Deep Learning

Supplementary slides

Week 1

Machine Learning

Input: X

Output: Y

Label”motorcycle”

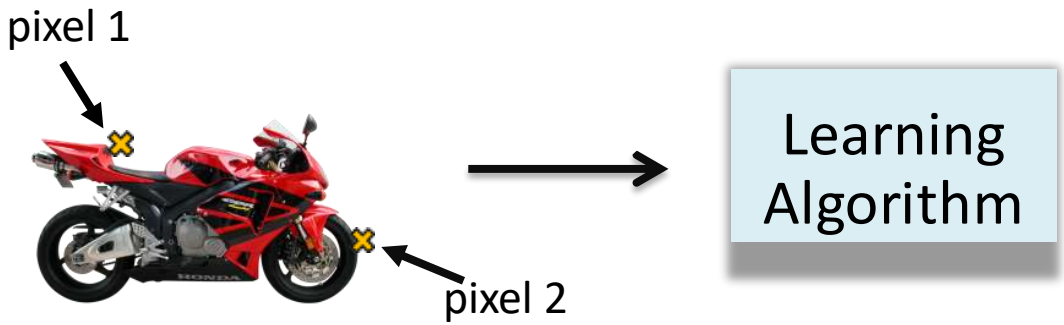
Why is it hard?

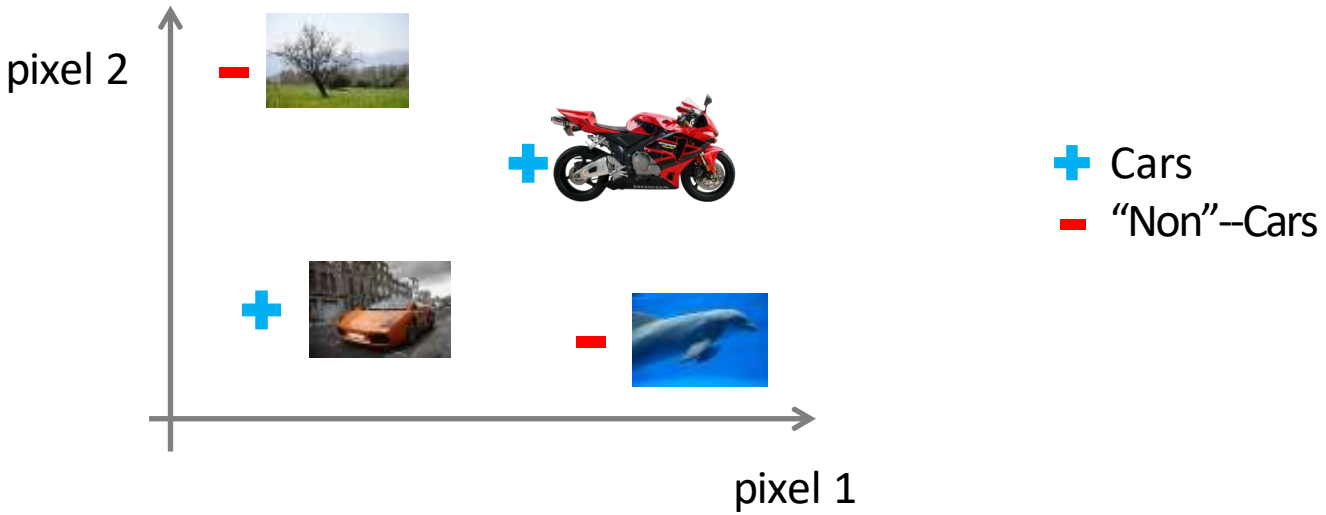
You see this

But the camera sees this:

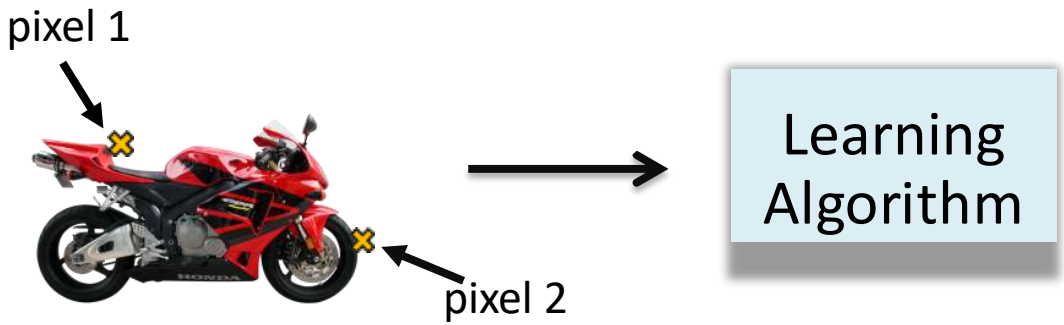
194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	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

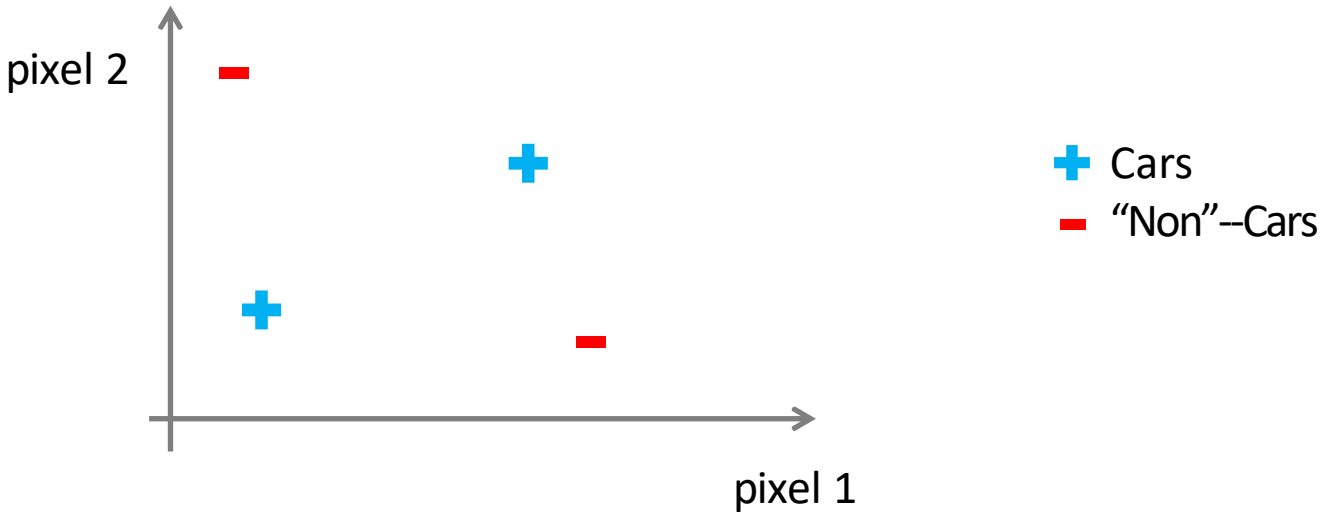
Raw Image Representation



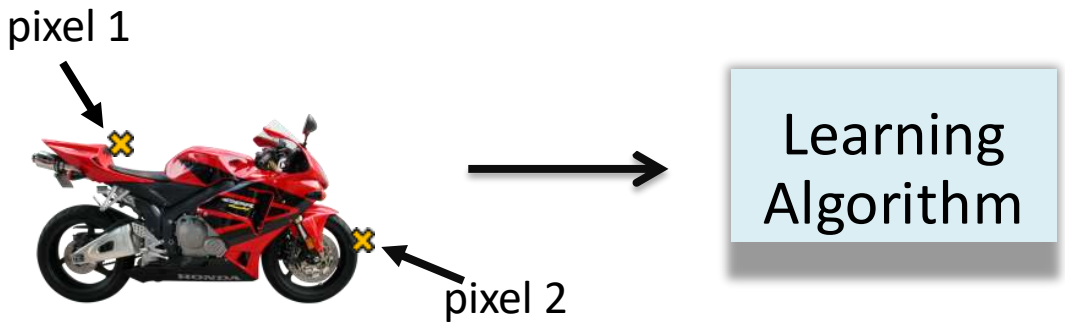


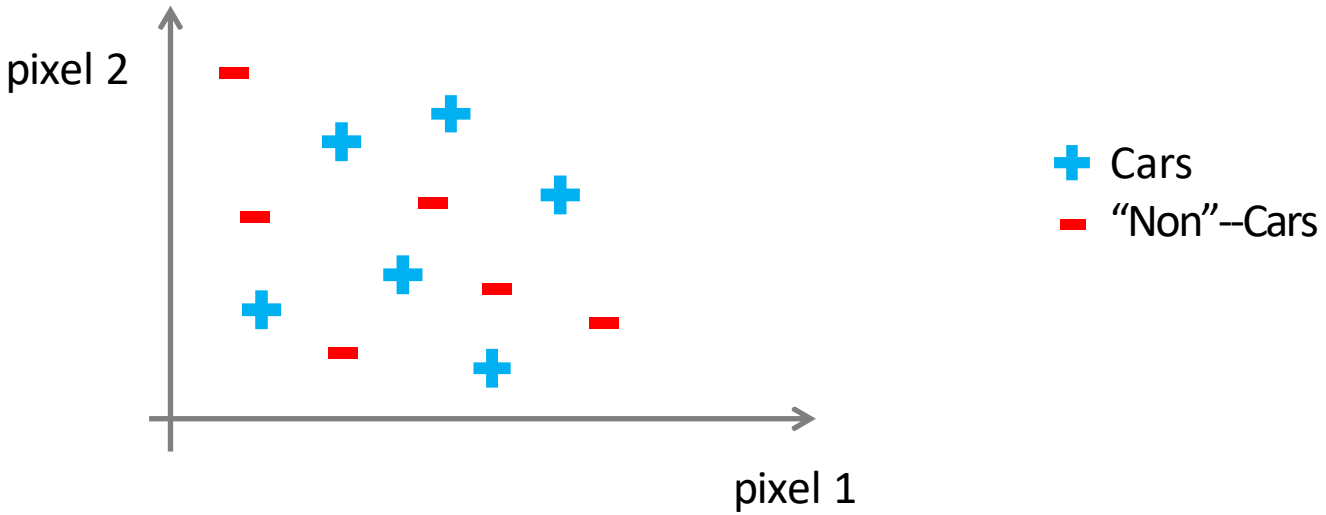
Raw Image Representation



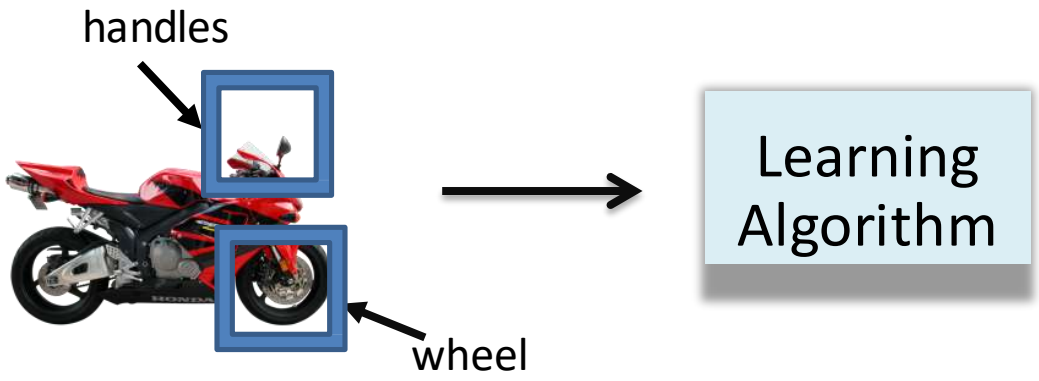


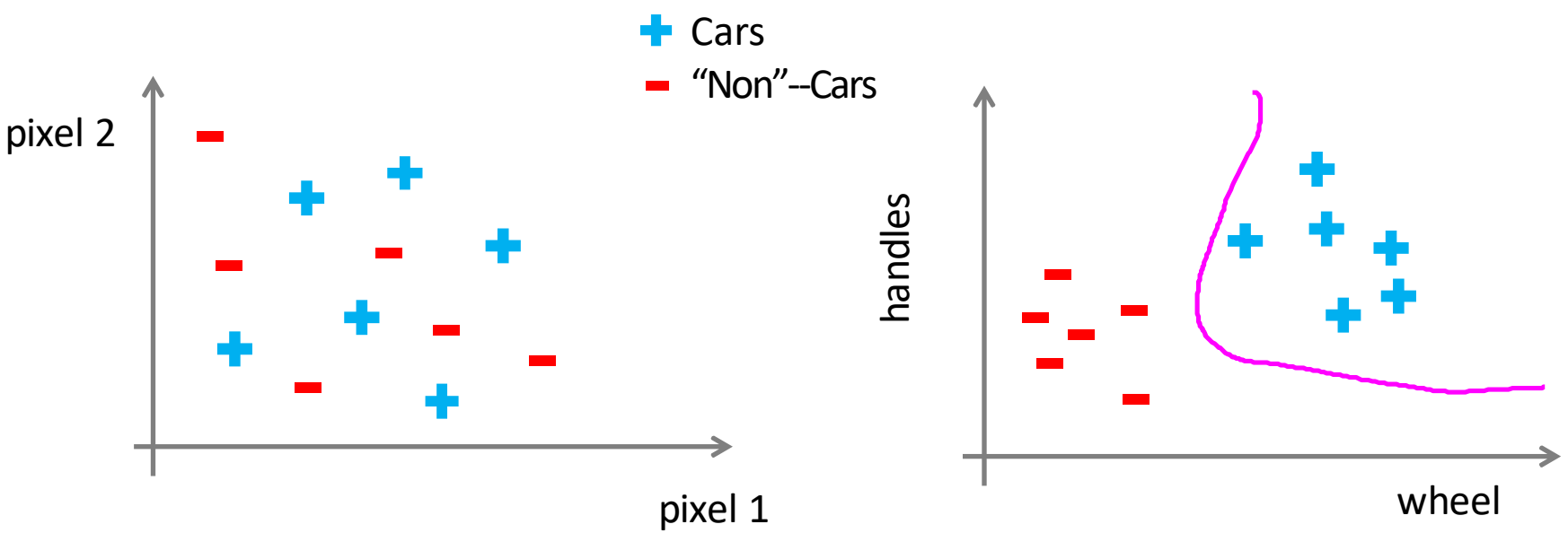
Raw Image Representation



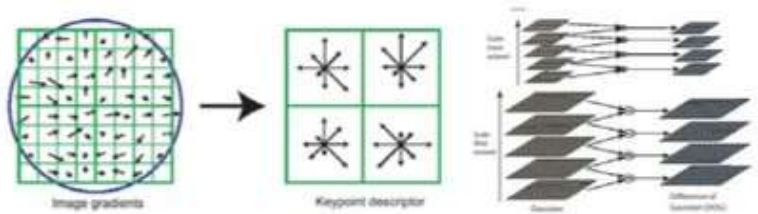


Better Feature Representation?

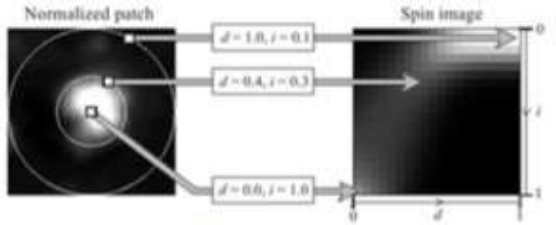




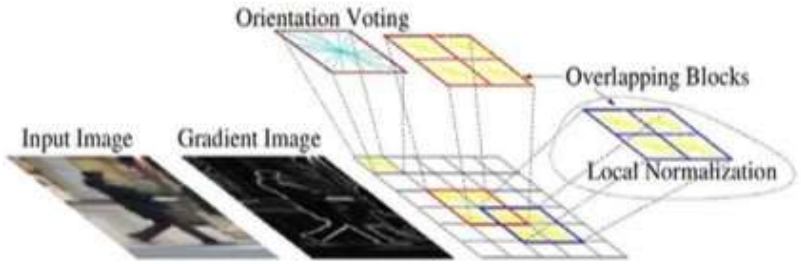
Feature Representations



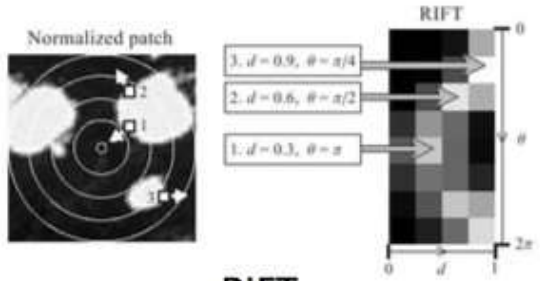
SIFT



Spin image

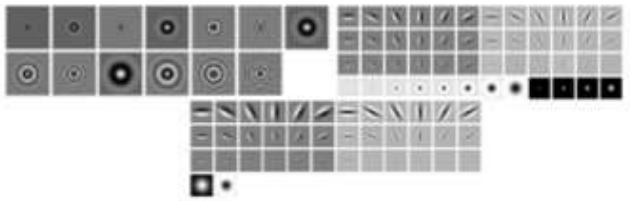


HoG

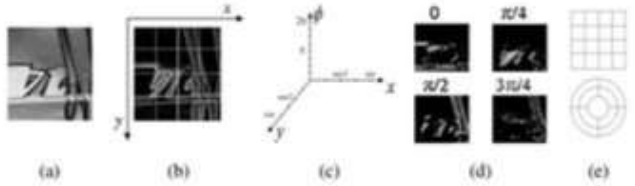


RIFT

Expert Knowledge!

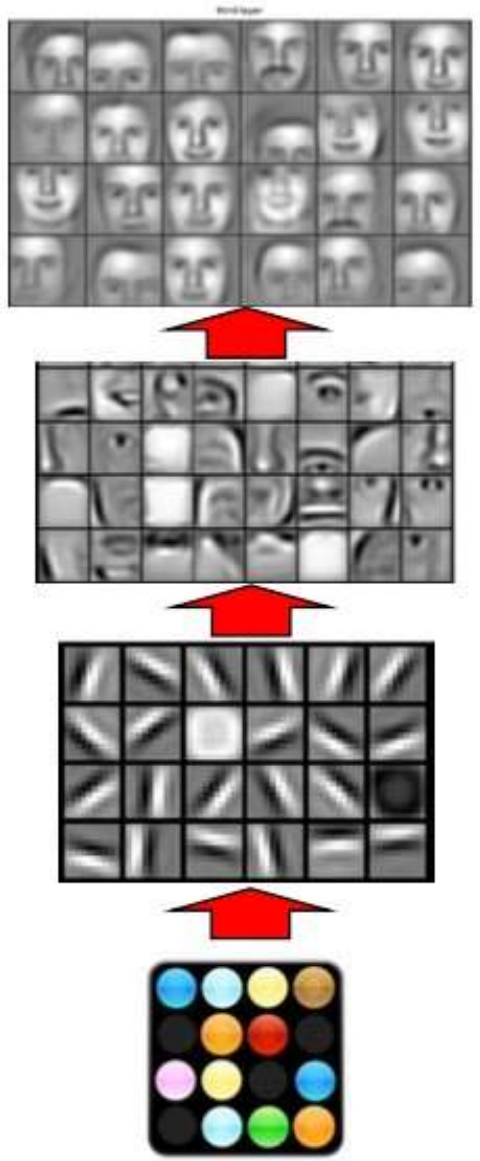


Textons



GLOH

Deep Learning: learn representations!



object models

object parts
(combination
of edges)

edges

pixels

What exactly is deep learning?

2. **why is it generally better** than other methods on image, speech and certain other types of data?

The short answers

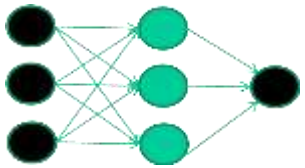
- **Deep Learning' means** using a **neural network** with **several layers of nodes** between input and output

- **The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.**

hmmm… OK, but:

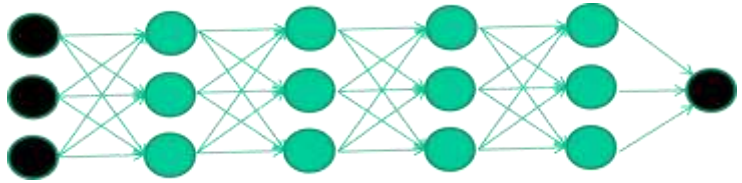
3. **multilayer neural networks have been around for 25 years. What's actually new?**

we have always had good algorithms for learning the weights in networks with 1 hidden layer

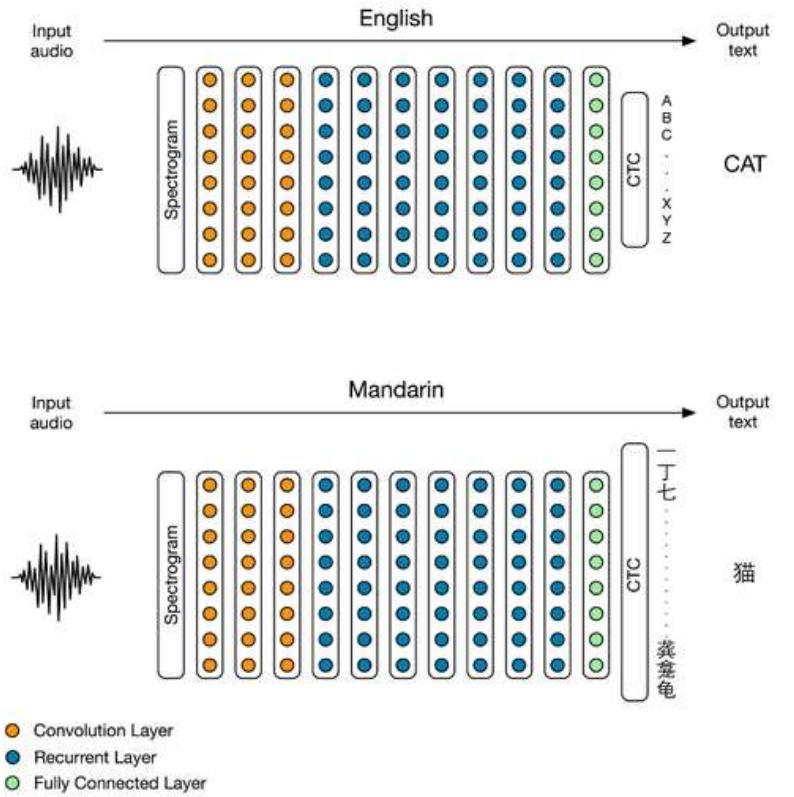


but these algorithms are not good at learning the weights for networks with more hidden layers

what's new is: **algorithms for training many-later networks**



DL Today: Speech to Text

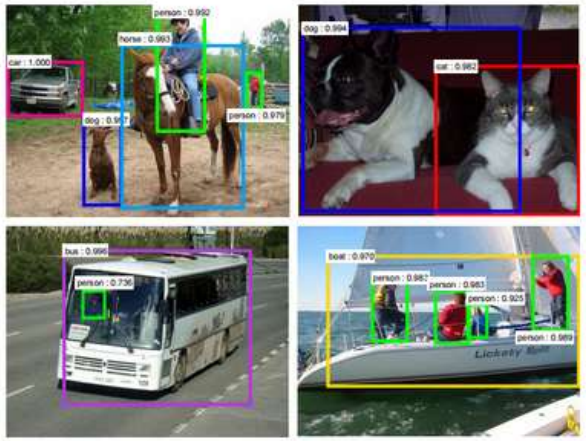


[Baidu 2014]

DL Today: Vision

[Krizhevsky 2012]

[Ciresan et al. 2013]

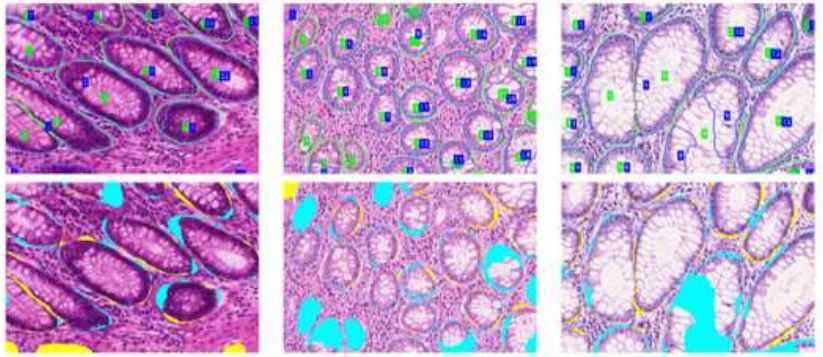


[Faster R-CNN - Ren 2015]

[NVIDIA dev blog]

DL Today: Vision

[Stanford 2017]



[Nvidia Dev Blog 2017]

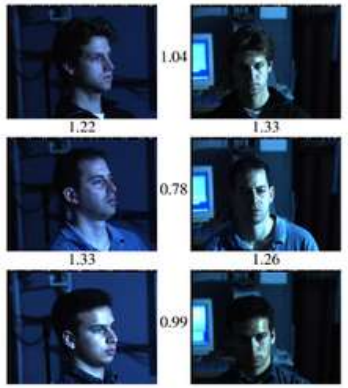
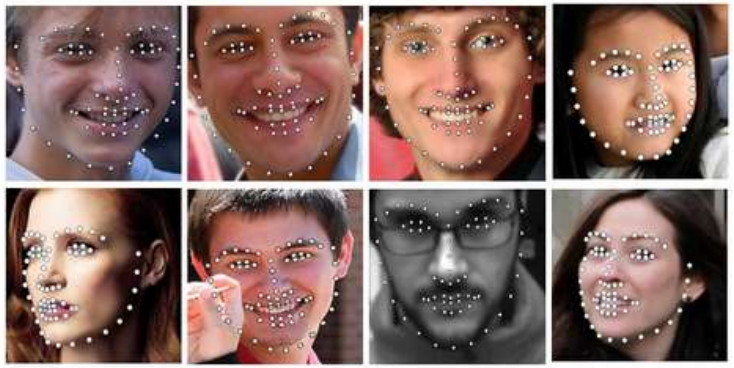


Figure 1. **Illumination and Pose invariance.**

[FaceNet - Google 2015]



[Facial landmark detection CUHK 2014]

DL Today: NLP

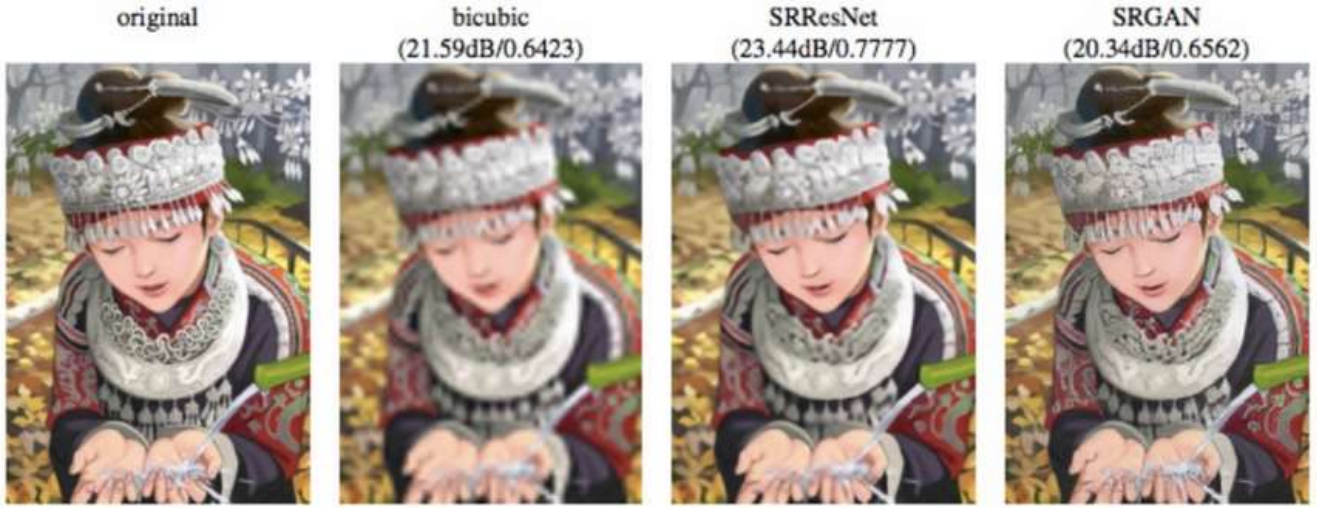
[Google Inbox Smart Reply]

[Amazon Echo / Alexa]

DL Today: Image Translation

[DeepDream 2015]

[Gatys 2015]



[Ledig 2016]

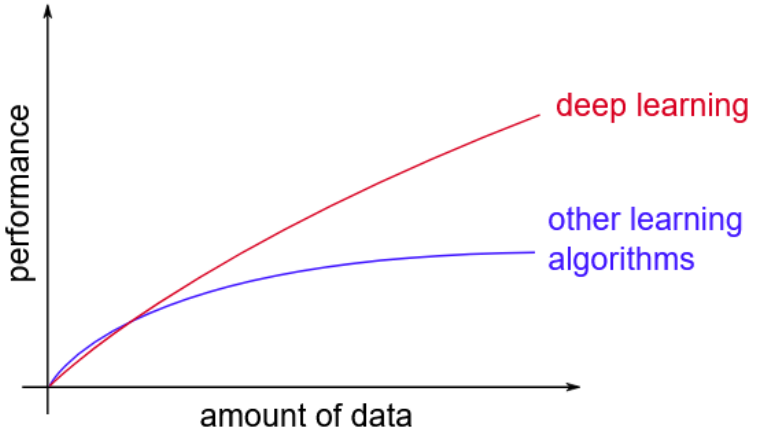
DL Today: Generative Models

Sampled celebrities [Nvidia 2017]

StackGAN v2 [Zhang 2017]

Why Deep Learning Now

- Better algorithms and Understanding
- Computing Power(GPUs, TPUs)
- Data with labels
- Tools and models



PYTORCH

Microsoft CNTK

Caffe2

dmlc mxnet

gensim

spaCy

theano