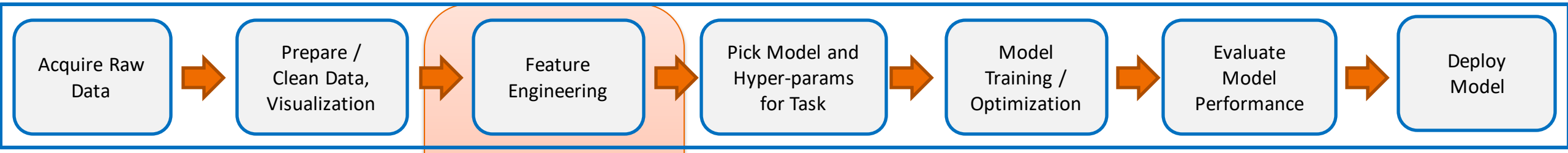
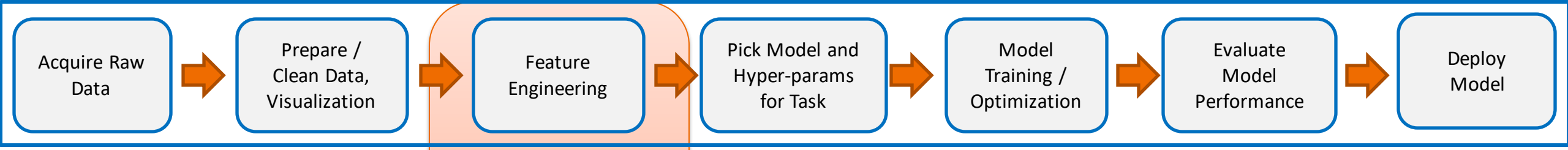
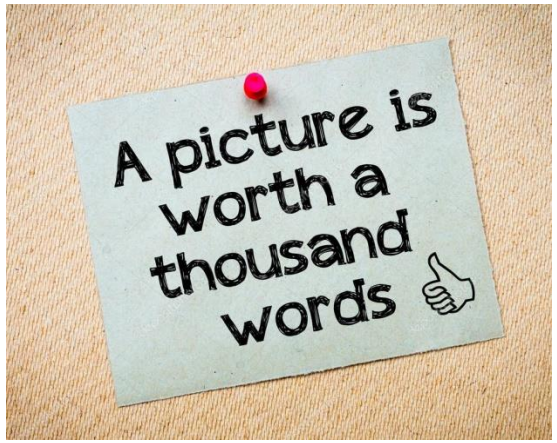
Focus for this lecture



- Image Representation
  - Hand-crafted features
  - Statistical features
  - Deep Learnt Features



AI and Problem of Perception



http://karpathy.github.io/2012/10/22/state-of-computer-vision/

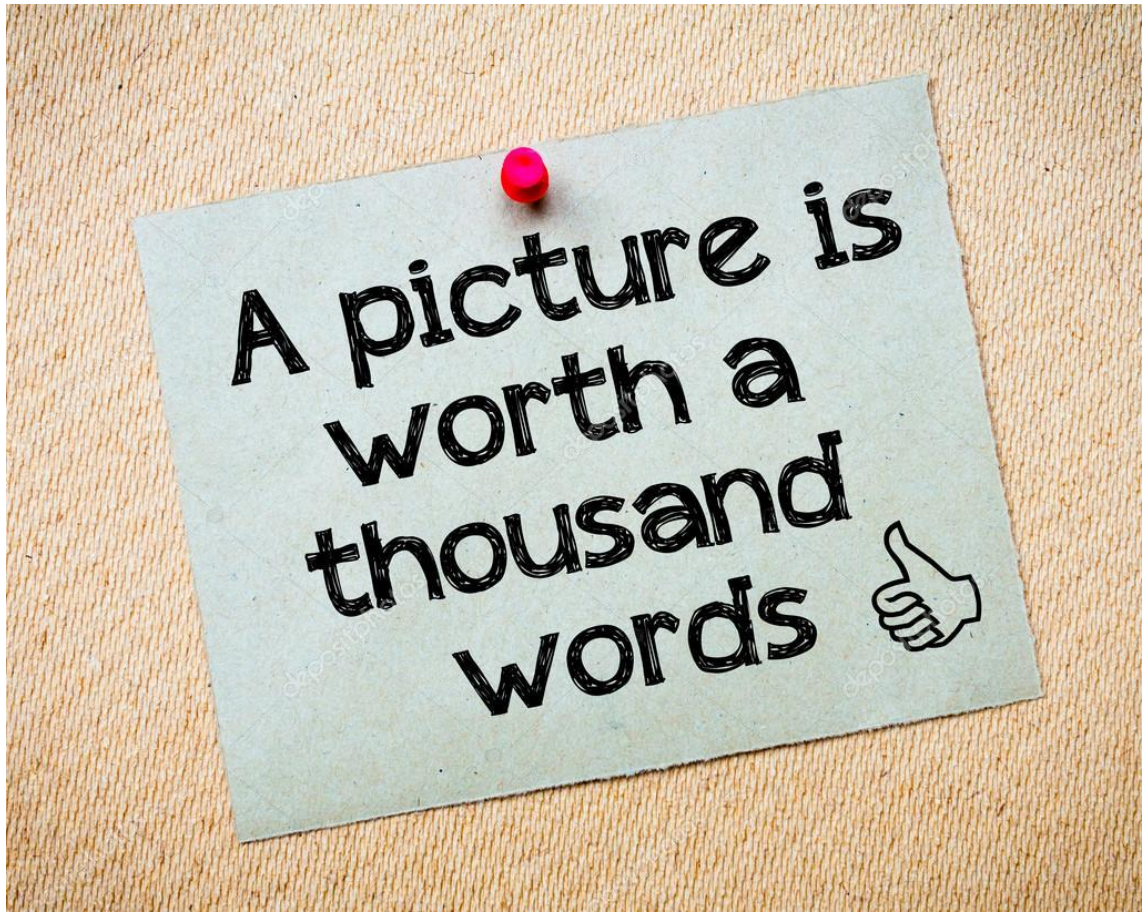
Why is it challenging?

Occlusions/Truncations

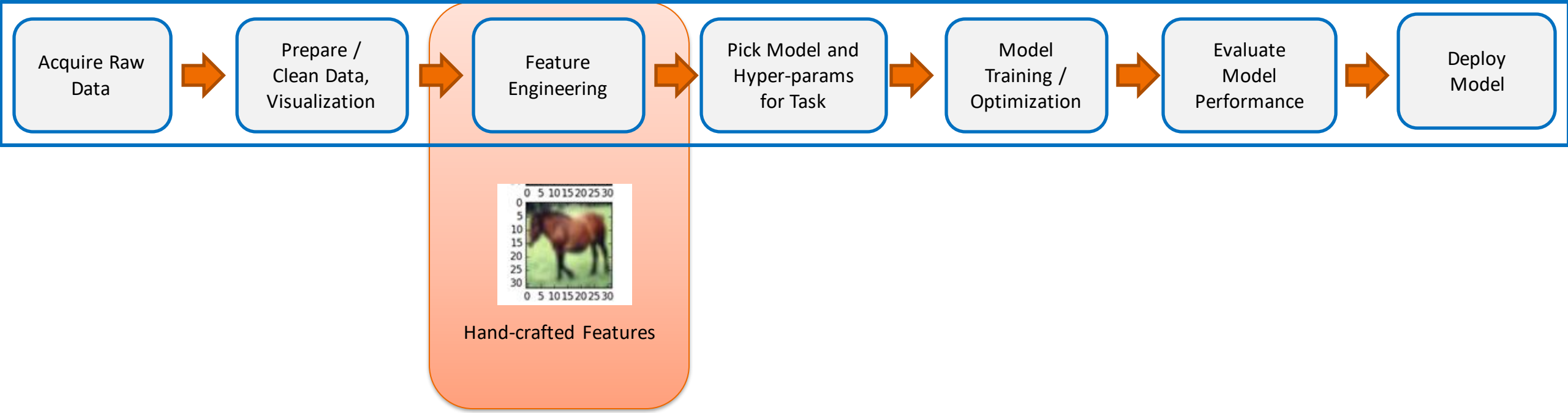
View Point Variation

Intra class variations

Inter class variations


A picture is worth a thousand words


TACO BELL
TACO BELL

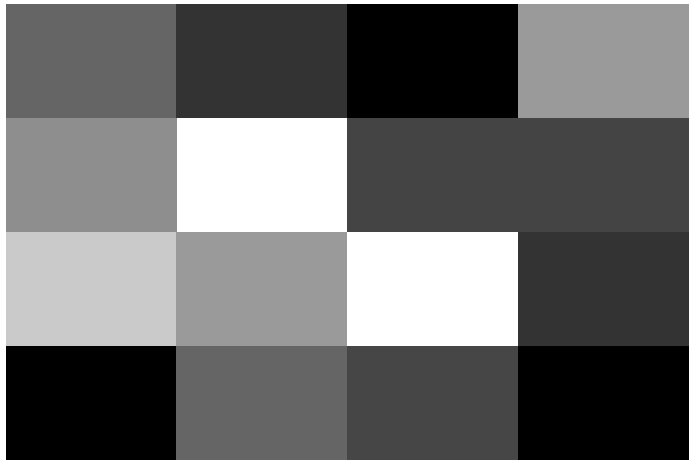


Hand-crafted features

What is a digital image?

- 2D matrix of intensities (gray or color values) or numbers

100	50	0	150
90	255	70	70
200	150	255	50
0	100	80	0



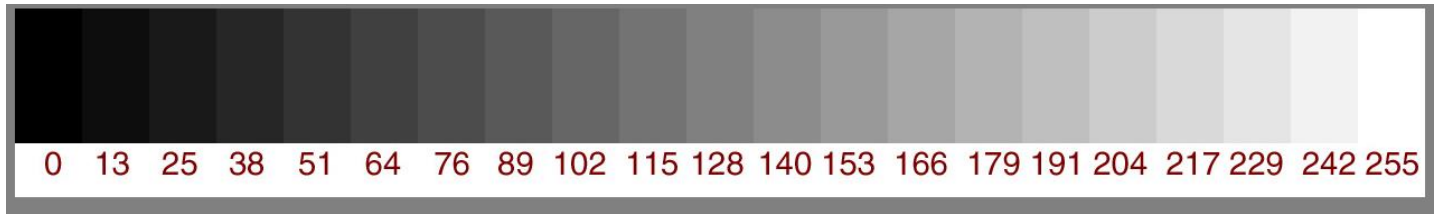
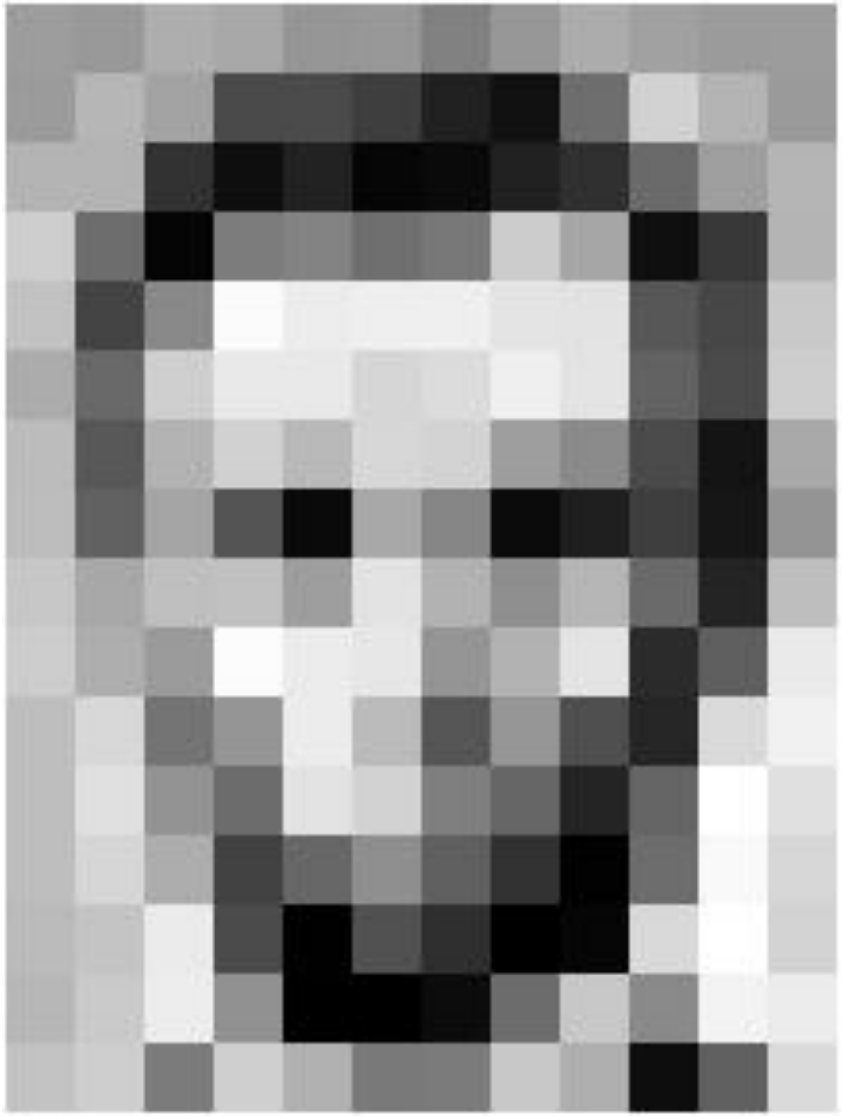
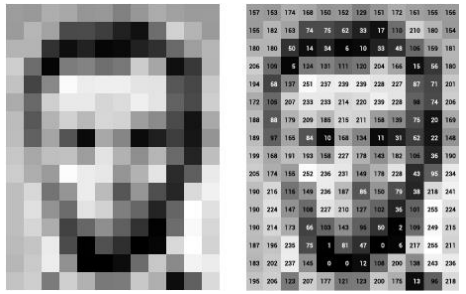


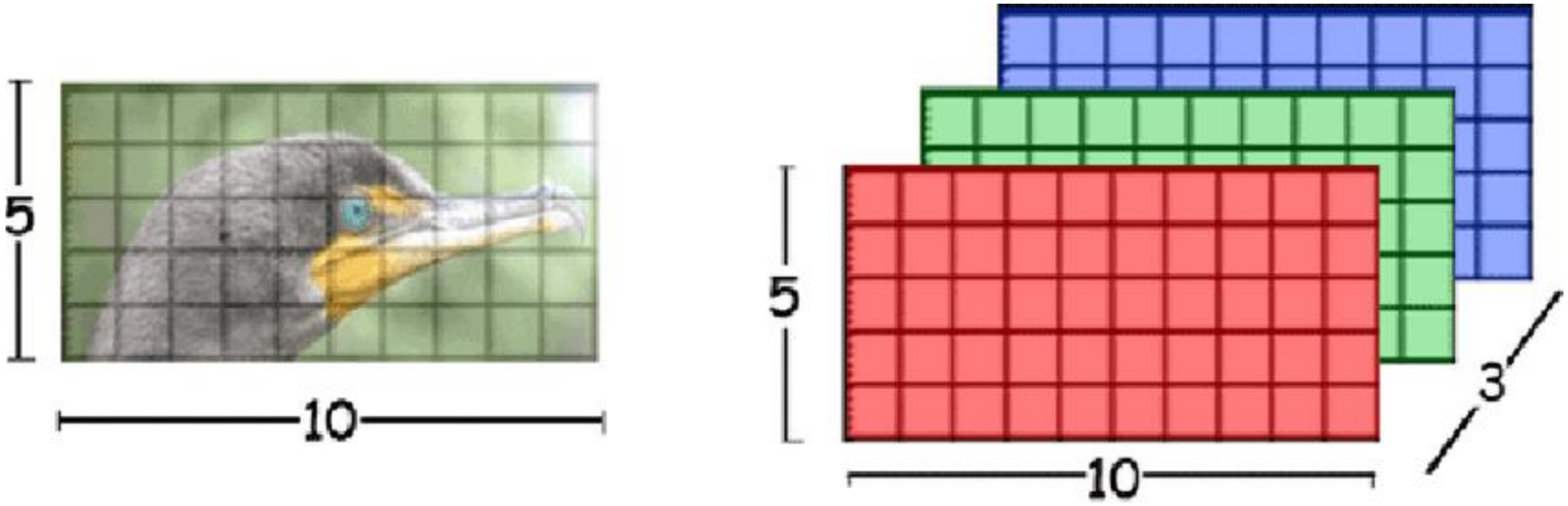
Image Representation



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	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

Image Representation





source: https://www.researchgate.net/figure/Color-image-representation-and-RGB-matrix_fig15_282798184

Example of an AI system which outputs high-level image description

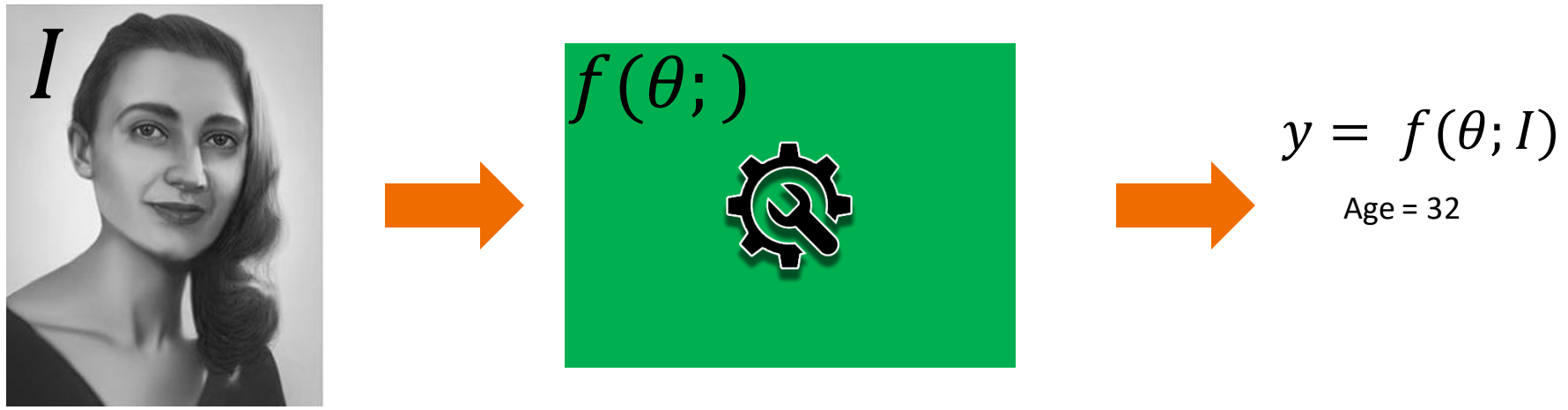
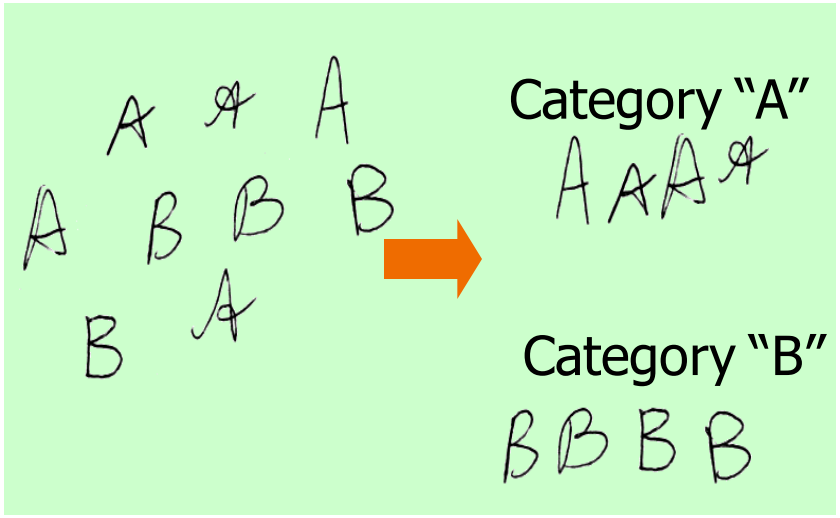


Image
or
Image Representation

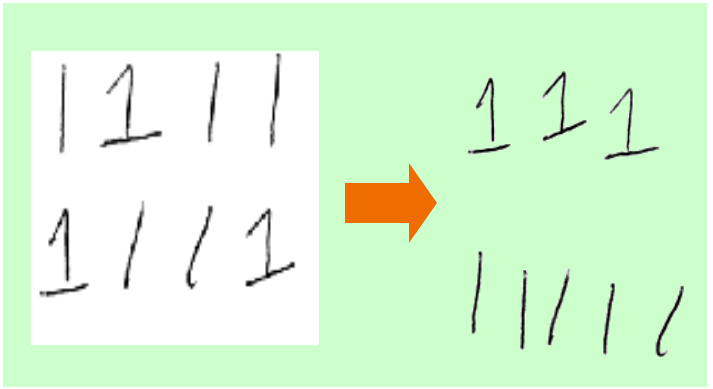
(High-level)
Image Description

Classification vs Clustering

- Classification (known categories)
- Clustering (unknown categories)

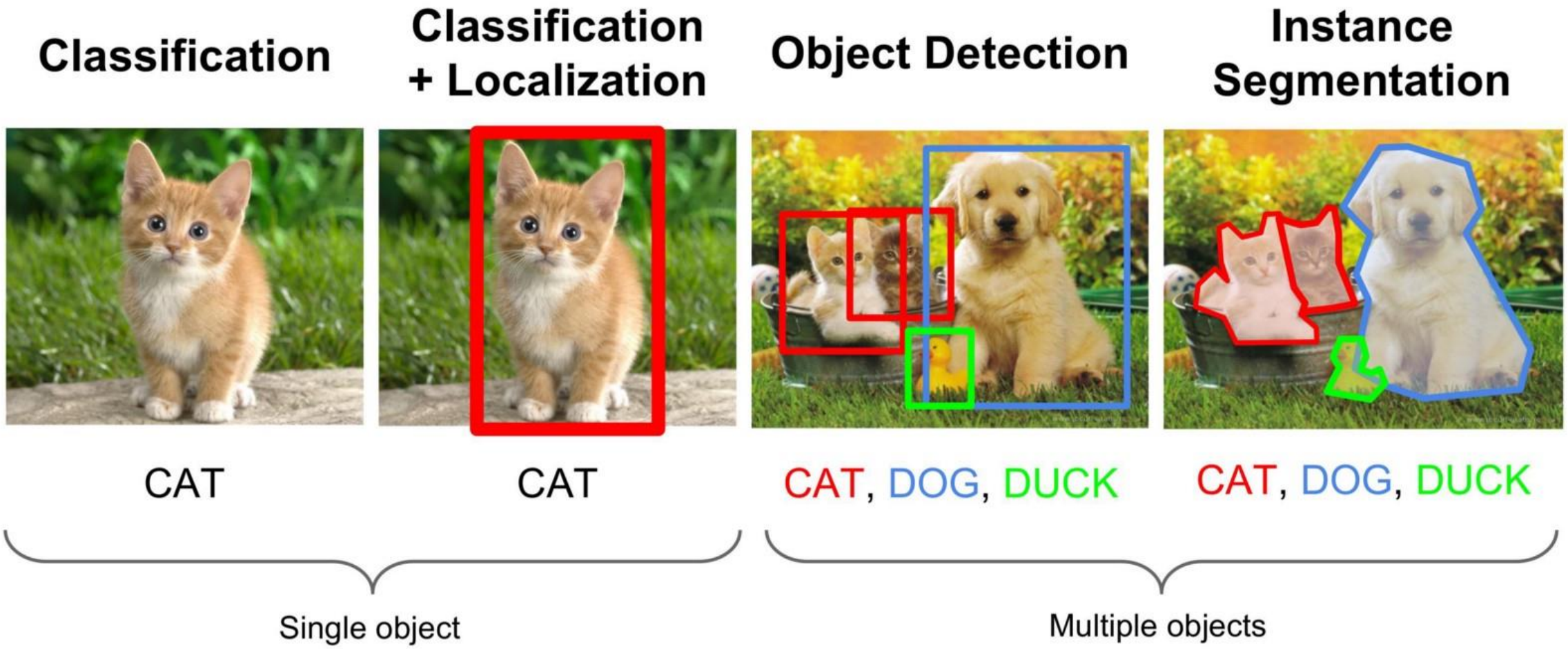


Classification (Recognition)
(Supervised Classification)



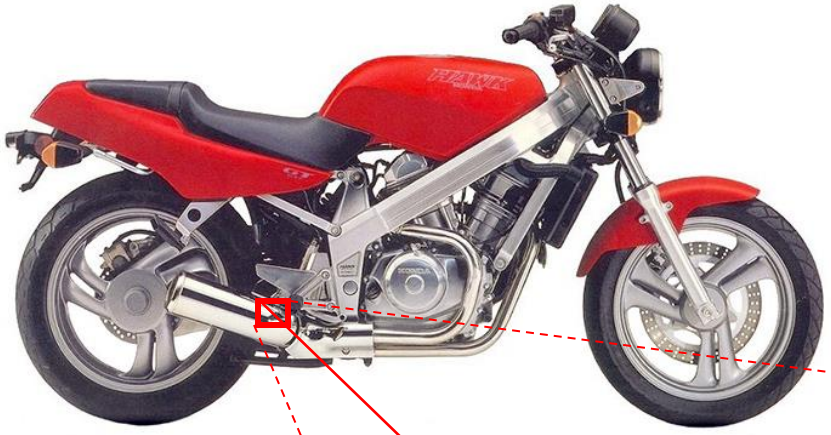
Clustering
(Unsupervised Classification)

Various kinds of ML problems involving images



https://cdn-images-1.medium.com/max/1600/1*Hz6t-tokG1niaUfmcysusw.jpeg

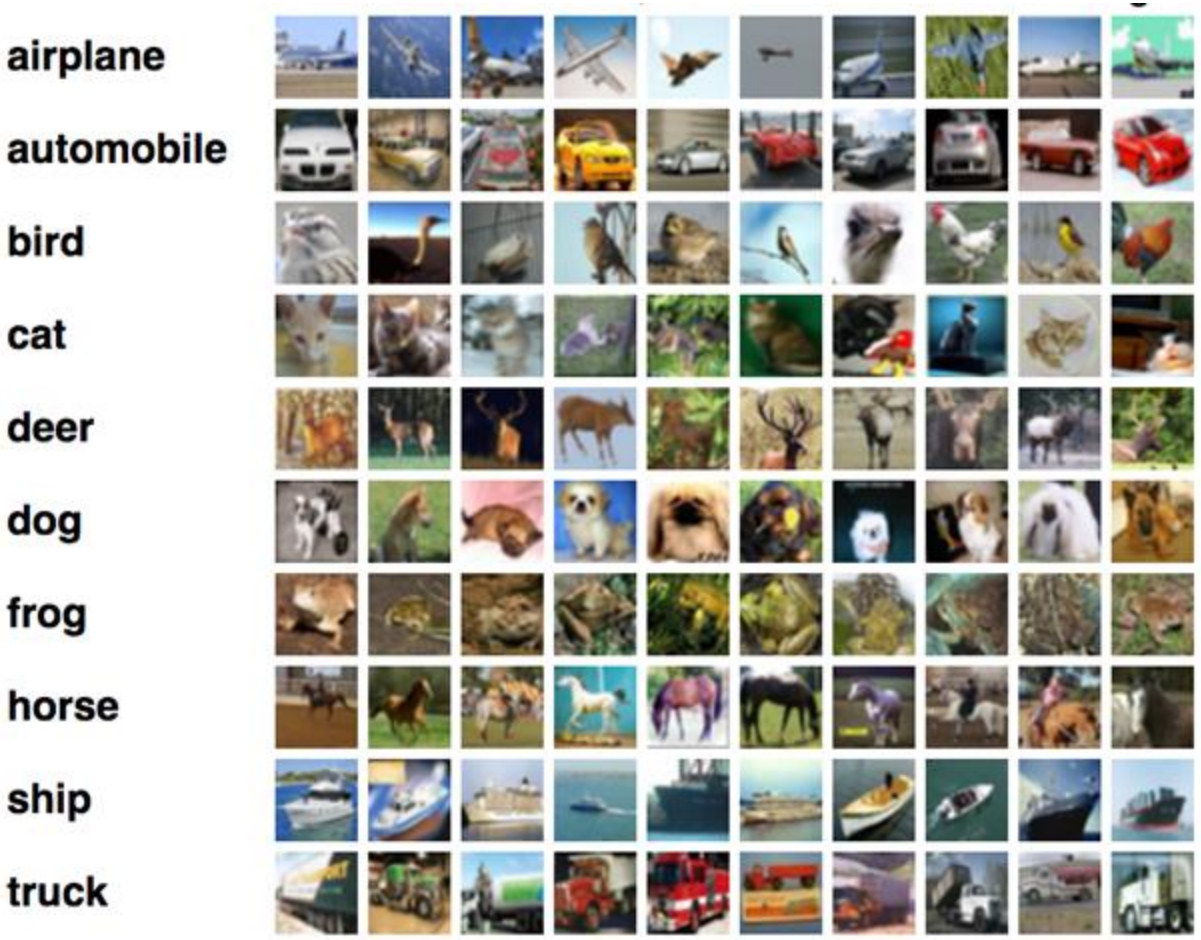
Why is this hard?



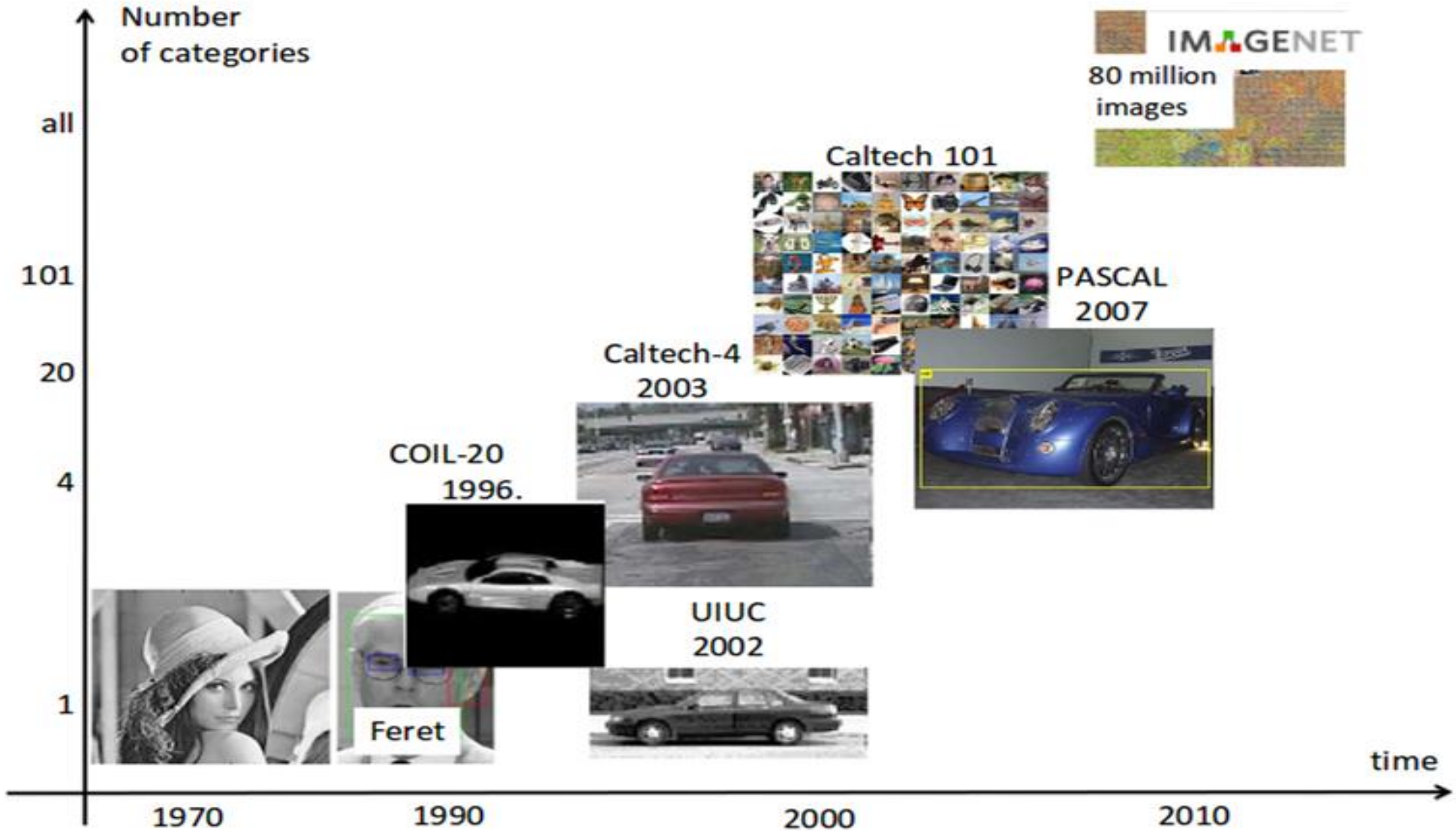
194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

A Typical Problem: CIFAR-10

- 10 classes. 50K Train. 10K Test



- Our “smaller” Lab problem
  - Automobile vs bird
  - Separate/recognize two classes
  - 10K Train 2K Test Samples


Number
of categories
all
101
20
4
1
IMAGENET
80 million
images
Caltech 101
PASCAL
2007
Caltech-4
2003
COIL-20
1996.
UIUC
2002
Feret
time
1970
1990
2000
2010

A Naïve Attempt

- If (image has green/grass)
  - It is an animal
    - if (it is tall) it is a horse
- If (image has blue)
  - It is either airplane or bird
    - If (…)
- And so on ..

- Any hope of this working? ☺

Human knowledge

Human knowledge

Possible Features: Handcrafting

MIN RED
MAX RED
MEAN RED
MIN GREEN
MAX GREEN
MEAN GREEN
MIN BLUE
MAX BLUE
MEAN BLUE

9 X 1
FEATURE VECTOR
PER IMAGE

Concerns:

- **Too naïve to capture the visual content?**
- **Too small to represent information?**

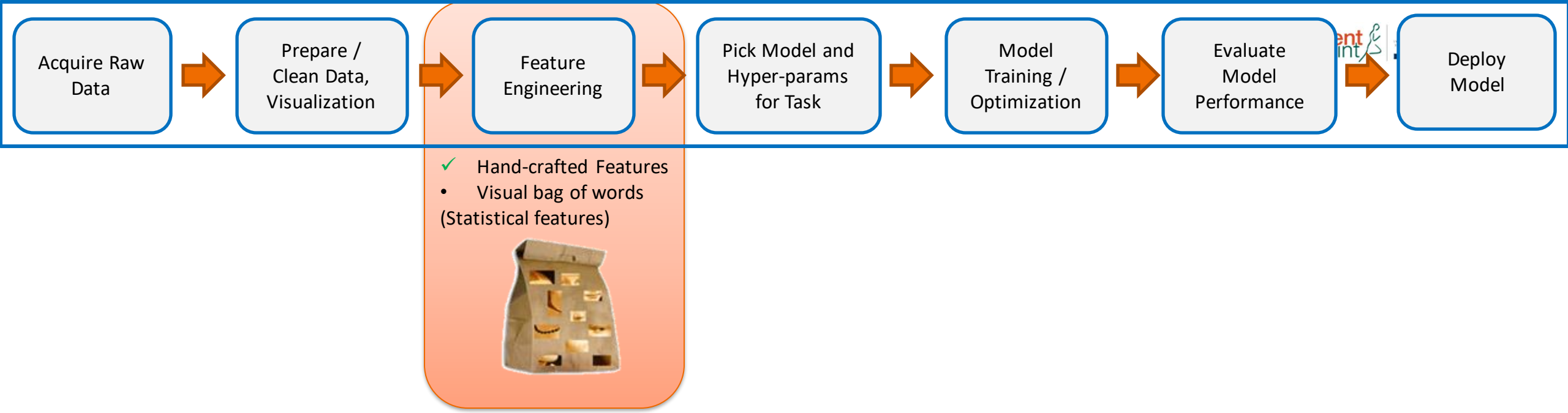
Possible Features: Raw Data Itself

3072 X 1
vector

FEATURE VECTOR
32 X 32 X 3 = 3072
DIMENSION
PER IMAGE (d = 3072)

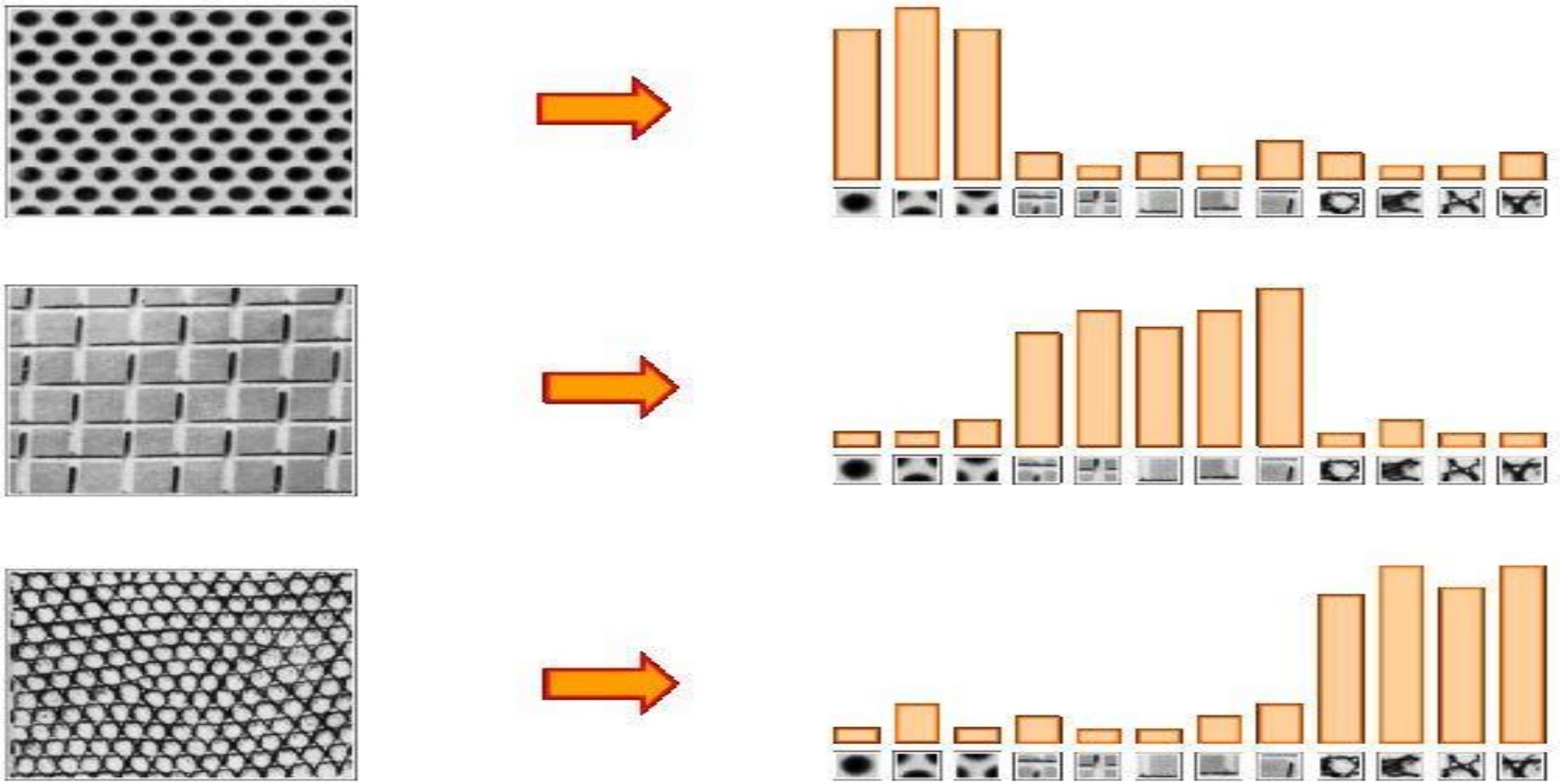
CONCERNS:

- Too big ?
- May be redundancy ?
- Too rigid?



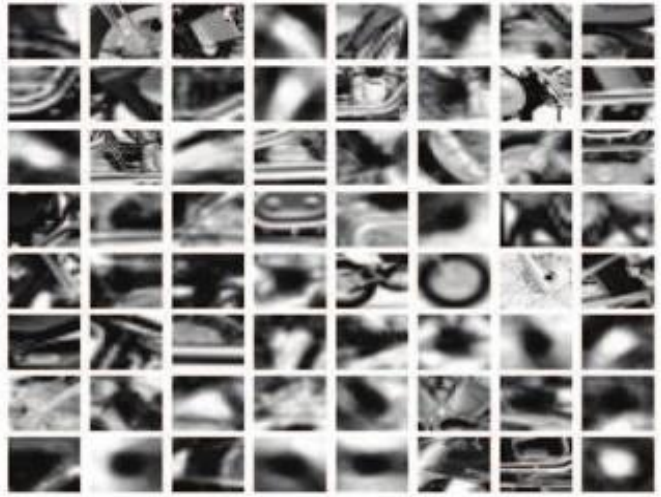
Visual Bag of Words

Visual BoW: Basic Idea



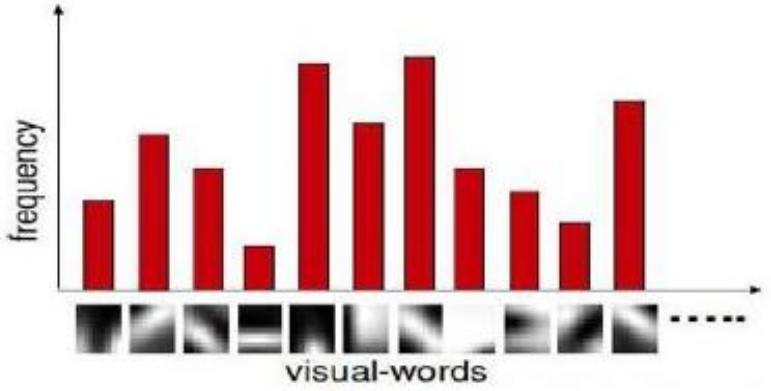
Source: wordpress.com/2013/04/24/research-bag-of-features-for-visual-recognition

Bag of Visual Words

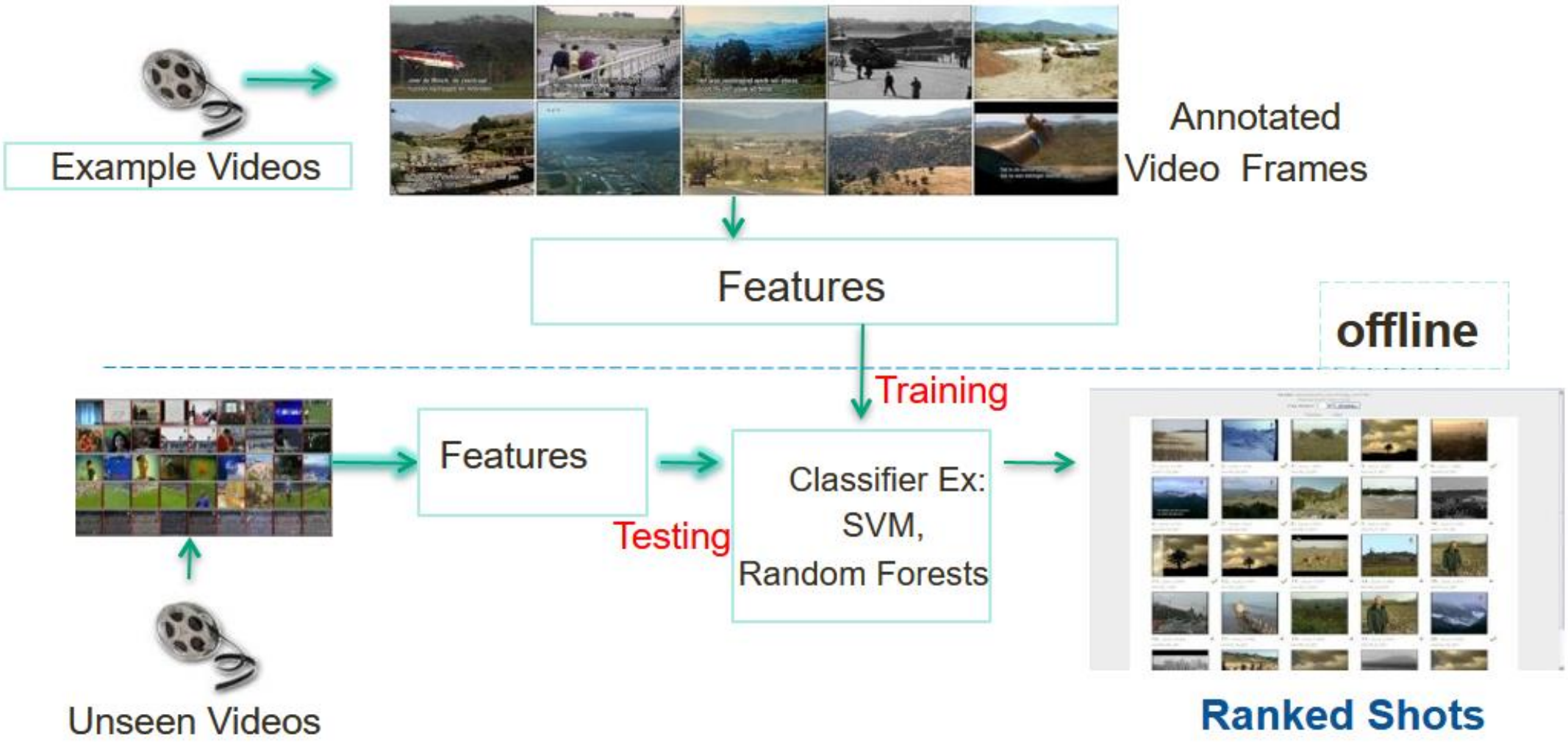


Learned Visual

Vocabulary



Example: Search in a huge video DBs



Person Riding Bicycle: Results

SVM Classifier, Intersection Kernel, BoW feature

Cityscape: Results

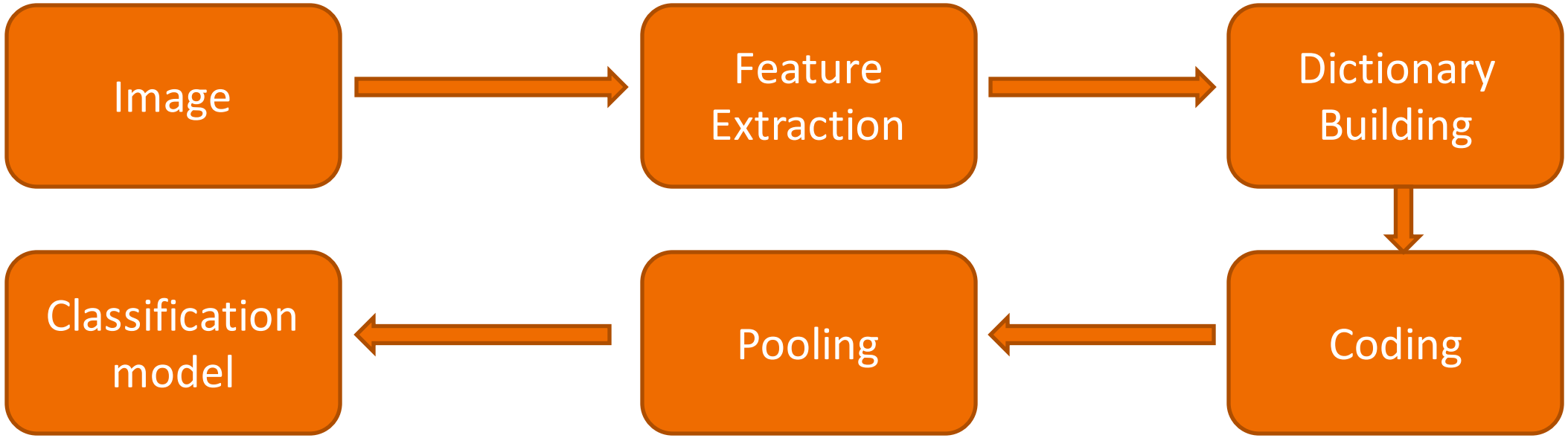
SVM Classifier, Intersection Kernel, BoW feature

Female Face Close-up: Results

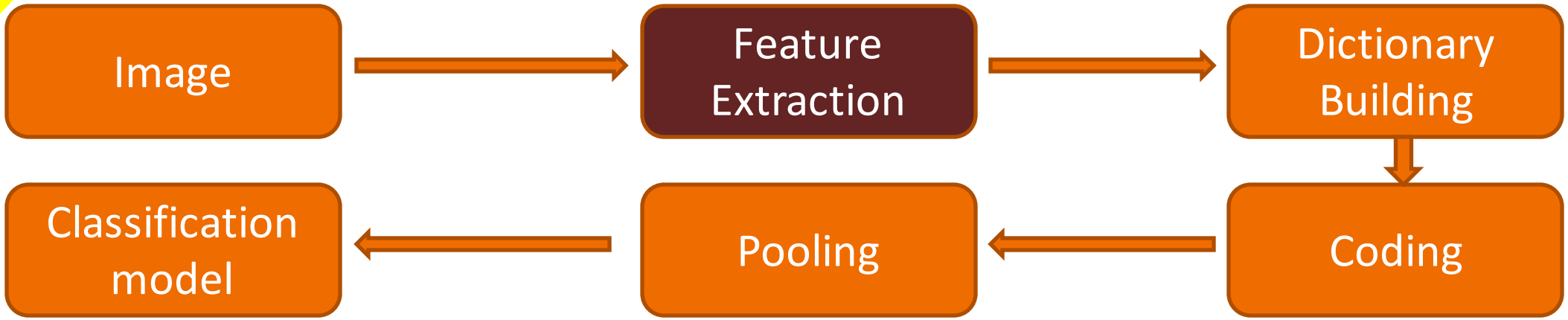
SVM Classifier, Intersection Kernel, BoW feature

Code Available

Bag of Words model

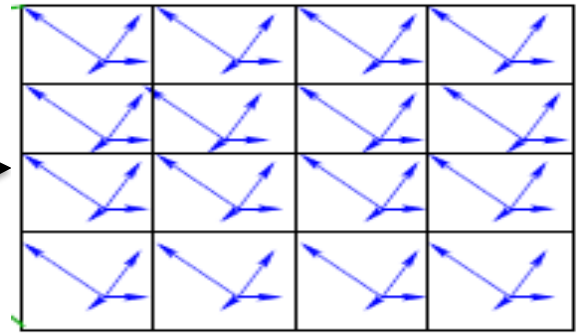


Code Available



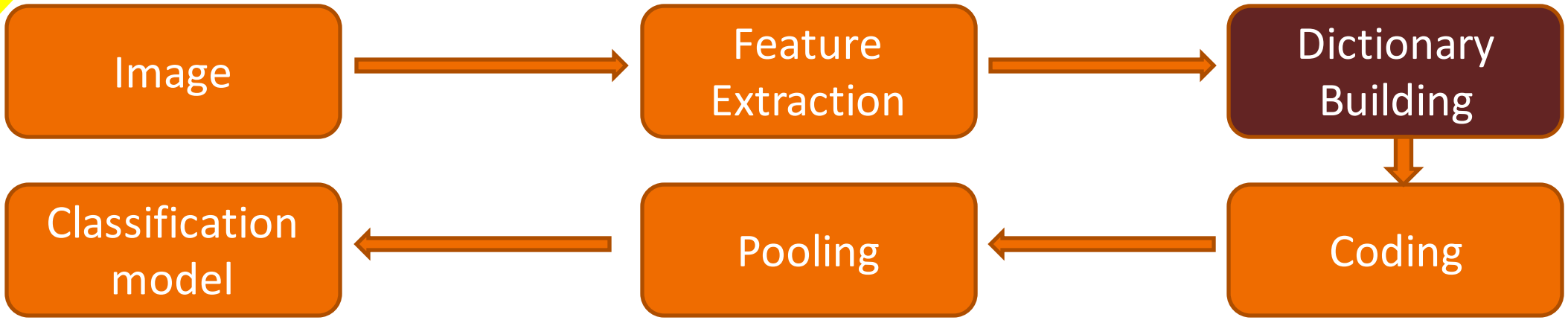
Image

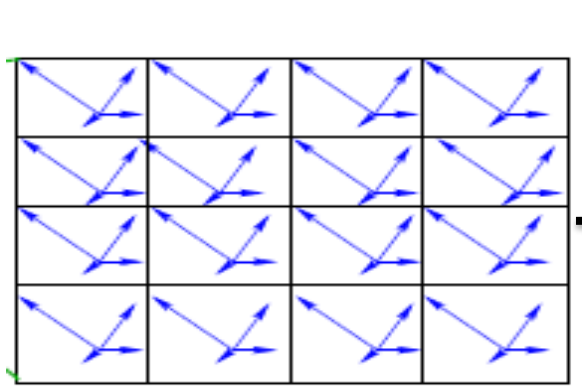
Interest point Detection



SIFT (128D) Features
for each interest point

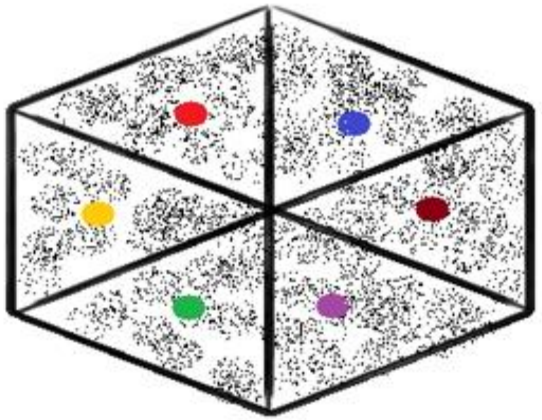
Code Available





Group/Cluster (eg. Kmeans)

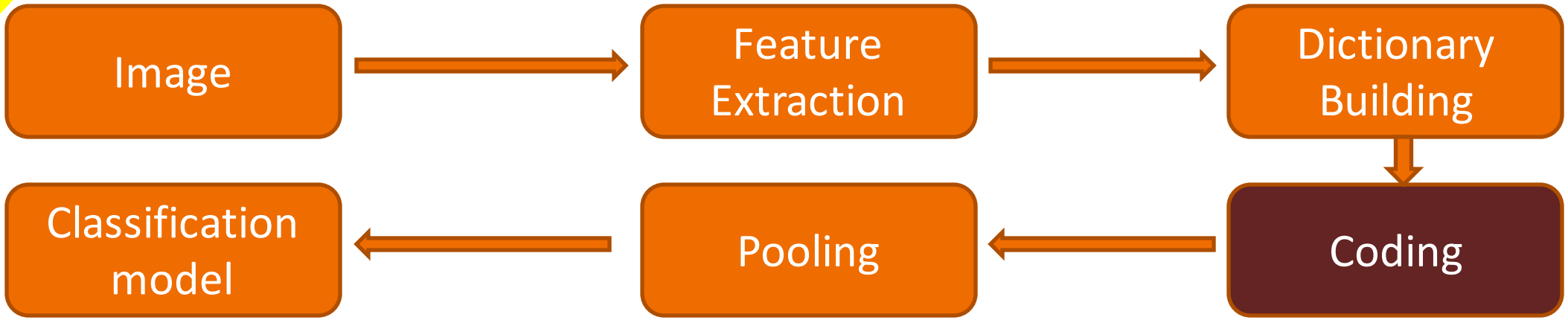
Discover Words

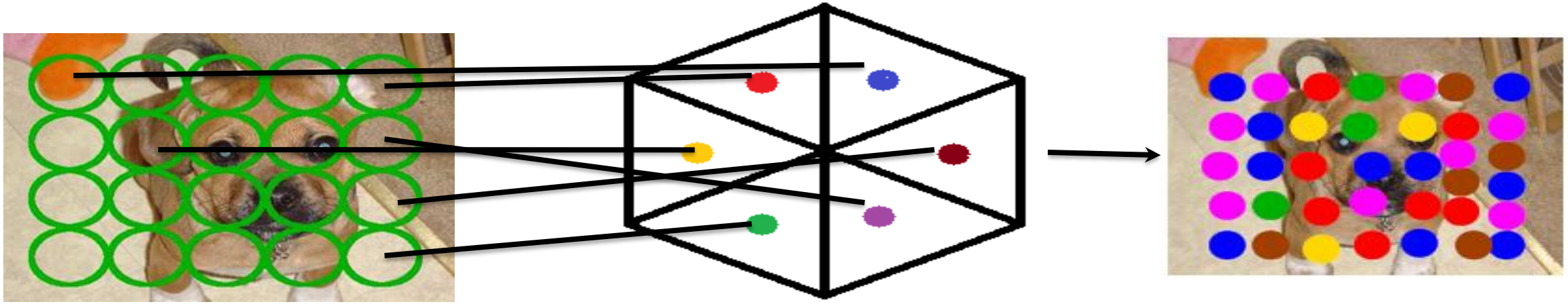


SIFT Features

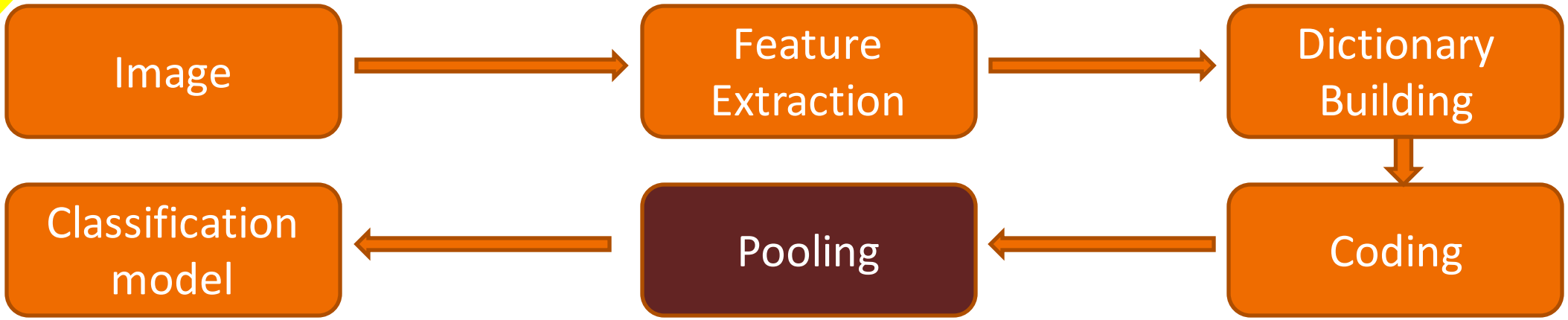
Vector Quantization

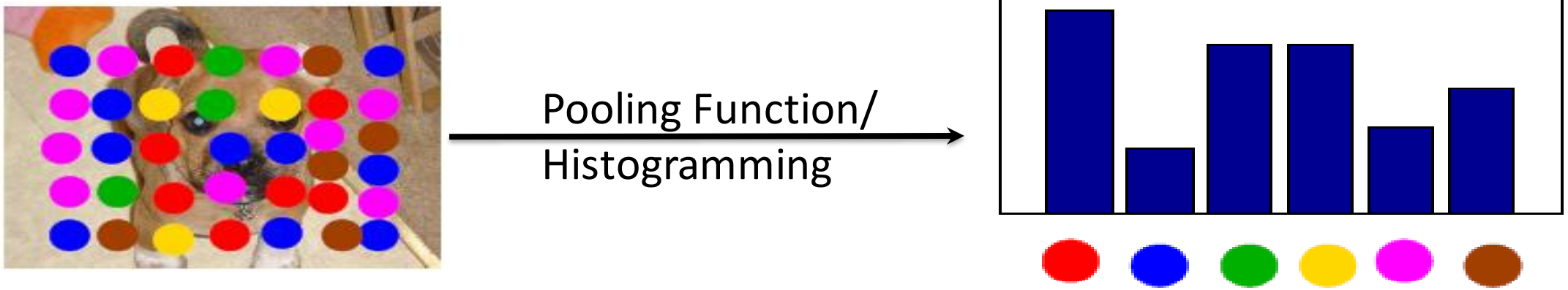
Code Available





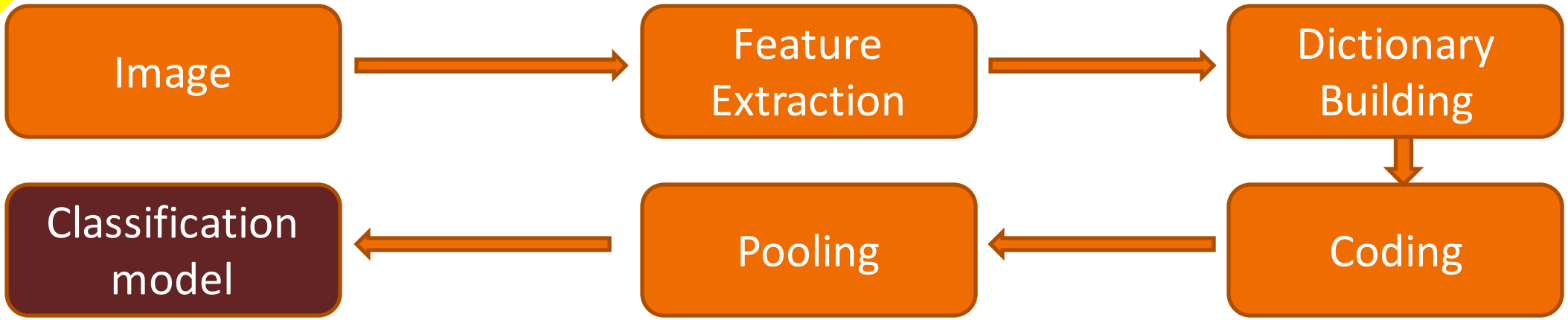
Dictionary/Codebook

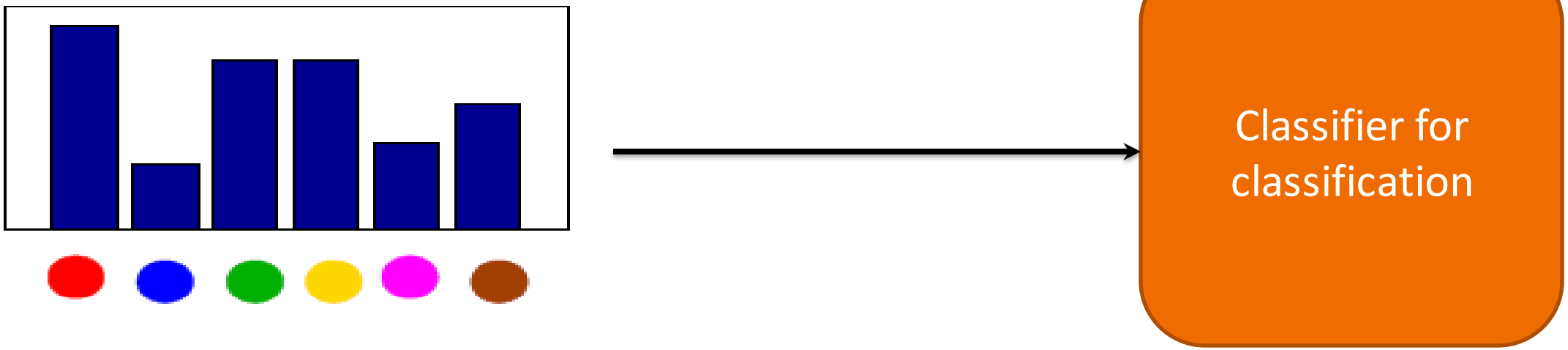




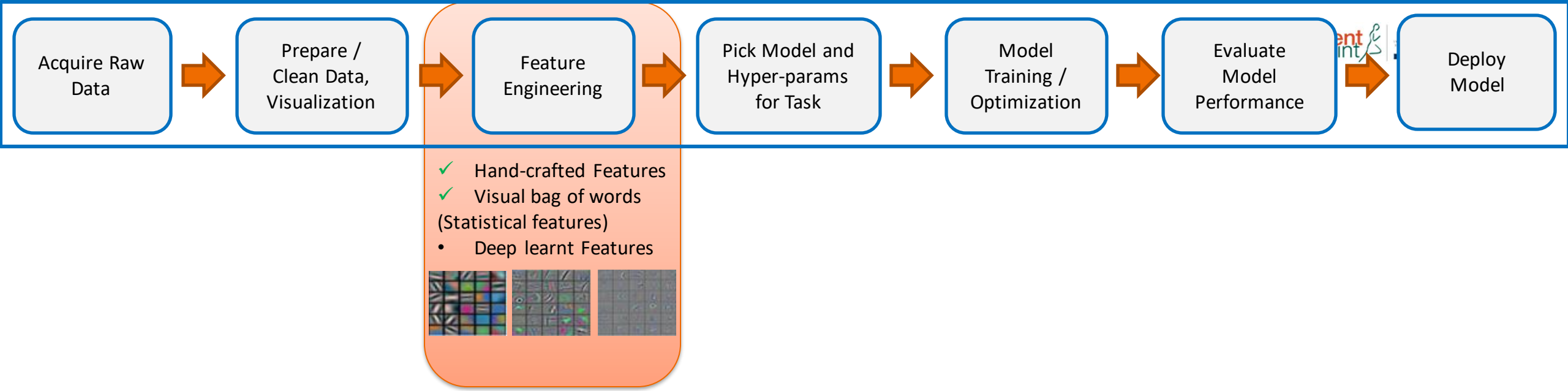
Histogram of Visual Words / Image Representation

Code Available





Histogram of Visual Words / Image Representation

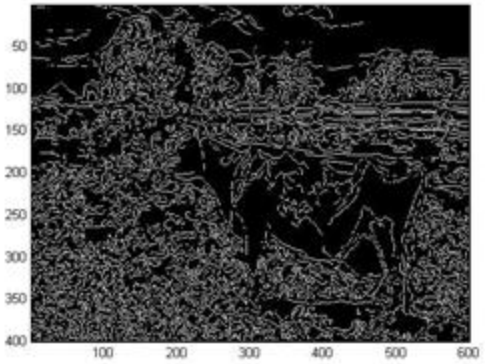


Deep Features

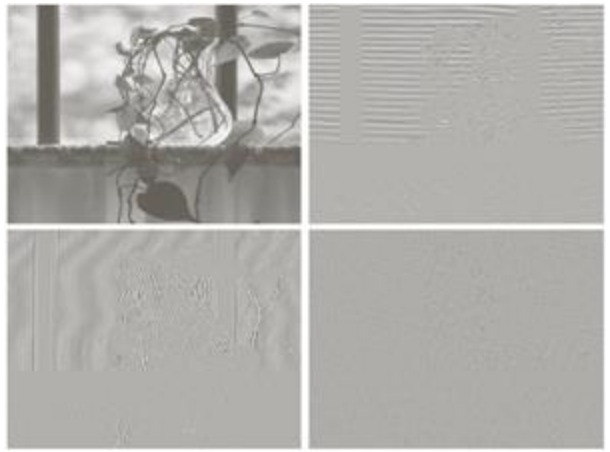
Neural Network, Deep Learning and Deep Features

Image Representation

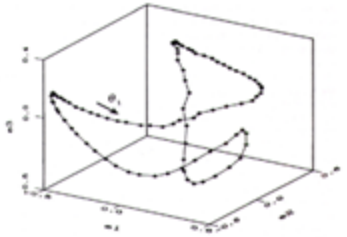
Features: Classical



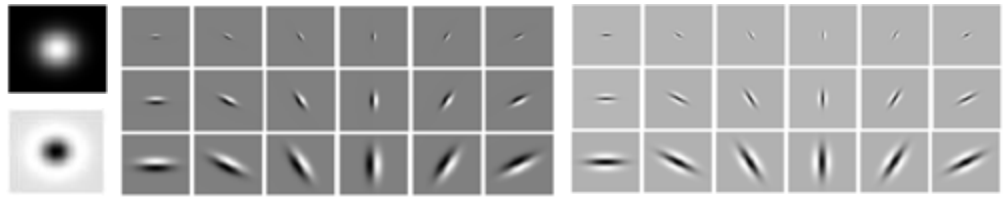
Edges and Corners: Sobel, LoG and Canny

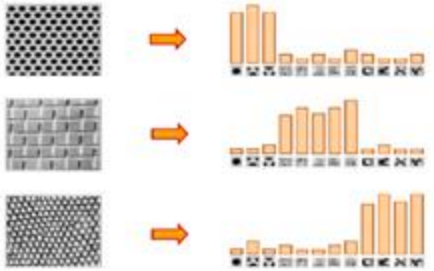


Fourier and Wavelet



PCA, Subspaces and Manifolds

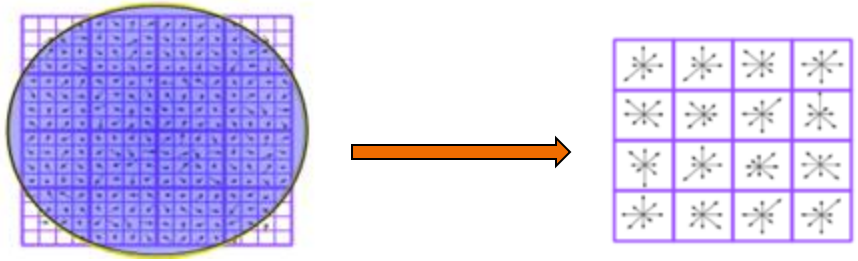




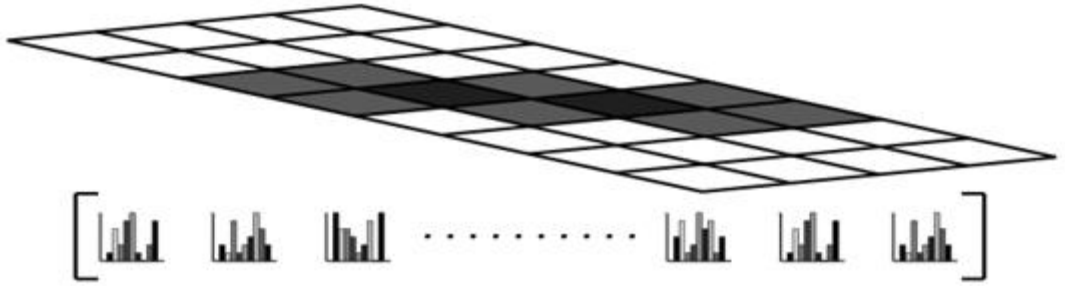
Texture; Filter bank; Histogram of responses

http://vision.stanford.edu/teaching/cs131_fall1718/files/13_notes.pdf

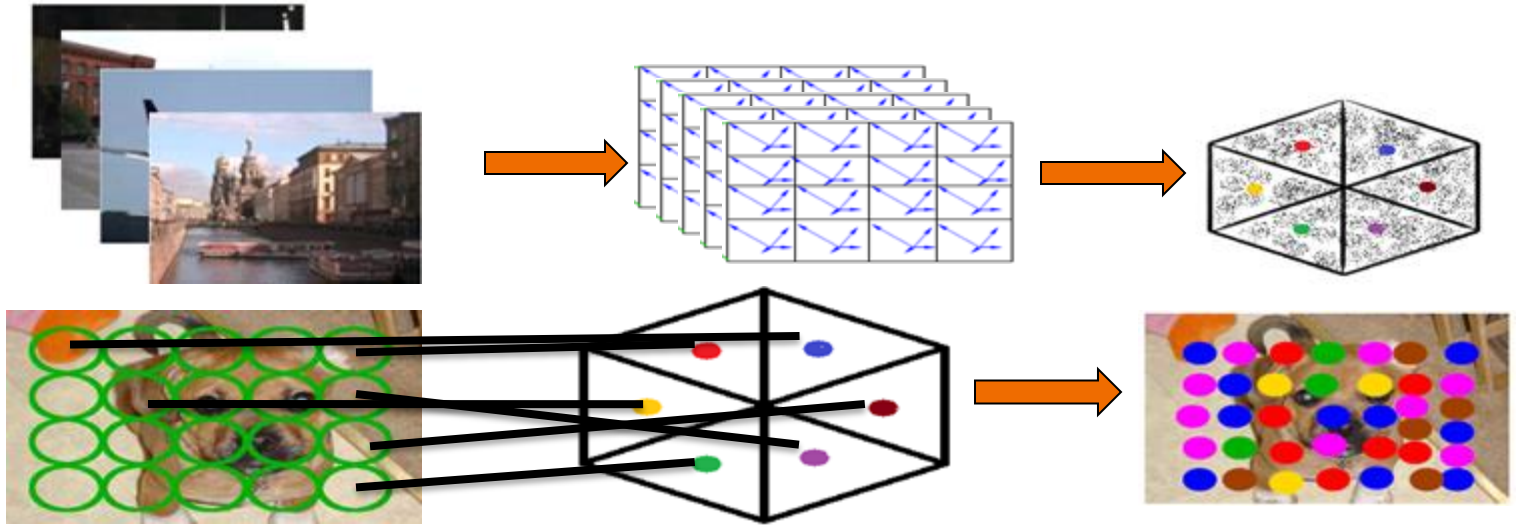
Well Engineered Features



SIFT (Lowe 1999, 2004)



HOG(Dalal and Triggs 2005)



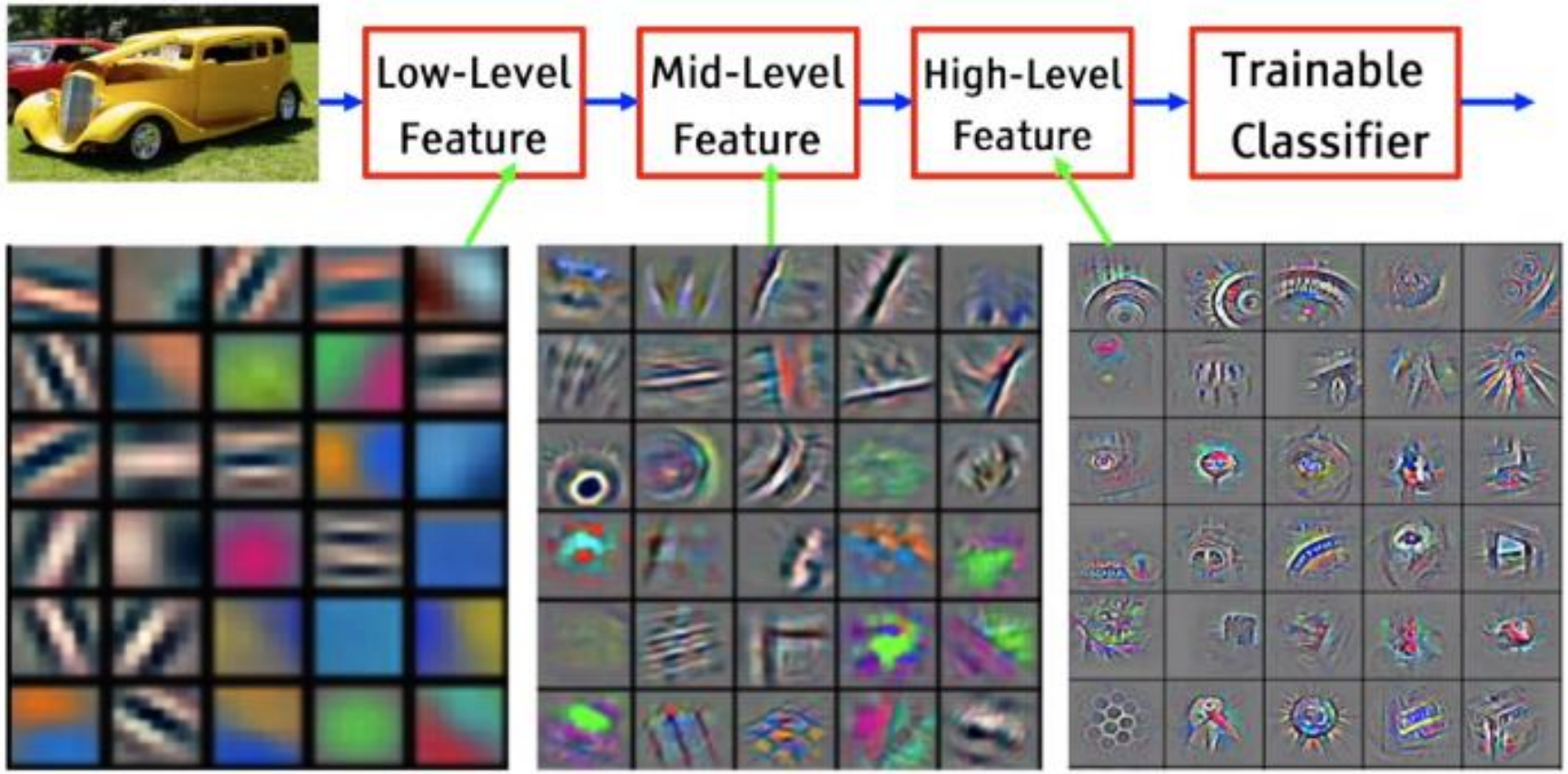
Bag of Words (Sivic and Zisserman 2003)

Focus: Dictionary Learning, Pooling and Coding

Mid-Level Features (2012-2013)

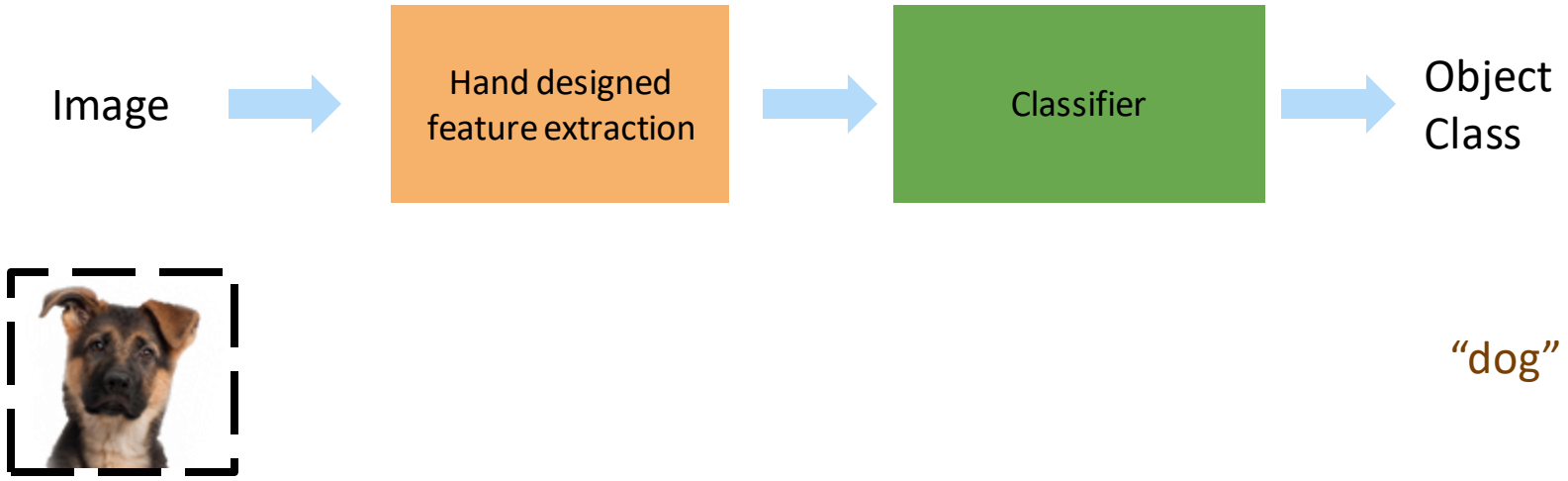
Deep Learnt Features (2013-XXX)

- It's deep if it has more than one stage of non-linear feature transformation.

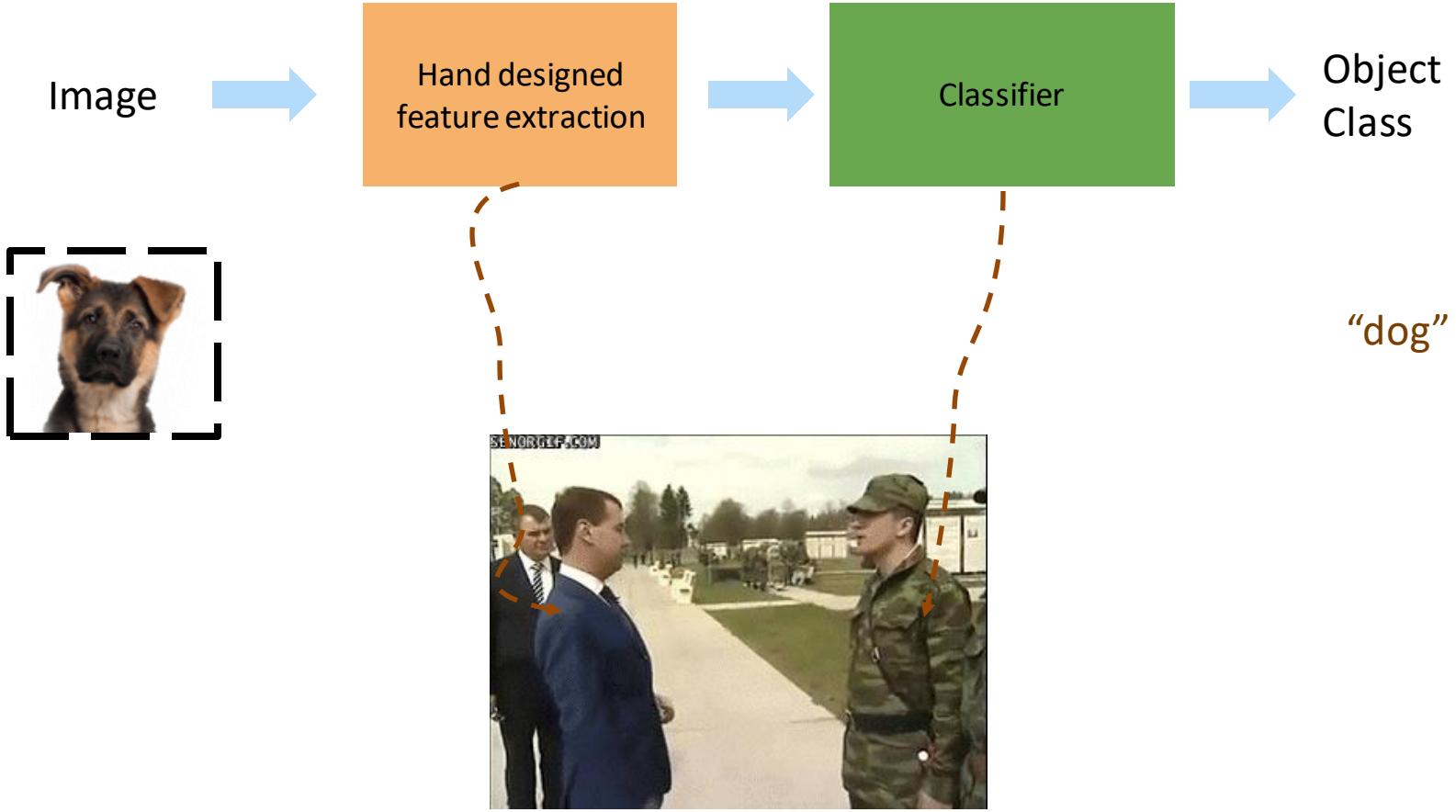


Source: Yann LeCun

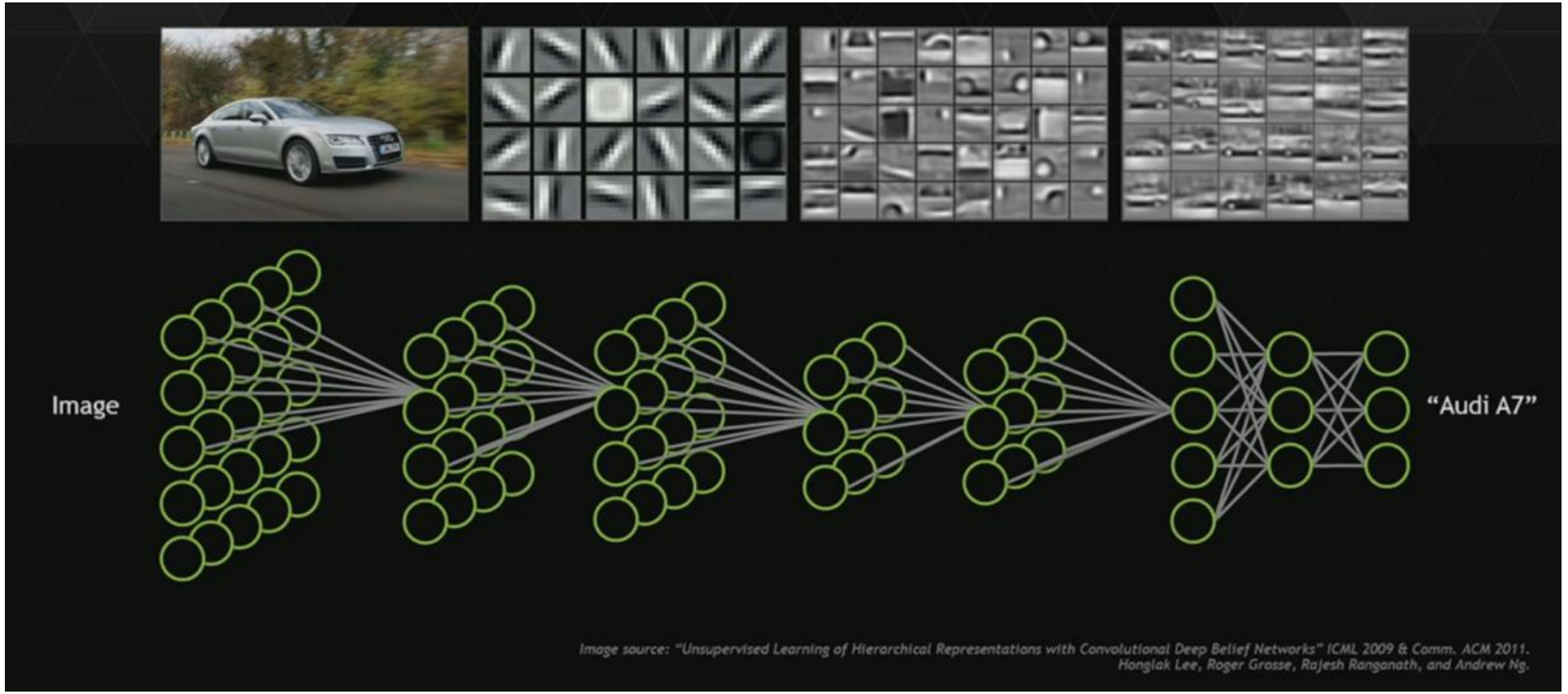
Object-recognition: conventional approach



Object-recognition: conventional approach

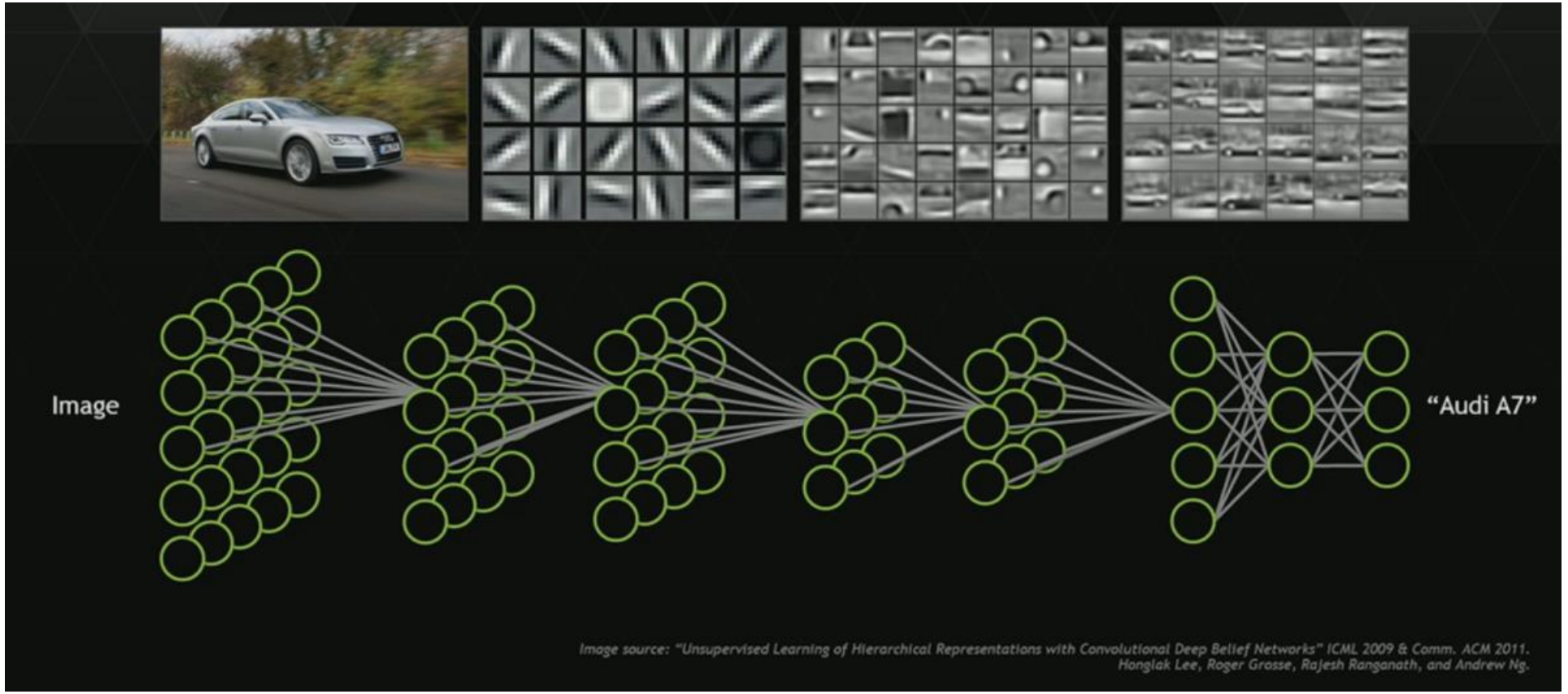


Object Recognition: Deep Neural Networks



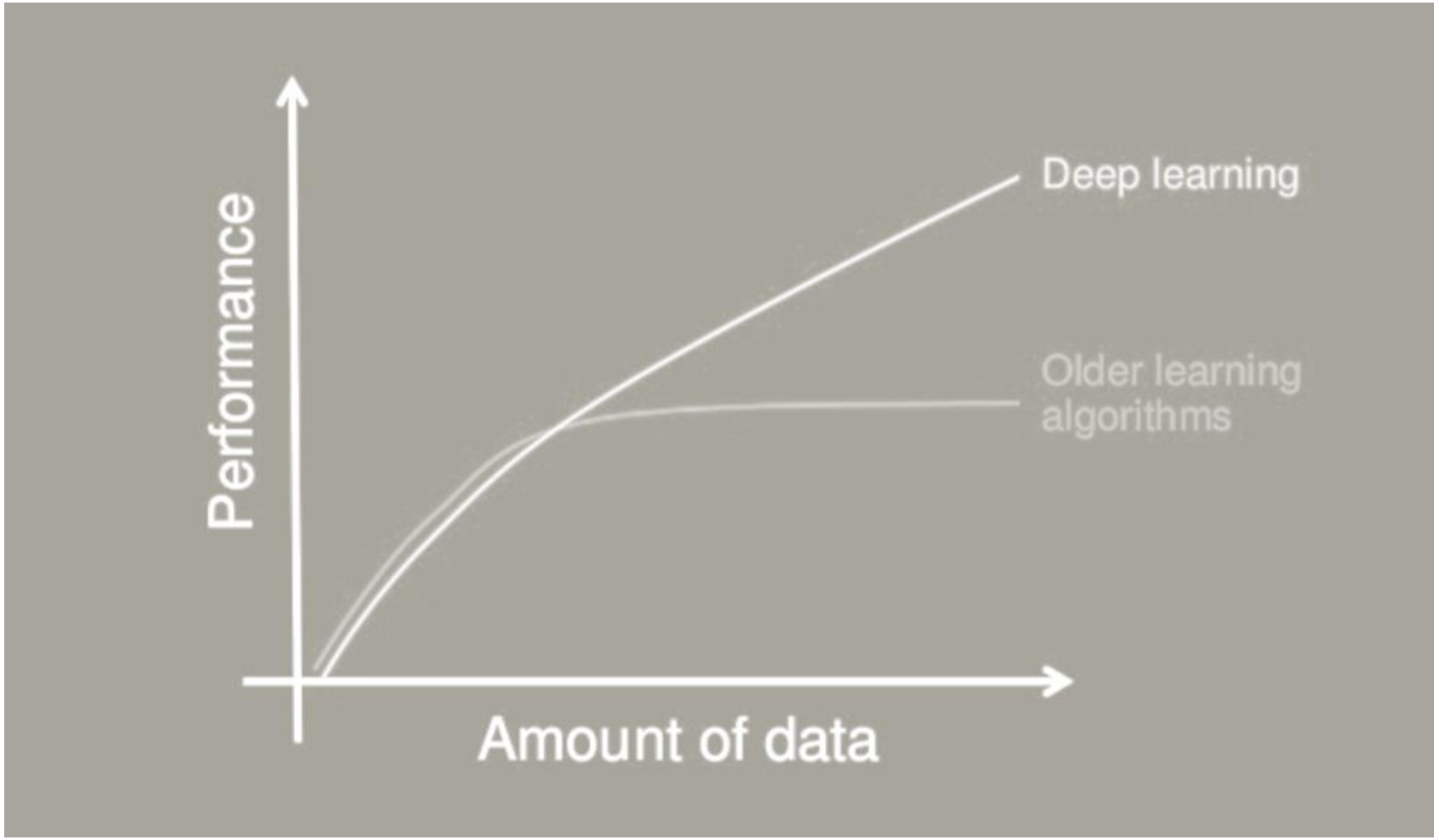
Data-driven, End-to-End learning

Object Recognition: Deep Neural Networks



Data-driven, End-to-End learning, Task-specific feature hierarchy

Why deep learning



Summary

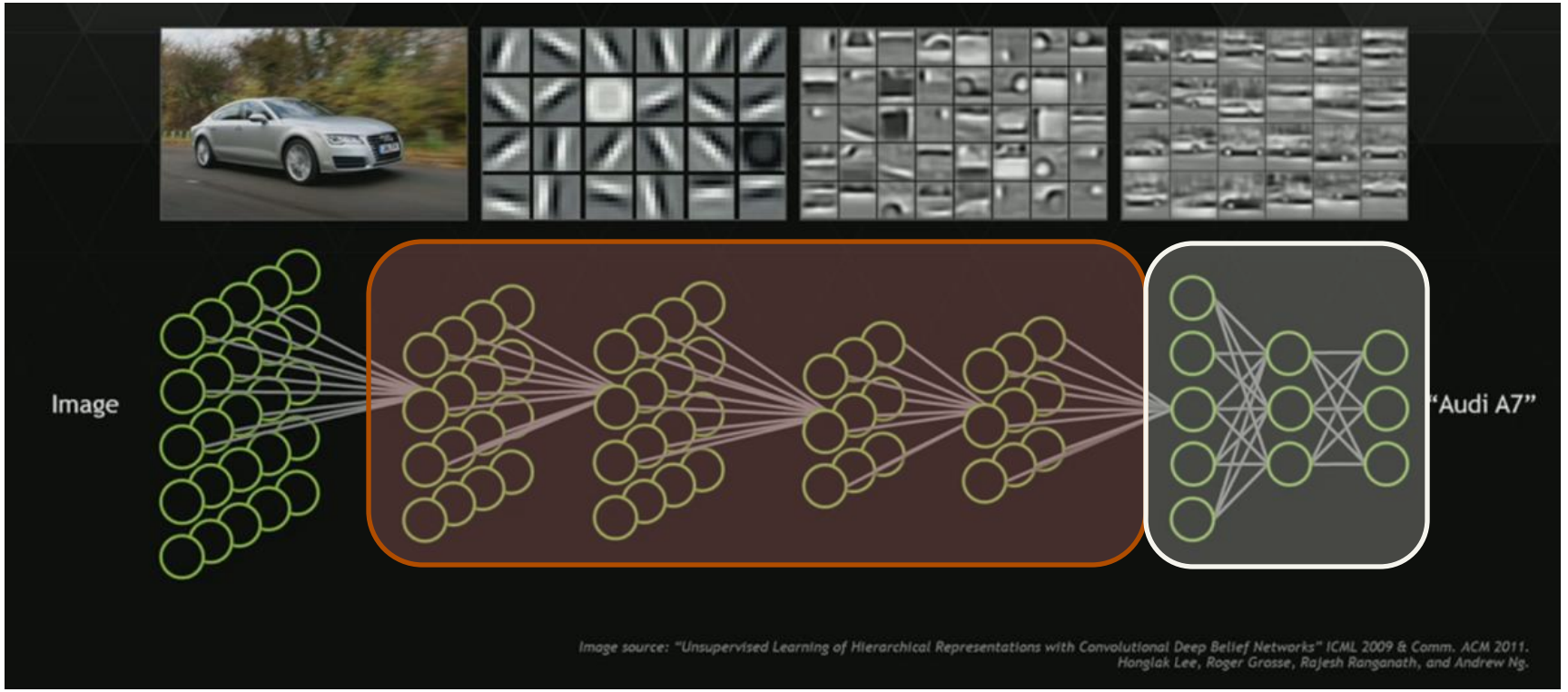
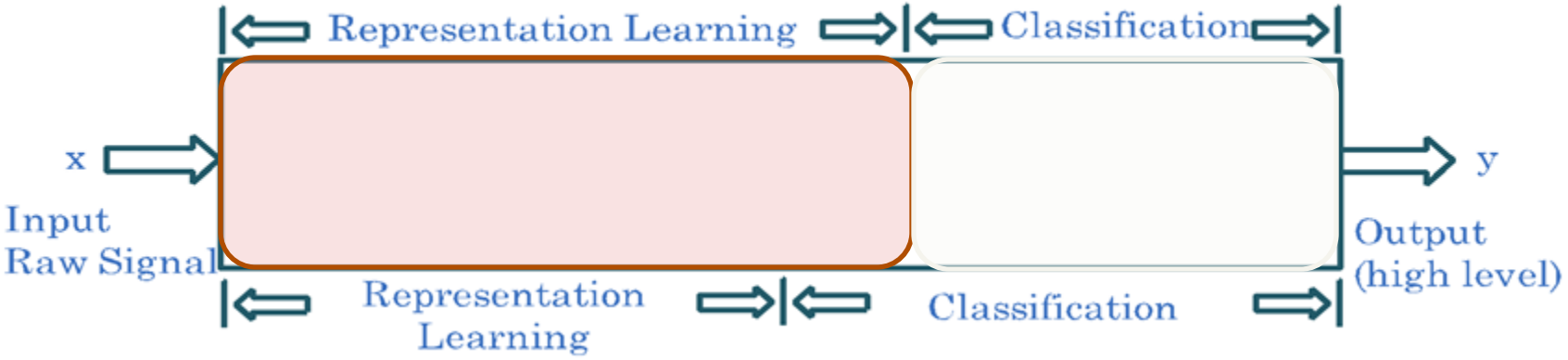
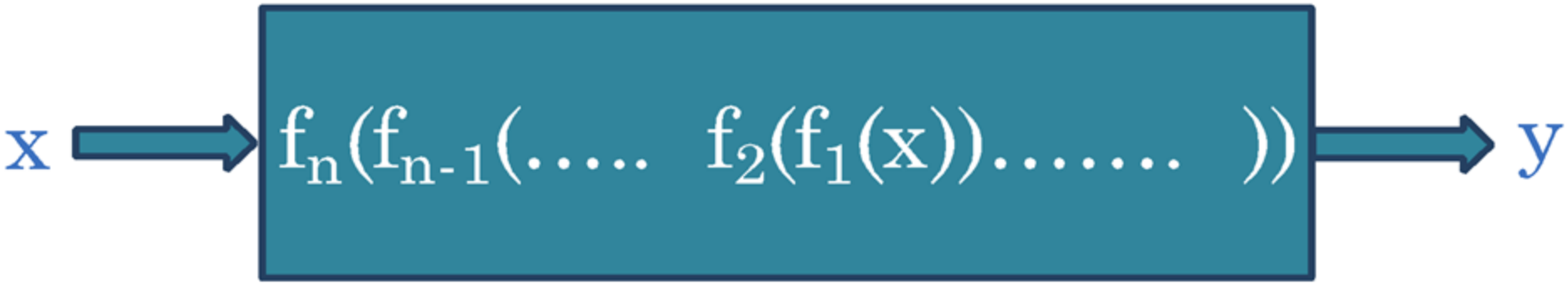


Image
"Audi A7"
Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011.
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.


Representation Learning
Classification
x
y
Input
Raw Signal
Output
(high level)
Representation
Learning
Classification

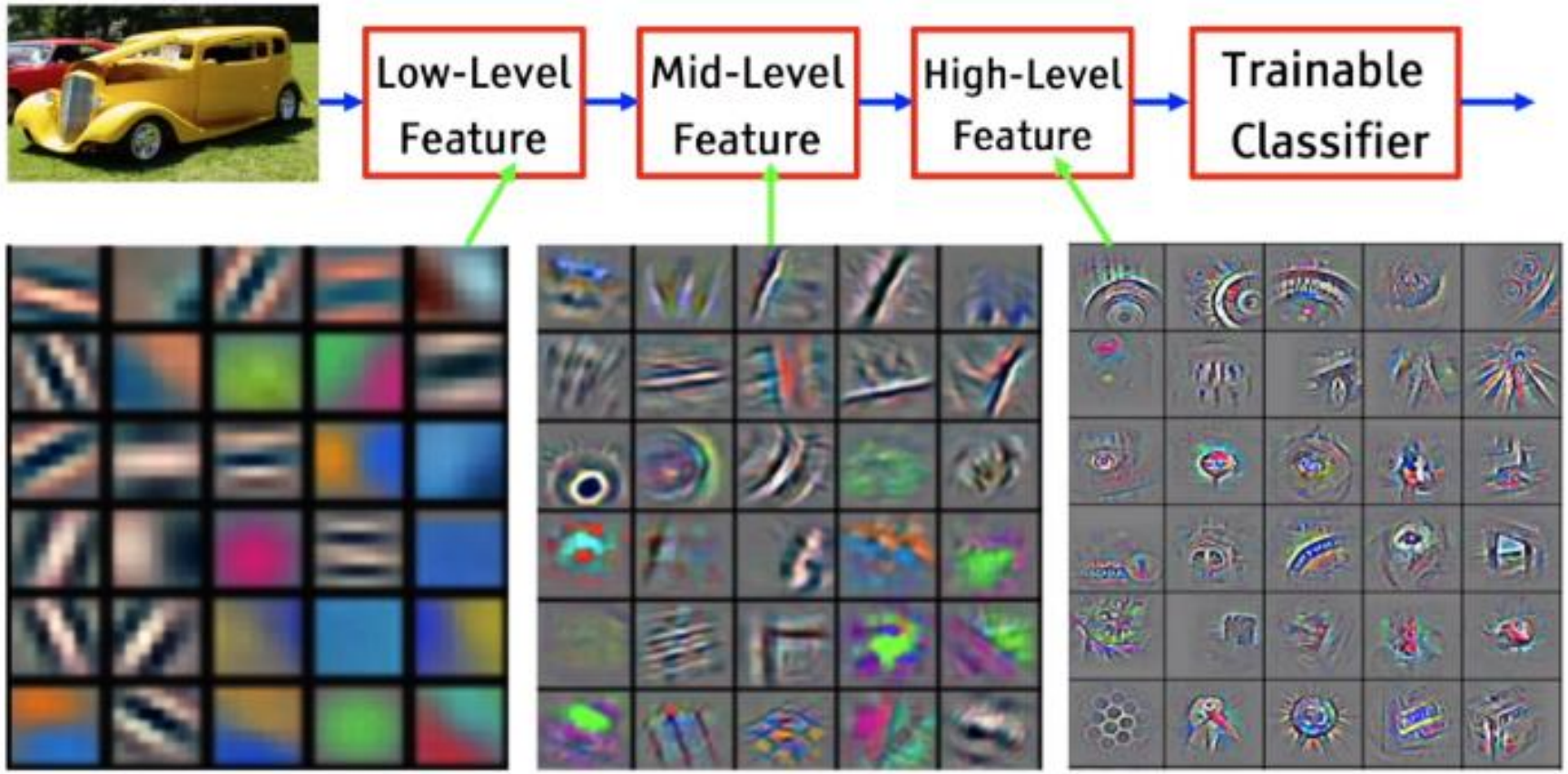
Deep Learning

Complex Functions and Richer Features



Deep Learnt Features (2013-XXX)

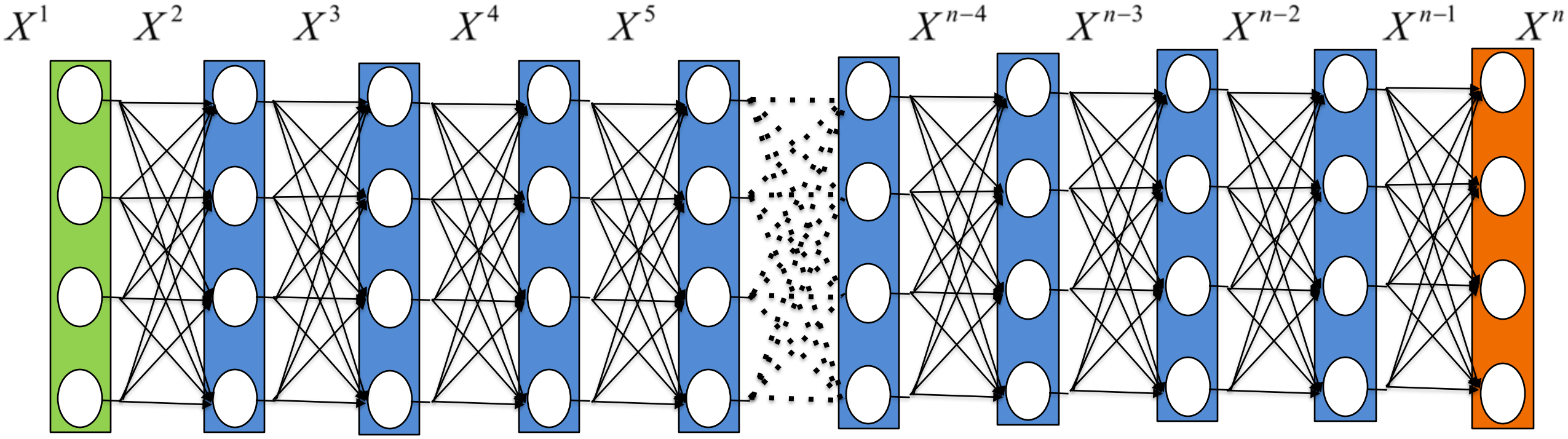
- It's deep if it has more than one stage of non-linear feature transformation.



Source: Yann LeCun

Deep Features

Cutting a Trained Neural Network



Input

Hidden layers

Output

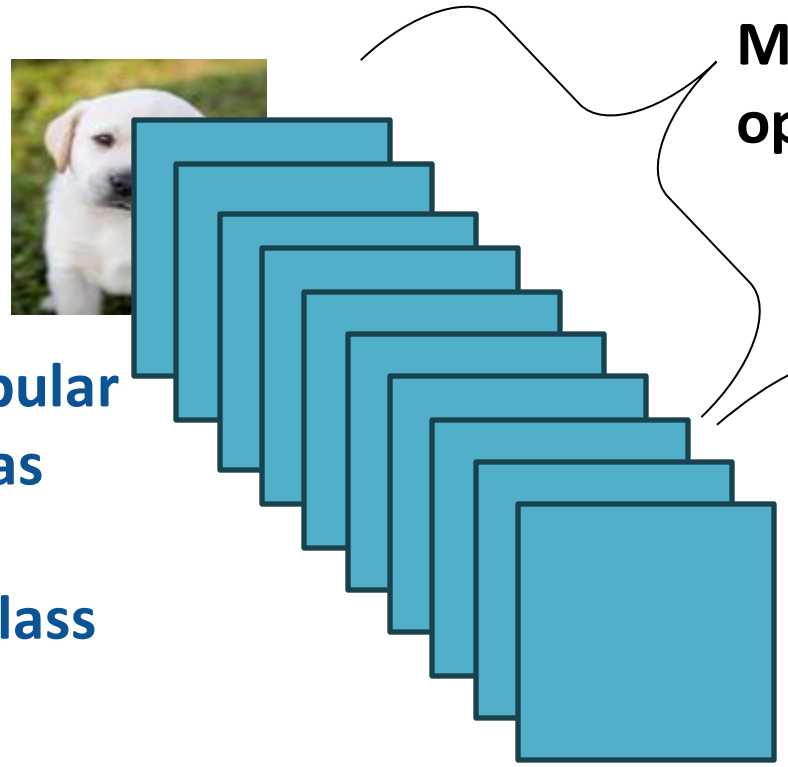
Cut after 2nd Hidden Layer

Cut after (n-2)th Hidden Layer

MLP Perspective

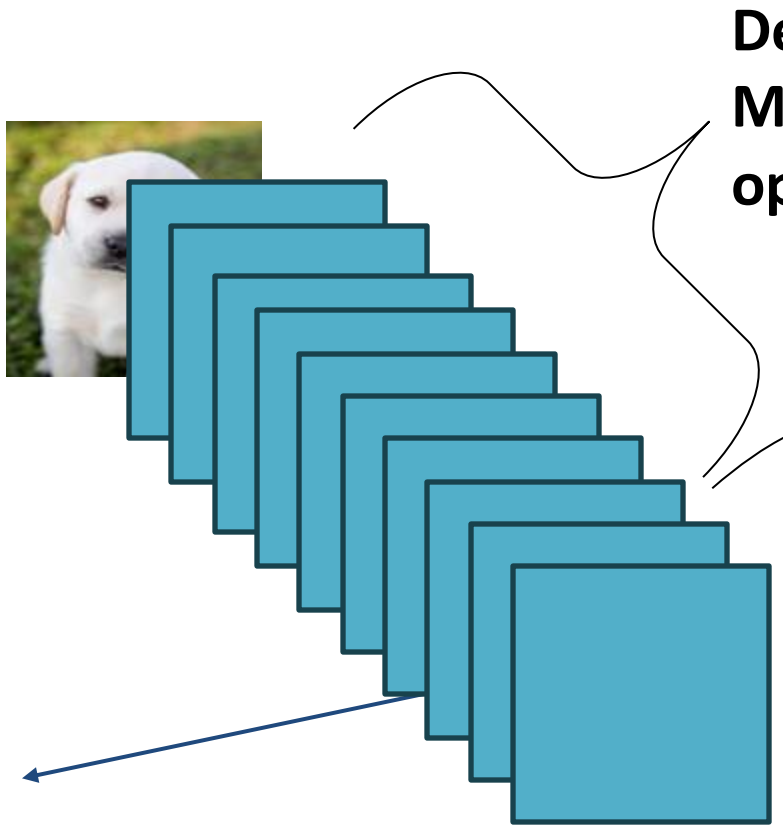
Deep Image Features

- Deep Learning = End to End Learning (Raw data to labels)
- Deep Learning = Feature Learning!!

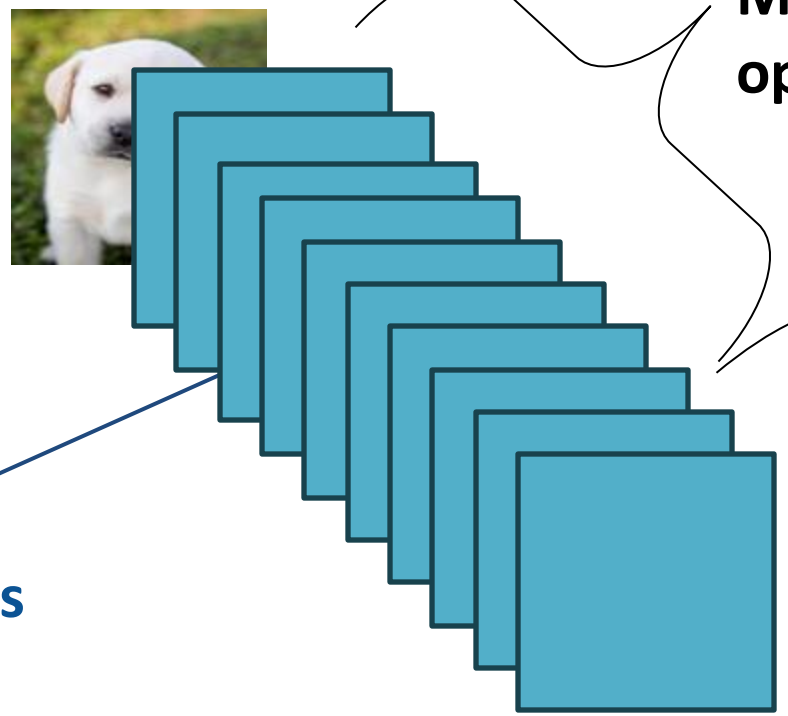


An intermediate representation from a popular "Deep VGGNet", which was designed and trained for solving a "general" 1000 class classification.

Deep Image Features



Deep Image Features

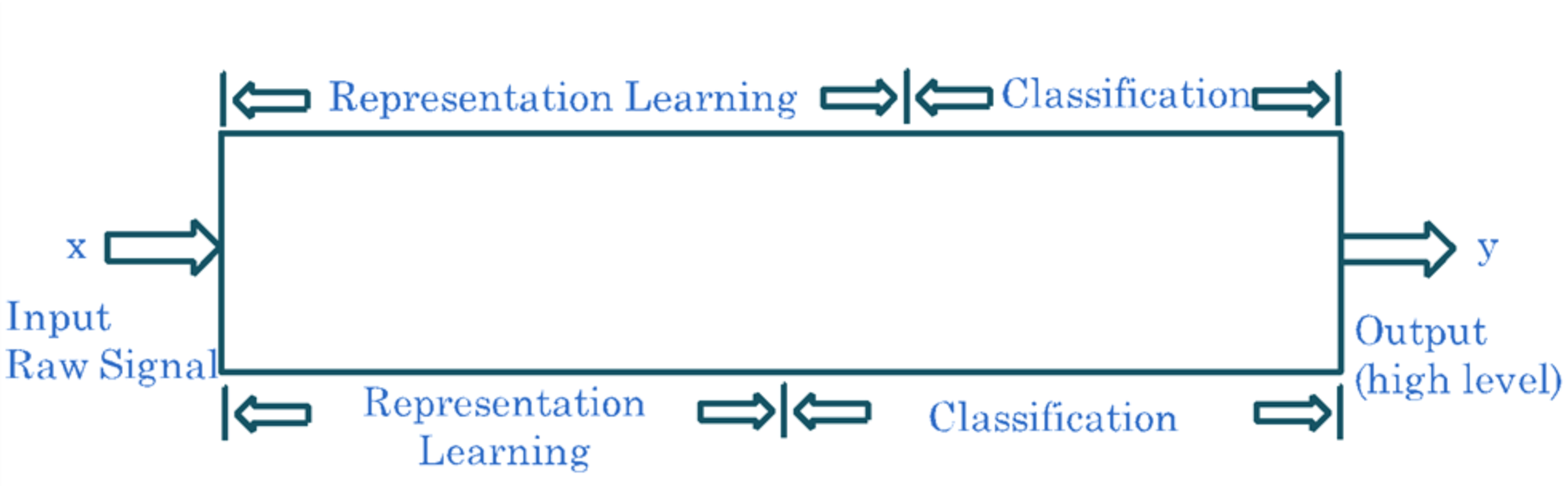


Deep Neural Networks
Many linear and nonlinear operations

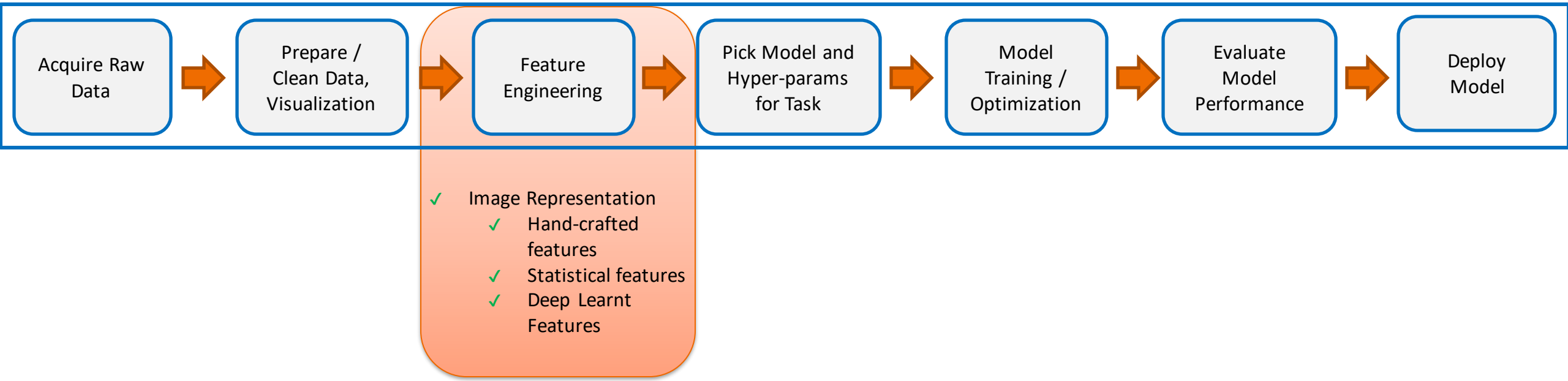
Final
Stages
Classifier

Lower semantics
information in features

Summary


Representation Learning
Classification
x
Input
Raw Signal
y
Output
(high level)
Representation
Learning
Classification

Summary


Acquire Raw Data
Prepare / Clean Data, Visualization
Feature Engineering
Pick Model and Hyper-params for Task
Model Training / Optimization
Evaluate Model Performance
Deploy Model
✓ Image Representation
✓ Hand-crafted features
✓ Statistical features
✓ Deep Learnt Features

Thanks!!!

Questions?