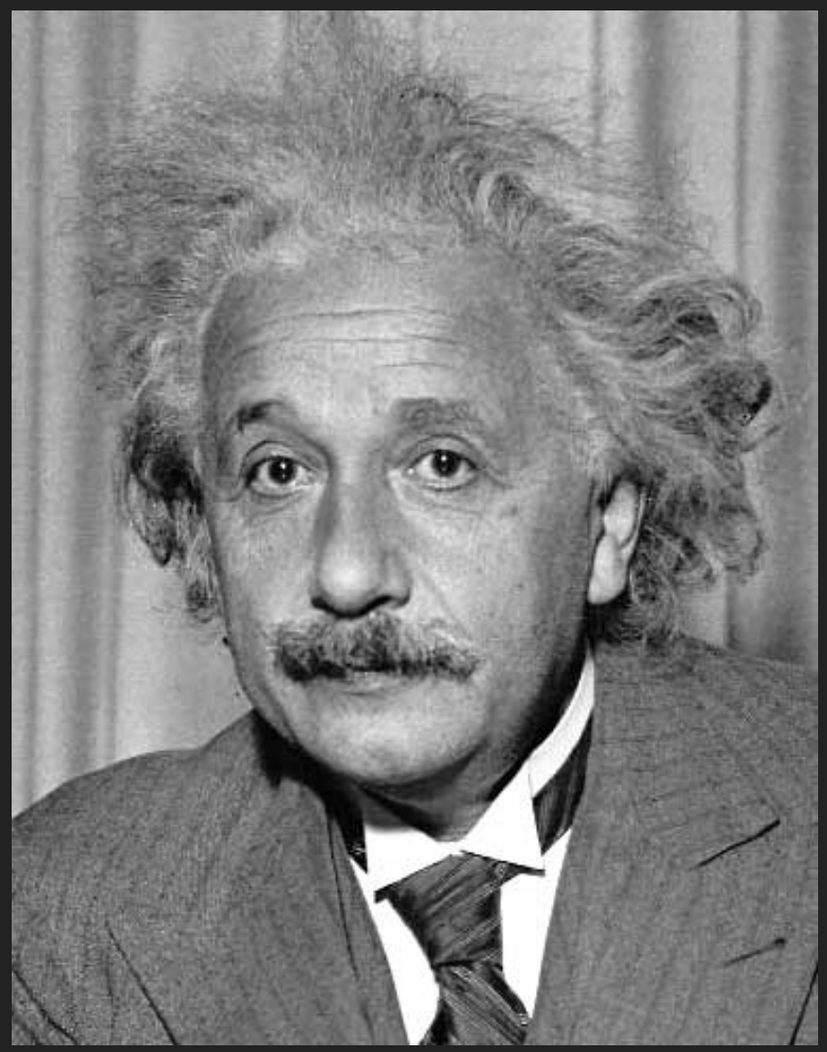


1	0	-1
2	0	-2
1	0	-1

Sobel



FEATURES: HORIZONTAL SOBEL OPERATOR



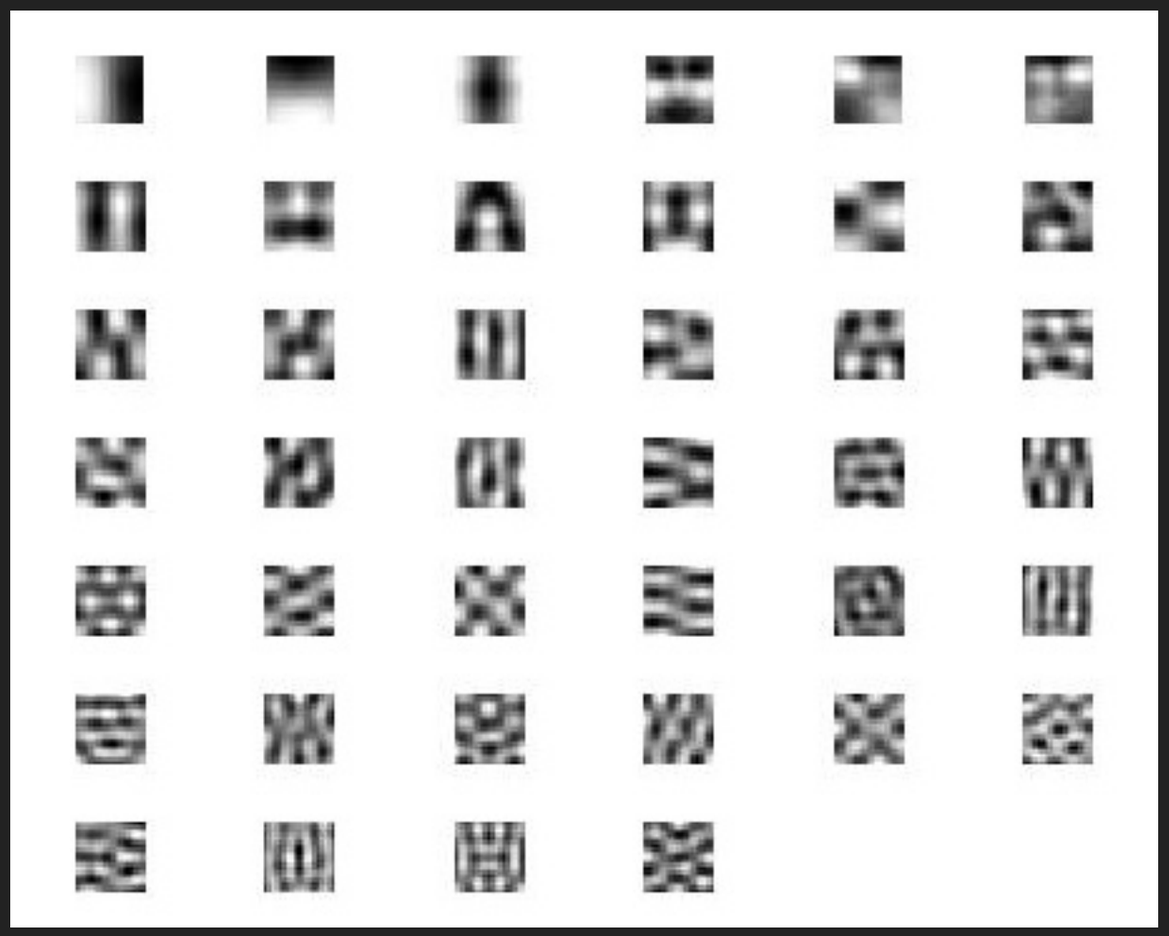
1	2	1
0	0	0
-1	-2	-1

Sobel

FEATURES: LAPLACIAN FILTER

-10	-5	-2	-1	-2	-5	-10
-5	0	3	4	3	0	-5
-2	3	6	7	6	3	-2
-1	4	7	8	7	4	-1
-2	3	6	7	6	3	-2
-5	0	3	4	3	0	-5
-10	-5	-2	-1	-2	-5	-10

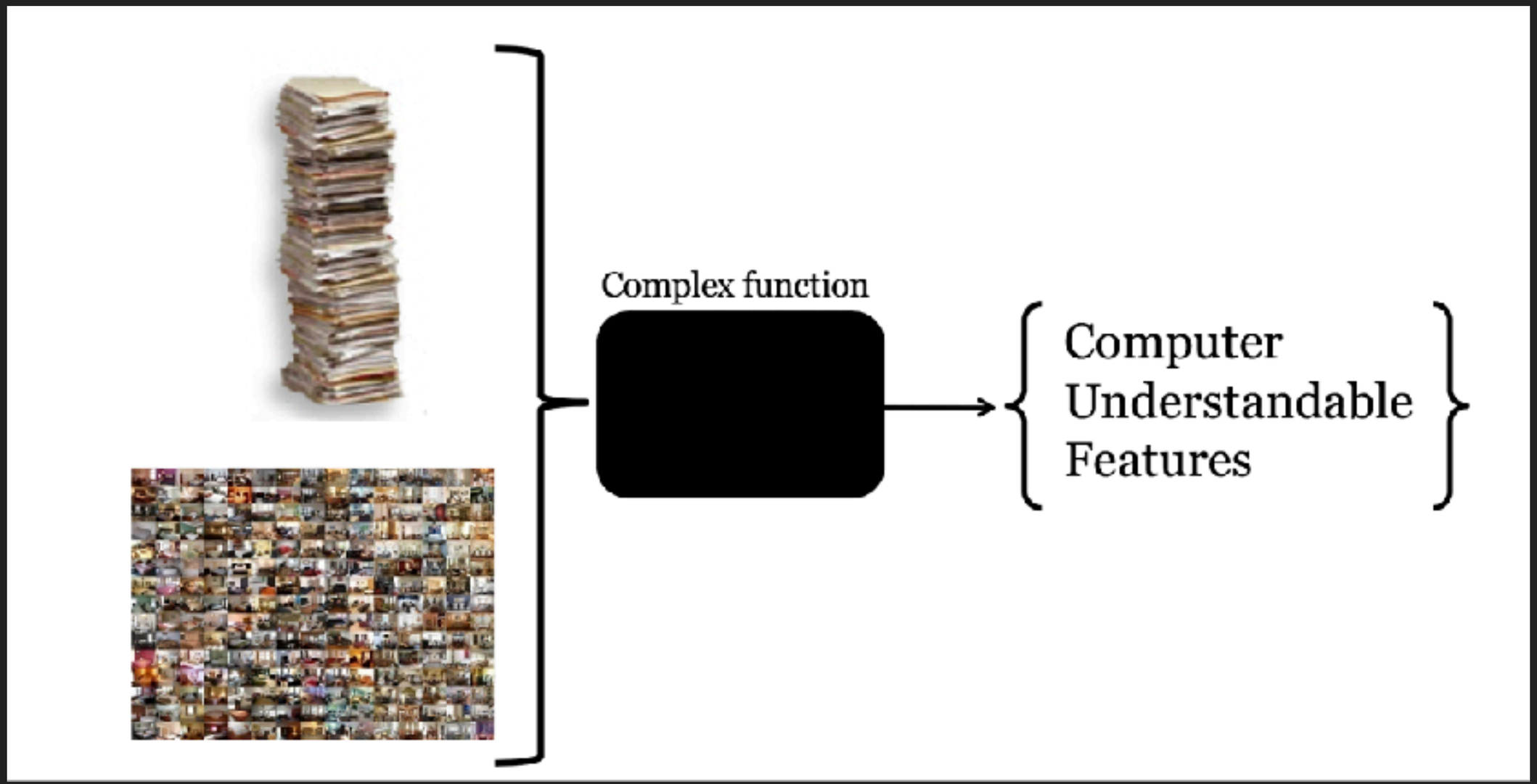
FEATURES: AND A FAMILY OF FILTERS . . .



- Each of these filters is a mathematical function that has to be manually defined
- Which filter works is a hit and trial method

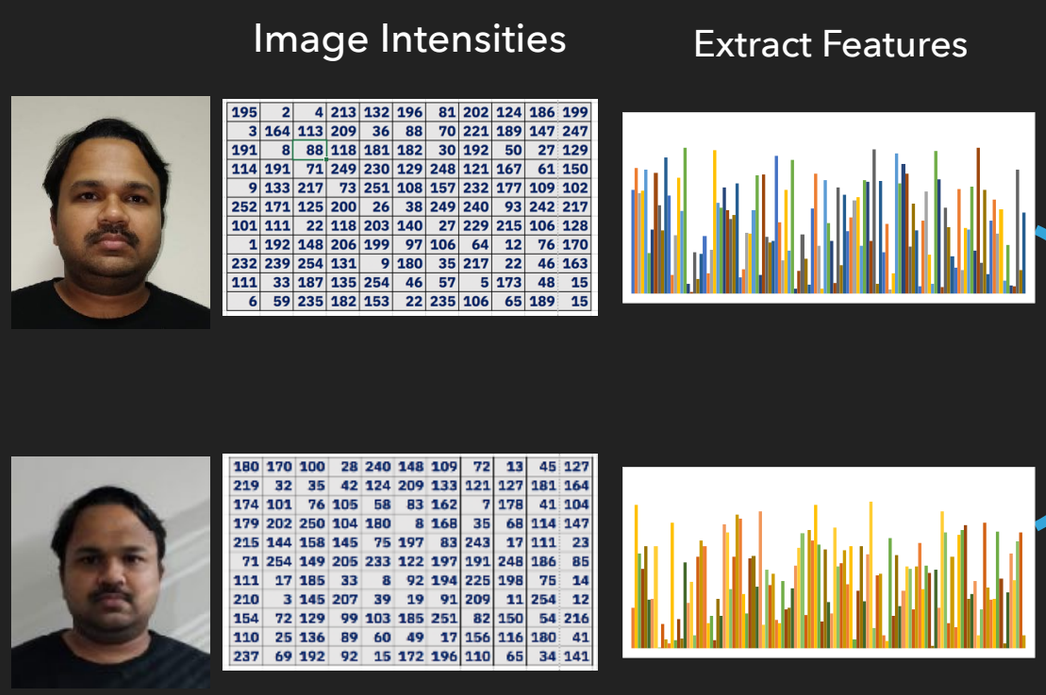
FOCUS IS ON THE FEATURES …

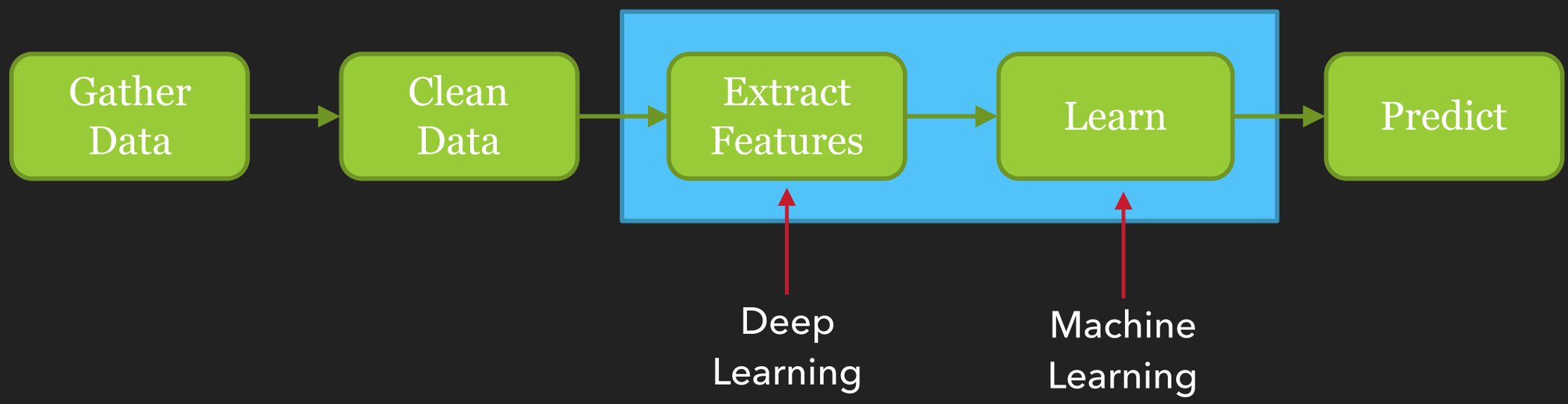
UNSUPERVISED FEATURE LEARNING



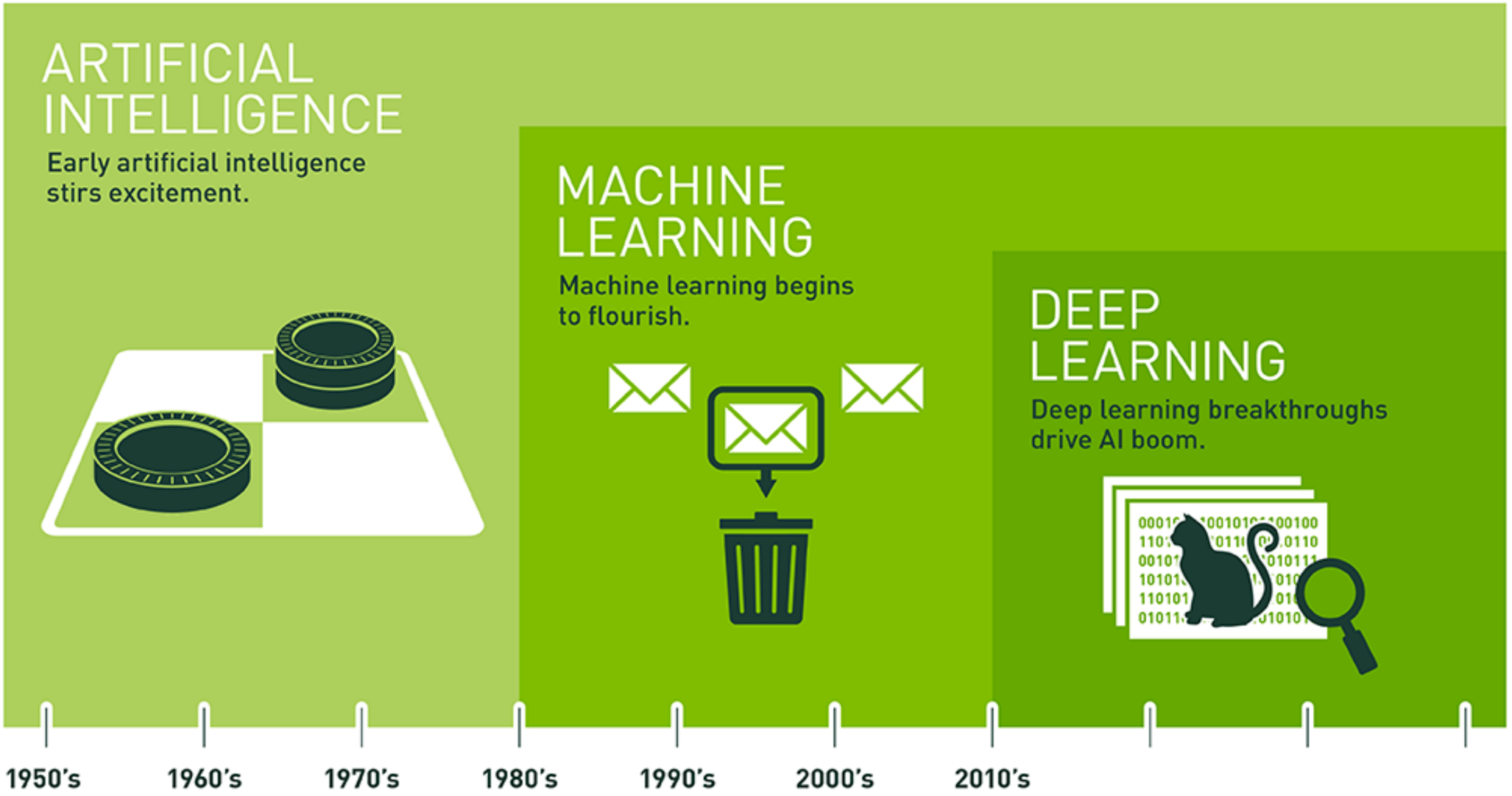
Deep Learning is one technique to perform Unsupervised Feature Learning

DEEP LEARNING VS. MACHINE LEARNING





EVOLUTION OF AI



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.