Introduction of Deep Learning using Deep Neural Network

Ruijiang Luo, June 2017

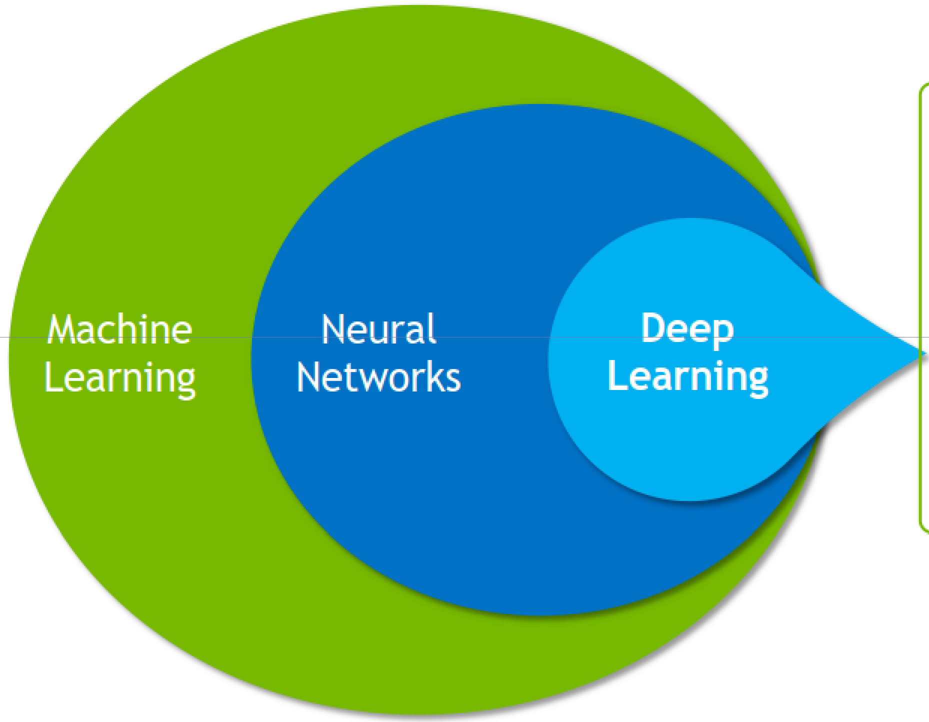
DESAY SV automotive

Confidential

Content

- Overview
- Machine Learning
- Neural Network
  - Shallow Architecture
  - Deep Architecture
  - Convolutional Neural Network (CNN)
- Big Data and Deep Neural Network
- DNN (CNN) for ADAS
- Demo

What is Deep Learning



Modern reincarnation of Artificial Neural Networks (concept originated around 1950s). Synonymous to term “Deep Neural Networks” and an advanced form of Machine Learning.

Self-learn new concepts & Powerful adaptive learning. Performance increases significantly as amount of training data increase. E.g. Google Brain able to form its own concept on what constitutes a cat without being explicitly programmed.

* NVIDIA, “Getting Started with Deep Learning”, 13 Mar. 2017

Machine Learning

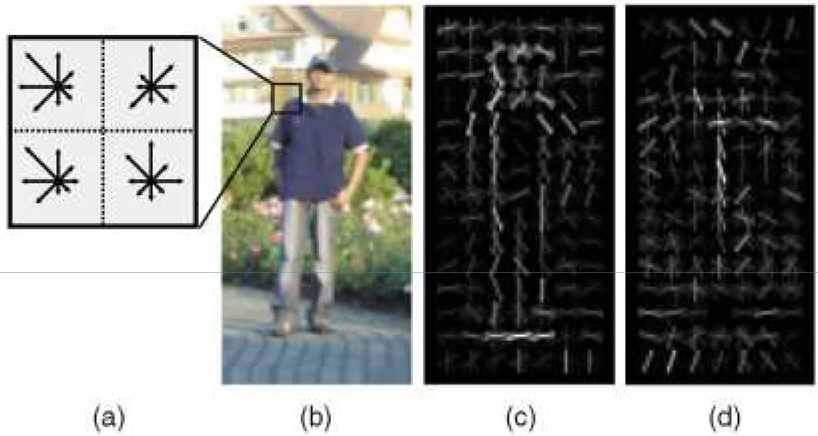


Figure 1. Histogram of Oriented Gradients

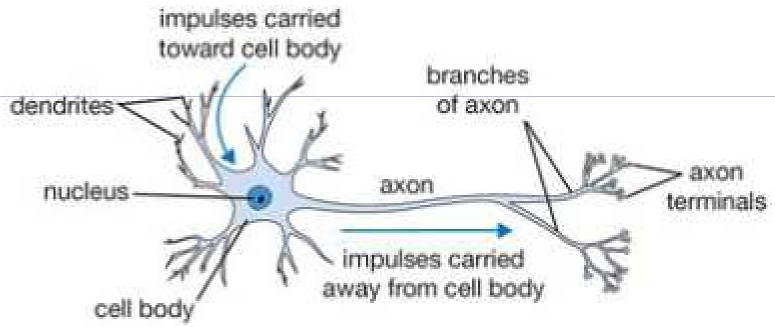
* Sarthak Ahuja, Prateekshit Pandey, “Pedestrian detection using HoG features

Content

- ❑ Overview
- ❑ Machine Learning
- ❑ Neural Network
  - ❑ Shallow Architecture
  - ❑ Deep Architecture
  - ❑ Convolutional Neural Network (CNN)
- ❑ Big Data and Deep Neural Network
- ❑ DNN (CNN) for ADAS
- ❑ Demo

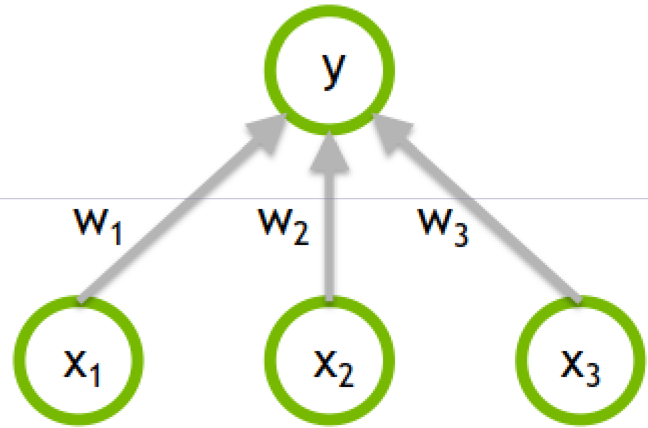
Neural Networks

Biological neuron



From Stanford cs231n lecture notes

Artificial neuron



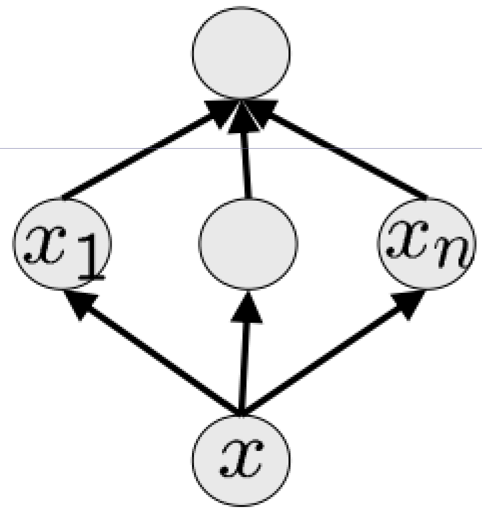
$$y=F(w_1x_1+w_2x_2+w_3x_3)$$

* NVIDIA, “Getting Started with Deep Learning”, 13 Mar. 2017

Shallow Neural Network – one hidden layer

- Most machine learning algorithm are kind of shallow NN*
  - SVM, kNN, Mog, KDE, Parzen Kernel regression, PCA, Perceptron

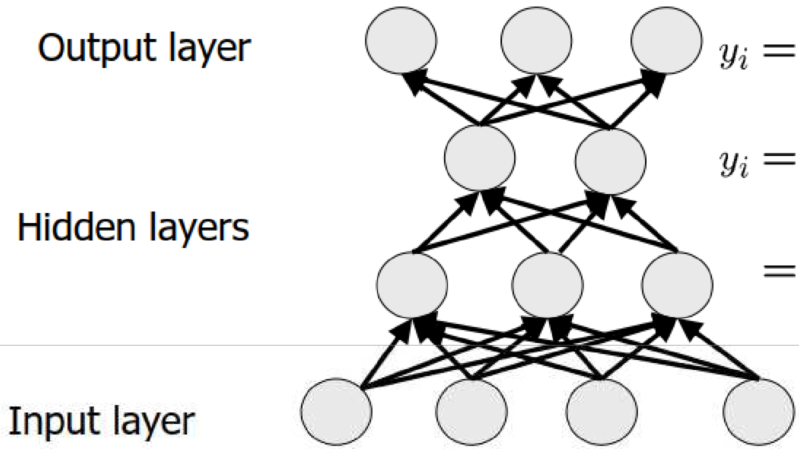
$$\text{SVM: } \hat{f}(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x})\right)$$



* Introduction to Machine Learning, CMU-10701, Deep Learning

Deeper Neural Networks – Many Hidden Layers

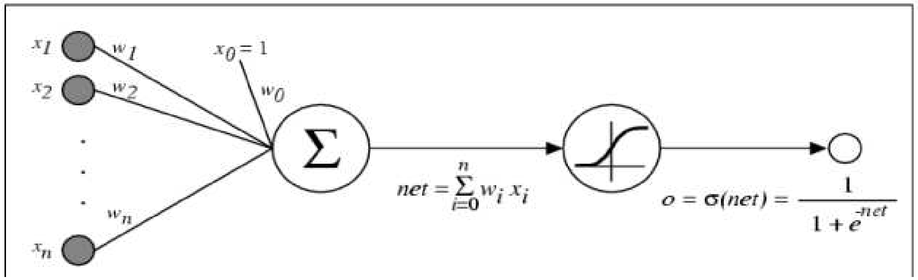
DESAY SV automotive



$$y_i = \tanh\left(\sum_j W_{ij} x_j + W_{i0}\right)$$

$$y_i = \sigma\left(\sum_j W_{ij} x_j + W_{i0}\right)$$

$$= \frac{1}{1 + \exp[-(\sum_j W_{ij} x_j + W_{i0})]}$$



4

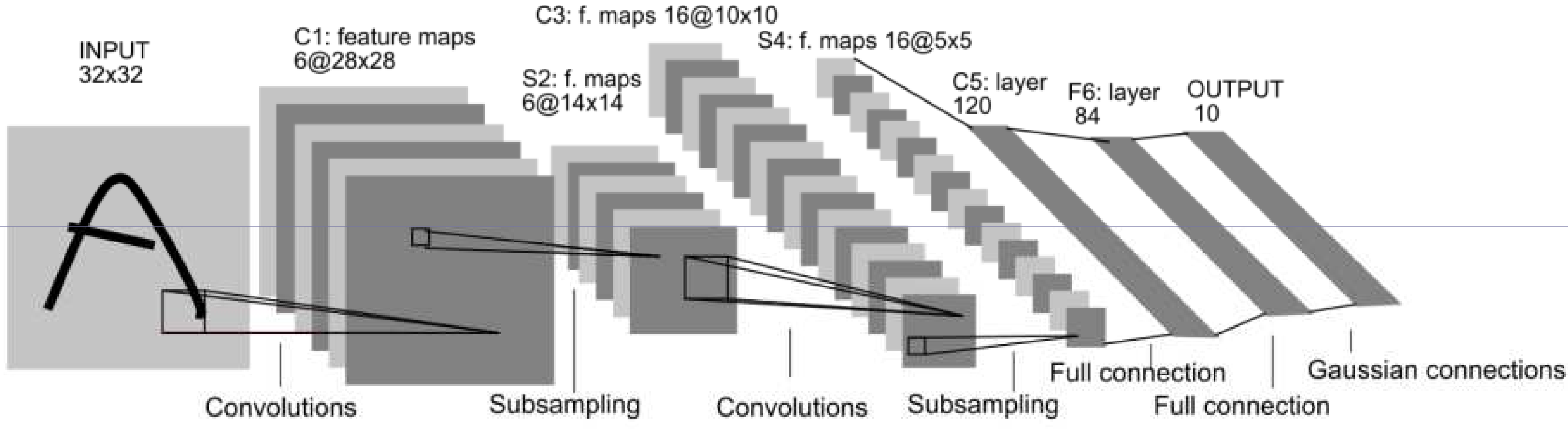
* Introduction to Machine Learning, CMU-10701, Deep Learning

Digression on Neural Networks

- Researchers reported positive experimental results with typically one or two hidden layers
- Training deeper networks consistently yielded poorer results.
- After Vapnik and his co-workers developed the Support Vector Machine (SVM) in 1993, many researchers abandoned neural networks with multiple adaptive hidden layers because SVMs worked better
- Exception: 7-layer Convolutional Neural Networks, CNN, LeCun Yann, 1998

* Introduction to Machine Learning, CMU-10701, Deep Learning

Convolutional Neural Network – LeNet



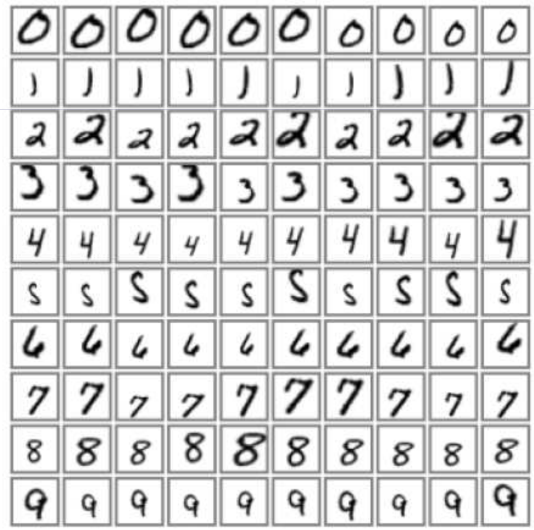
* Y. LeCun, L.Bottou, Y.Bengio and P.Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 86(11):2278-2324, November 1998

Convolutional Neural Network – cont.

- Very successful in MINIST Dataset on hand-writing digits recognition

60,000 original datasets

Test error: 0.95%



540,000 artificial distortions

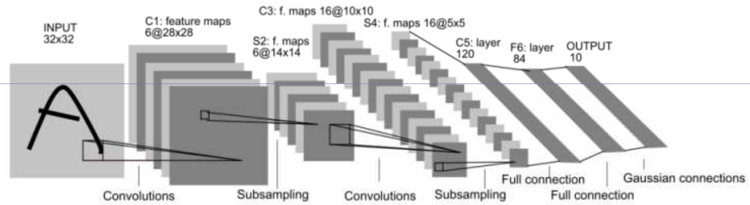
+ 60,000 original

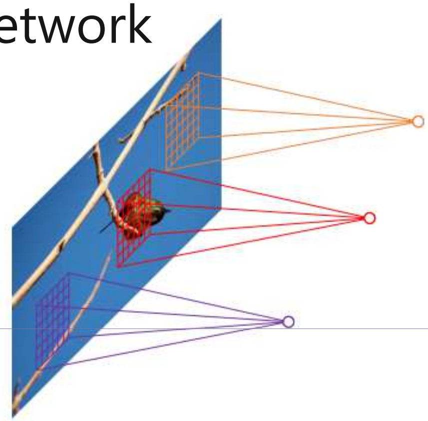
Test error: 0.8%

* Introduction to Machine Learning, CMU-10701, Deep Learning

Convolutional Neural Network – cont.

- Compared to standard NN, Convolutional Neural Network
  - Have much fewer connections and parameters
  - Easier to train
  - Shift invariant, rotation invariant, and noise resistant





- However, when used on image recognition, CNN's performance dropped significantly
  - As a result, CNN did not gain much attention, until big-data (huge amount of images) became available

Content

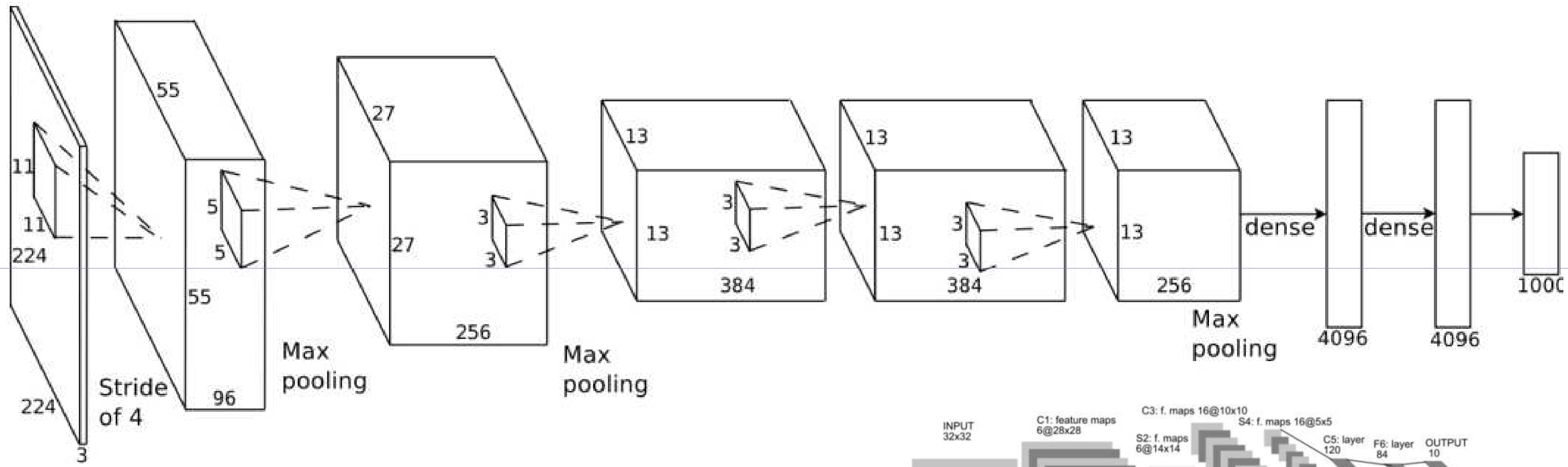
- ❑ Overview
- ❑ Machine Learning
- ❑ Neural Network
  - ❑ Shallow Architecture
  - ❑ Deep Architecture
  - ❑ Convolutional Neural Network (CNN)
- ❑ Big Data and Deep Neural Network
- ❑ DNN (CNN) for ADAS
- ❑ Demo

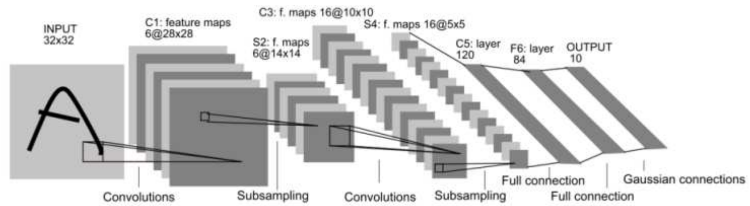
Big Data – ImageNet by Stanford University

- First ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2010)
  - 1K categories
  - 1.2M training image (~1000 per category)
  - 50,000 validation images
  - 150,000 testing images

* Introduction to Machine Learning, CMU-10701, Deep Learning

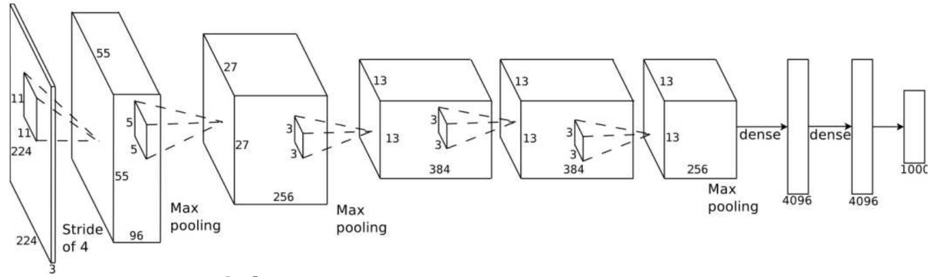
Deep Convolutional Neural Network Architecture





* Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton, "ImageNet Classification with Deep Convolutional Neural Networks" , NIPS 2012

Deep Convolutional Neural Network Facts



- 650,000 neurons
- 60,000,000 parameters
- 630,000,000 connections
- 5 convolutional layers, 3 fully connected layers
- Final feature layer: 4096 dimensional
- Output: 1000 categories
- Trained:
  - With stochastic gradient descent
  - On two NVIDIA GTX 580 3GB GPUs
  - For about a week

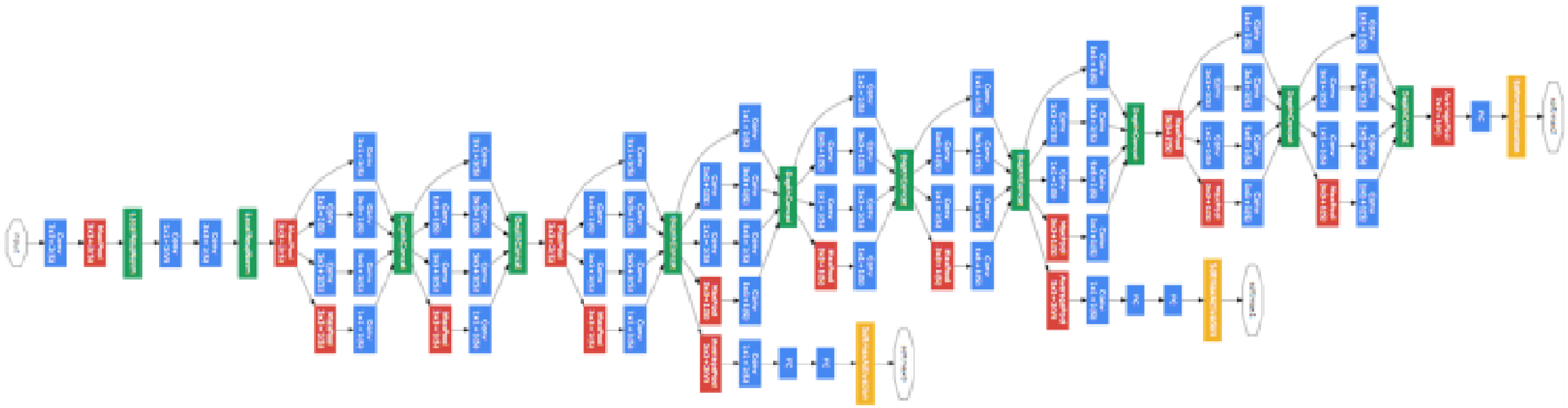
* Introduction to Machine Learning, CMU-10701, Deep Learning

ImageNet – on going

- 15M images, collected from Web
- 22,000 categories
- Human labelers (Amazon's Mechanical Turk crowd-sourcing)

* Introduction to Machine Learning, CMU-10701, Deep Learning

Deeper and Deeper CNNs



* GoogLeNet, VGG16, winner of ILSVRC2014

Deeper and Deeper CNNs – cont.

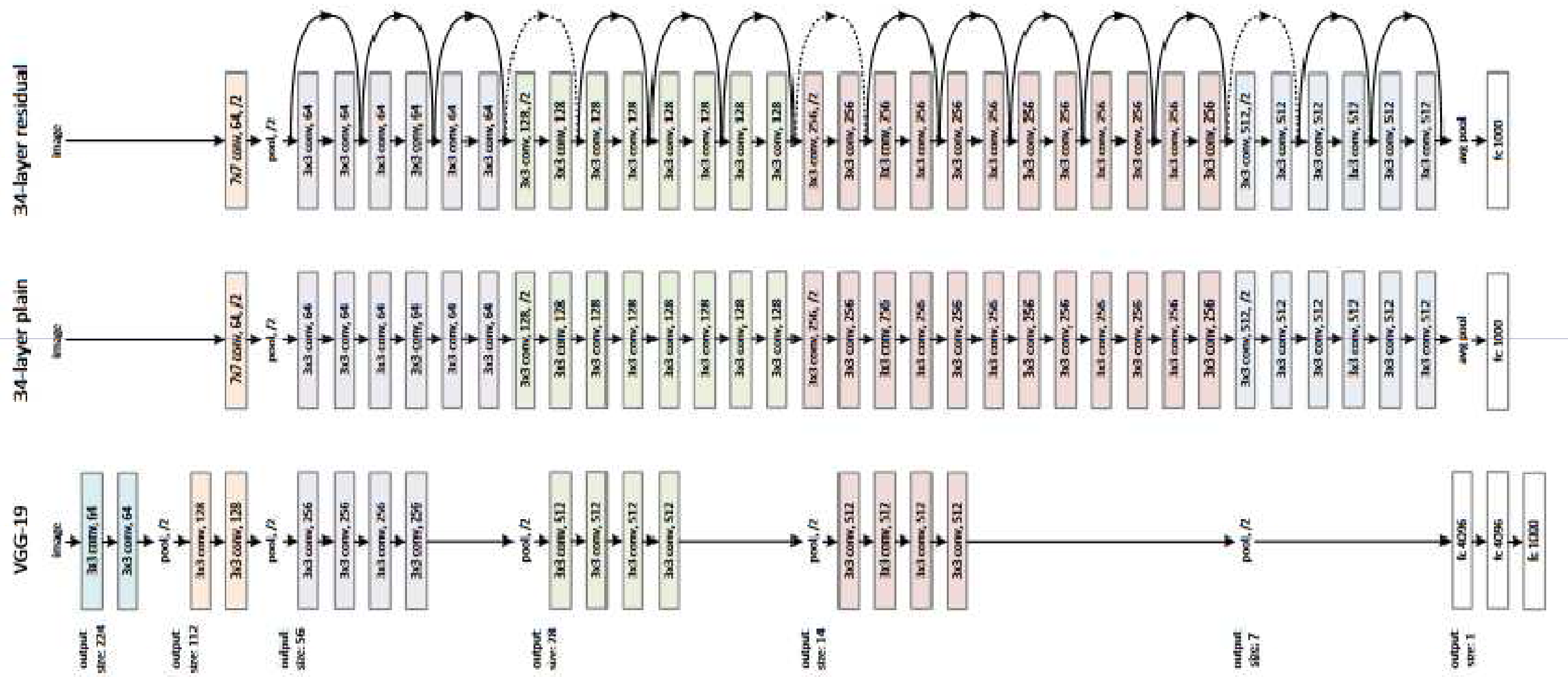
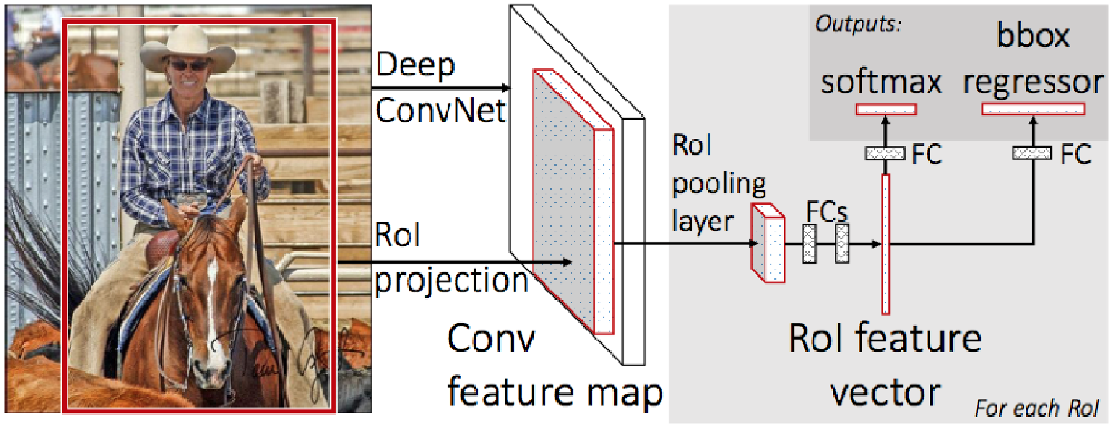


Figure 3. Example network architectures for ImageNet. **Left**: the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle**: a plain network with 34 parameter layers (3.6 billion FLOPs). **Right**: a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

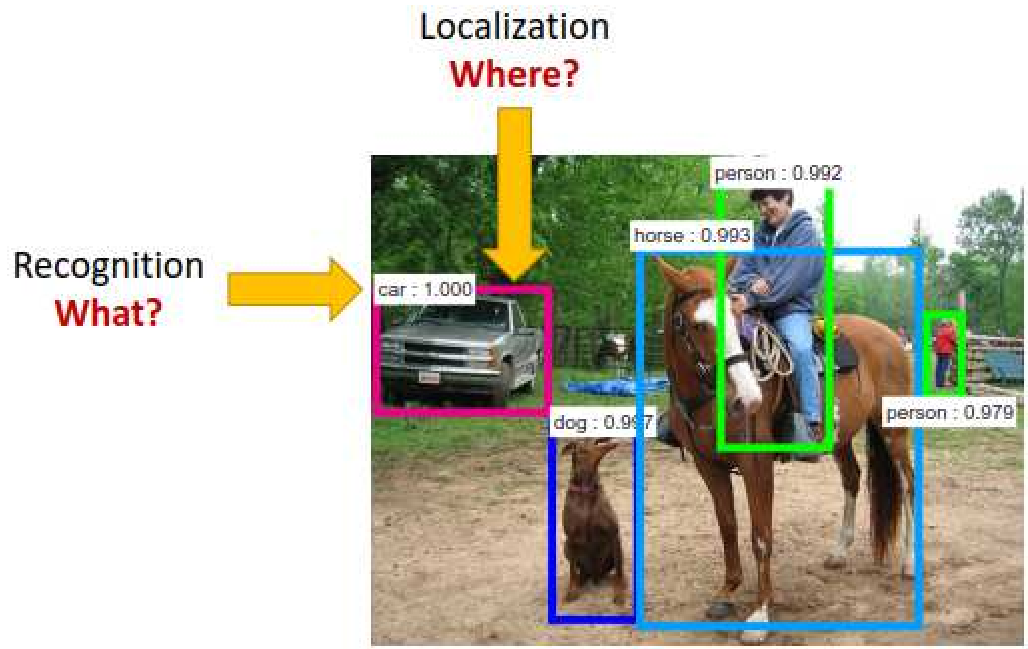
* Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", Microsoft Research, arXiv:1512.03385v1 [cs.CV] 10 Dec 2015

R-CNN for Object Detection



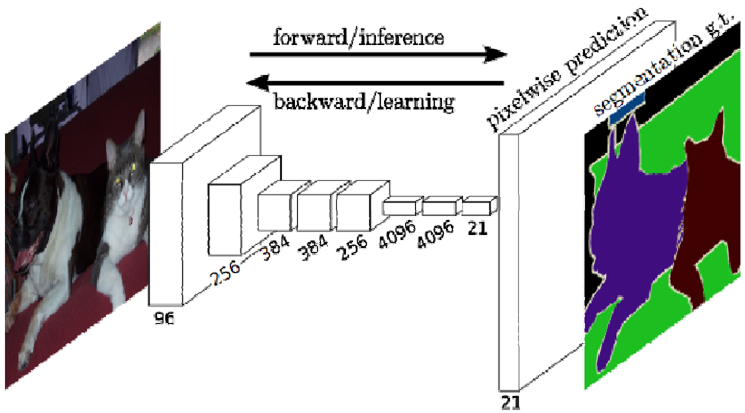
* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, "DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe", Univ. of Berkeley, BerkeleyVision

R-CNN for Object Detection



* Keiming He, “Convolutional Feature Maps”, Microsoft Research Asia (MSRA)

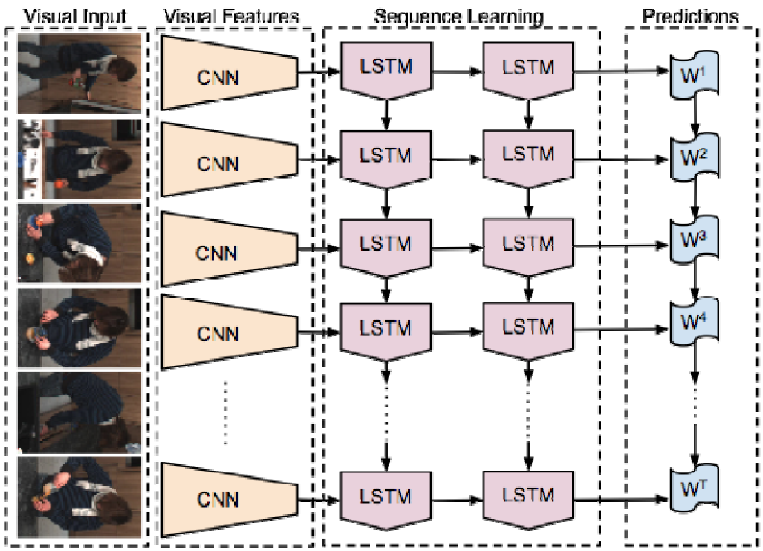
FCN for Pixelwise Prediction



* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, “DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe”, Univ. of Berkeley, BerkeleyVision

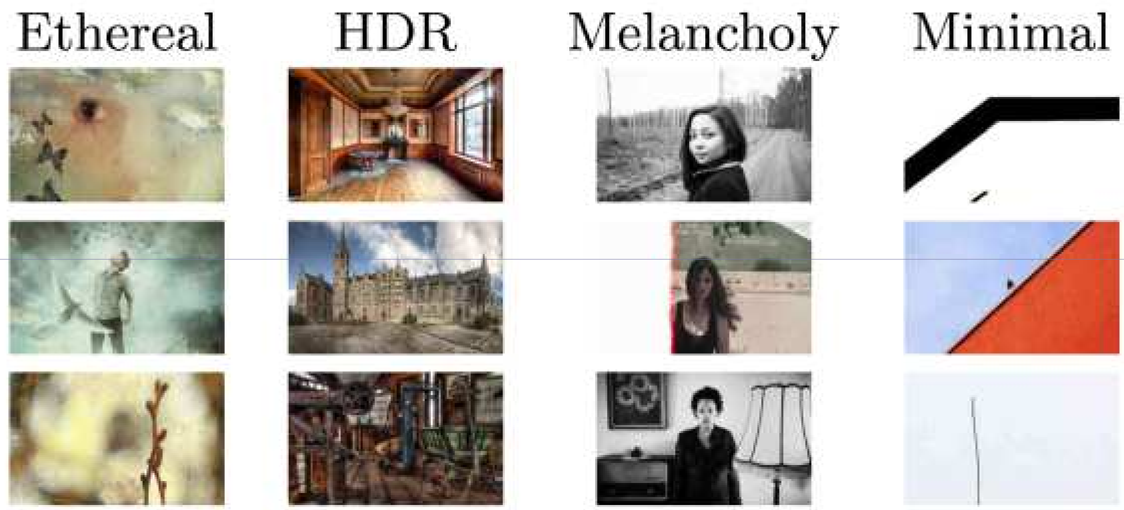
Recurrent Networks for Sequences

- LRCN: Long-term Recurrent Convolutional Network



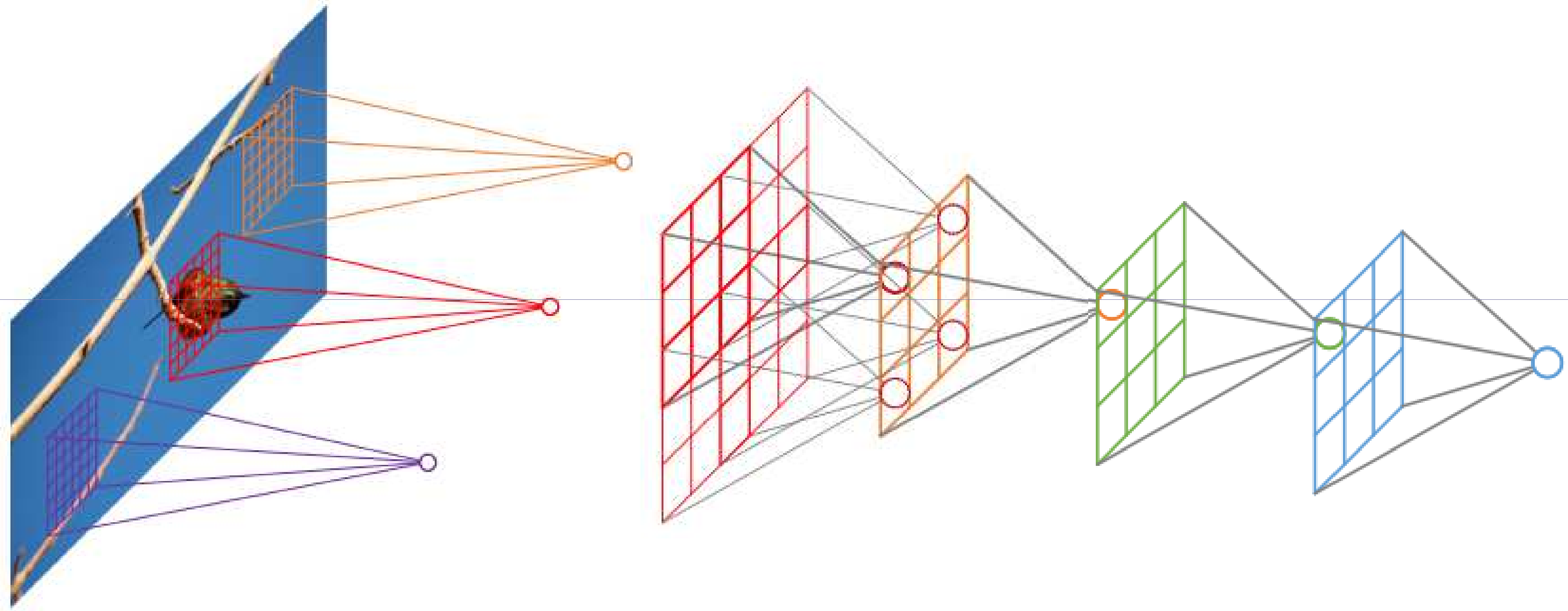
* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, "DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe", Univ. of Berkeley, BerkeleyVision

And Even Visual Style Recognition



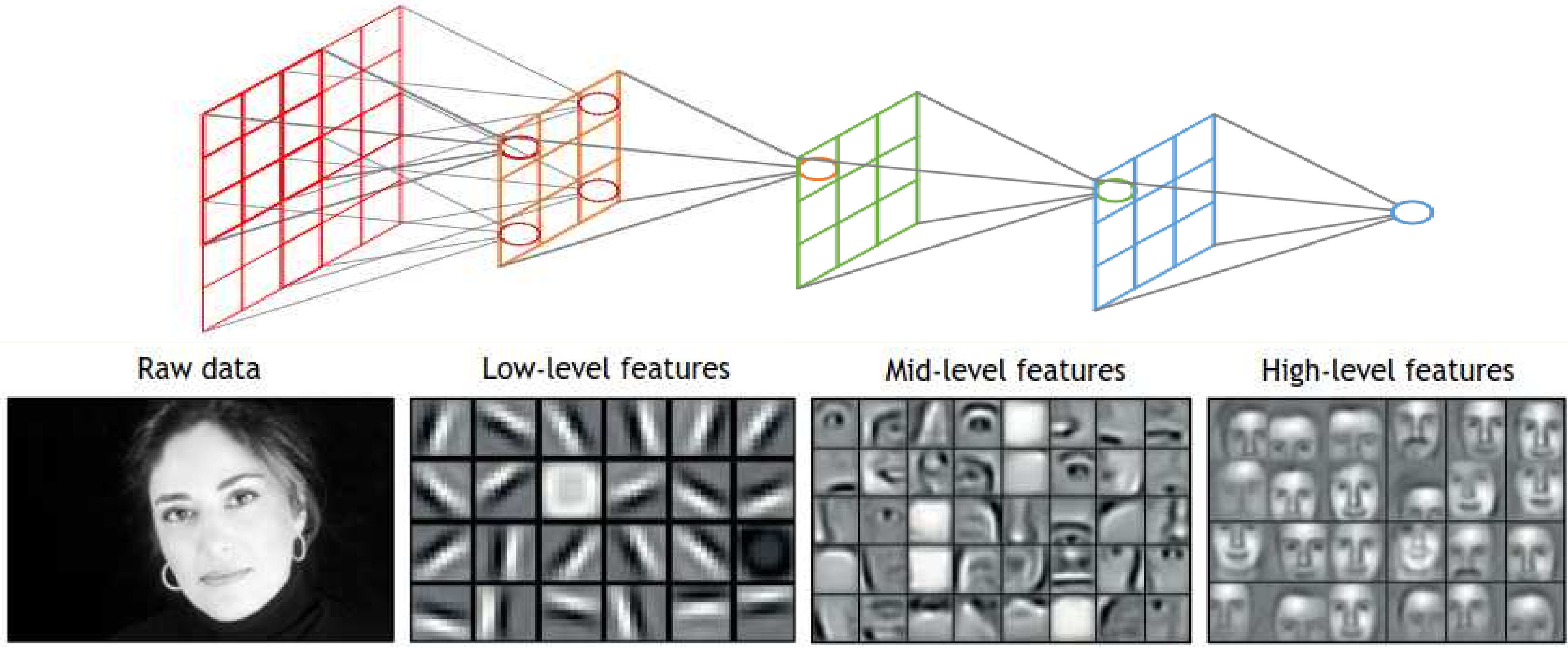
* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, “DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe”, Univ. of Berkeley, BerkeleyVision

Receptive Field of Convolutional Layers



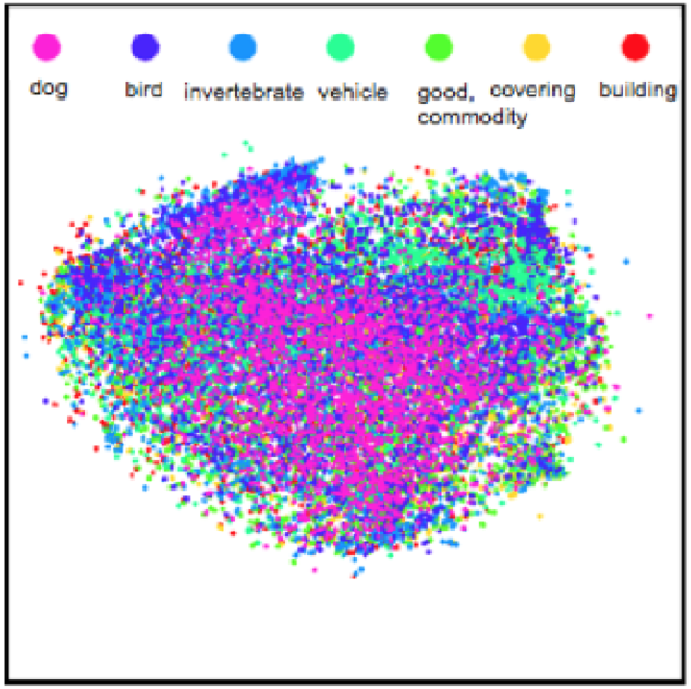
* Keiming He, “Convolutional Feature Maps”, Microsoft Research Asia (MSRA)

Why Deep Learning – Global Feature Extraction

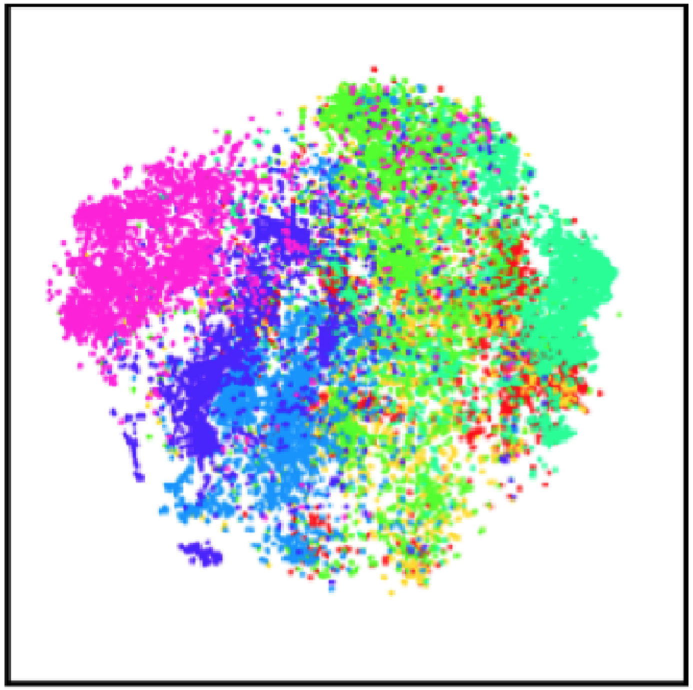


* NVIDIA, “Getting Started with Deep Learning”, 13 Mar. 2017

Why Deep Learning – Better Feature Separation



Low-level: $Pool_1$



High-level: FC_6

* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, “DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe”, Univ. of Berkeley, BerkeleyVision

Content

- ❑ Overview
- ❑ Machine Learning
- ❑ Neural Network
  - ❑ Shallow Architecture
  - ❑ Deep Architecture
  - ❑ Convolutional Neural Network (CNN)
- ❑ Big Data and Deep Neural Network
- ❑ Deep NN (Deep CNN) for ADAS
- ❑ Demo

ADAS: Advanced Driver Asistance Systems

* TI Deep Learning Library And Semantic Segmentation Demo

ADAS using Deep CNN – Pedestrian Detection

DESAY SV automotive

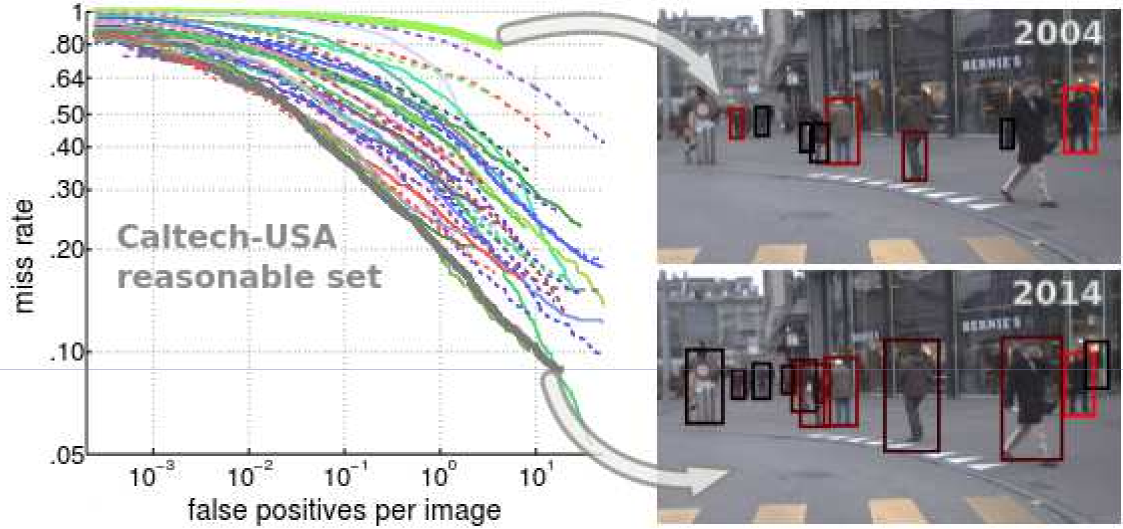


Figure 1: The last decade has shown tremendous progress on pedestrian detection. What have we learned out of the 40+ proposed methods?

* Rodrigo Benenson, Mohamed Omran, Jan Hosang, Bernt Schiele, "Ten Years of Pedestrian Detection, What Have We Learned", arXiv:1411.4304v1 [cs.CV] 16 Nov 2014

ADAS using Deep CNN – More Object Detection

Table 7: Detailed detection results on the PASCAL VOC 2007 test set.

method	data	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Faster R-CNN	07+12	76.4	79.8	80.7	76.2	68.3	55.9	85.1	85.3	89.8	56.7	87.8	69.4	88.3	88.9	80.9	78.4	41.7	78.6	79.8	85.3	72.0
Faster R-CNN+++	07+12+CO	85.6	90.0	89.6	87.8	80.8	76.1	89.9	89.9	89.6	75.5	90.0	80.7	89.6	90.3	89.1	88.7	65.4	88.1	85.6	89.0	86.8
R-FCN	07+12	79.5	82.5	83.7	80.3	69.0	69.2	87.5	88.4	88.4	65.4	87.3	72.1	87.9	88.3	81.3	79.8	54.1	79.6	78.8	87.1	79.5
R-FCN ms train	07+12	80.5	79.9	87.2	81.5	72.0	69.8	86.8	88.5	89.8	67.0	88.1	74.5	89.8	90.6	79.9	81.2	53.7	81.8	81.5	85.9	79.9
R-FCN ms train	07+12+CO	83.6	88.1	88.4	81.5	76.2	73.8	88.7	89.7	89.6	71.1	89.9	76.6	90.0	90.4	88.7	86.6	59.7	87.4	84.1	88.7	82.4

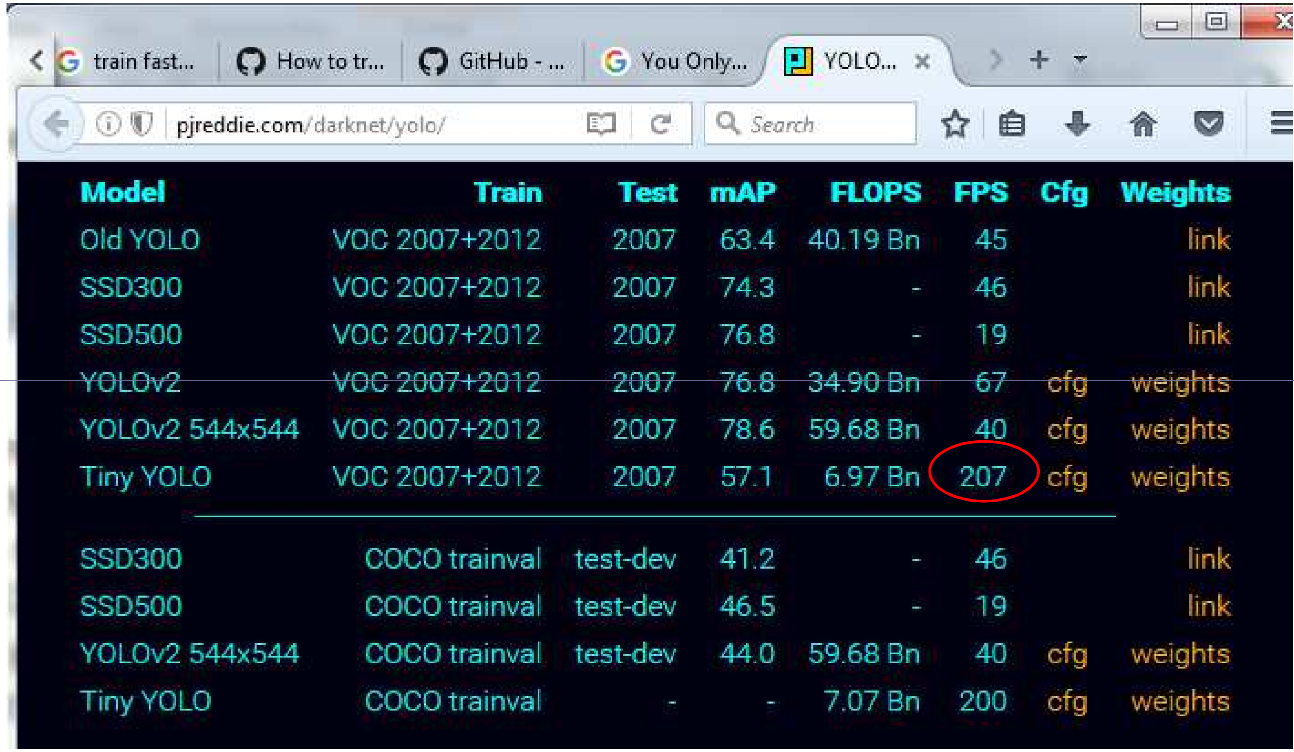
Table 8: Detailed detection results on the PASCAL VOC 2012 test set. †: http://host.robots.ox.ac.uk:8080/anonymous/44L5HI.html ‡: http://host.robots.ox.ac.uk:8080/anonymous/MVCM2L.html

method	data	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Faster R-CNN	07++12	73.8	86.5	81.6	77.2	58.0	51.0	78.6	76.6	93.2	48.6	80.4	59.0	92.1	85.3	84.8	80.7	48.1	77.3	66.5	84.7	65.6
Faster R-CNN+++	07++12+CO	83.8	92.1	88.4	84.8	75.9	71.4	86.3	87.8	94.2	66.8	89.4	69.2	93.9	91.9	90.9	89.6	67.9	88.2	76.8	90.3	80.0
R-FCN ms train†	07++12	77.6	86.9	83.4	81.5	63.8	62.4	81.6	81.1	93.1	58.0	83.8	60.8	92.7	86.0	84.6	84.4	59.0	80.8	68.6	86.1	72.9
R-FCN ms train‡	07++12+CO	82.0	89.5	88.3	83.5	70.8	70.7	85.5	86.3	94.2	64.7	87.6	65.8	92.7	90.5	89.4	87.8	65.6	85.6	74.5	88.9	77.4

- Jifeng Dai, Yi Li, Kaiming He, and Jian Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks”, arXiv: 1605.06409v2 [cs.CV] 21 Jun 2016 (Improved version, faster-R-CNN+++)

ADAS using Deep CNN – More Object Detection

DESAY SV automotive



Model	Train	Test	mAP	FLOPS	FPS	Cfg	Weights
Old YOLO	VOC 2007+2012	2007	63.4	40.19 Bn	45		link
SSD300	VOC 2007+2012	2007	74.3	-	46		link
SSD500	VOC 2007+2012	2007	76.8	-	19		link
YOLOv2	VOC 2007+2012	2007	76.8	34.90 Bn	67	cfg	weights
YOLOv2 544x544	VOC 2007+2012	2007	78.6	59.68 Bn	40	cfg	weights
Tiny YOLO	VOC 2007+2012	2007	57.1	6.97 Bn	207	cfg	weights
SSD300	COCO trainval	test-dev	41.2	-	46		link
SSD500	COCO trainval	test-dev	46.5	-	19		link
YOLOv2 544x544	COCO trainval	test-dev	44.0	59.68 Bn	40	cfg	weights
Tiny YOLO	COCO trainval	-	-	7.07 Bn	200	cfg	weights

- Joseph Redmon, Santosh Divvala, "You Only Look Once: Unified, Real-Time Object Detection", CVPR 2016
- https://github.com/xingwangsfu/caffe-yolo

More SoC Hardwares Are Available

Realtime Object Detection with SSD on Nvidia Jetson TX1

Nov 27, 2016

Realtime object detection is one of areas in computer vision that is still quite challenging performance-wise. When it comes to mobile/embedded application, GPUs certainly make a whole lot of difference allowing to achieve practically useful speeds. For example, SSD model described below runs at ~8.5 FPS on GPU and 0.03 FPS in CPU-only mode on TX1 board.

- Deep Learning Practitioner's Blog, https://myurasov.github.io/2016/11/27/ssd-tx1.html

More SoC Hardwares Are Available – cont.

- www.nvidia.com

More SoC Hardwares Are Available – cont.

DESAY SV automotive

Tesla Model S

From Wikipedia, the free encyclopedia

"Model S" redirects here. For the fighter aircraft, see Curtiss Model S.

The **Tesla Model S** is a full-sized all-electric five-door, luxury liftback, produced by Tesla Inc., and introduced in June 2012.[10] It scored a perfect 5.0 NHTSA automobile safety rating, and, as of March 2017, the P100D variant holds the record for the fastest acceleration of any production vehicle from a standstill to 60 mph in Motor Trend tests.[11] The United States Environmental Protection Agency (EPA) official range for the 2012 Model S Performance model equipped with an 85 kWh (310 MJ) battery pack is 265 miles (426 km), higher than any other electric car at the time.[12][13][14] EPA rated the 2012 85kWh Model S's energy consumption at 237.5 watt-hours per kilometer (38 kWh/100 mi or 24 kWh/100 km) for a combined fuel economy of 89 miles per gallon gasoline equivalent (2.64 L/100 km or 107 mpg_{-imp}).[12][15]

In 2016, Tesla updated the design of the Model S to closely match that of the Model X. As of January 2017, the following versions are available: 60, 60D, 75, 75D, 90D and P100D. Owners of the earlier 70 and 70D Model S have the option to unlock the 75 kWh capacity via a software update, adding up to 19 miles (31 km) per charge.[16] The 60 and 60D Model S, reintroduced in June 2016, have a US$9,000 option to unlock the full 75 kWh capacity via a software update any time after purchase.[17] In August 2016, Tesla introduced the P100D as the new top-level model. The P100D model has a 100 kWh battery, a 0–60 mph (0–97 km/h) time of 2.28 seconds, and over 300 miles (485 km) of EPA rated range.[18]

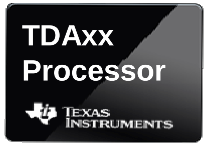
Tesla Model S	

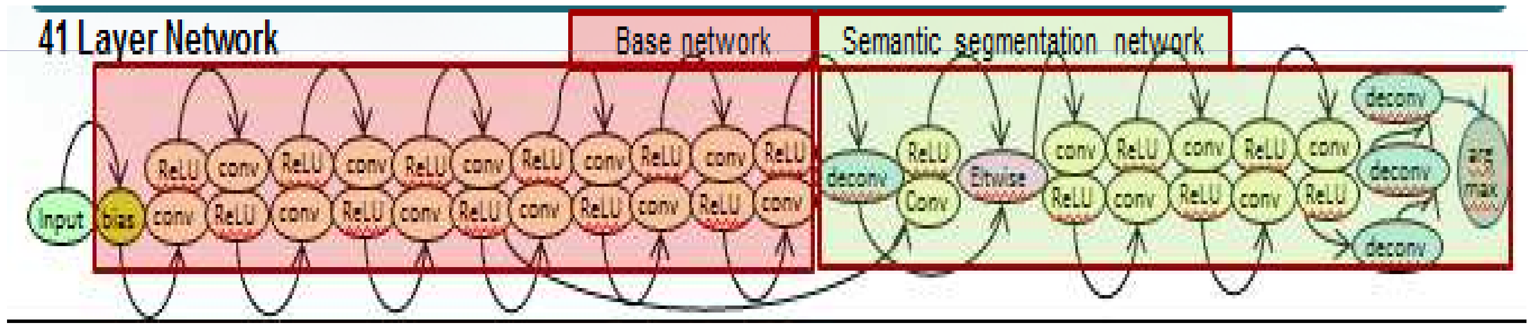
Overview	
Manufacturer	Tesla Inc.
Also called	Code name: WhiteStar[1][2][3]
Production	2012–present

- https://en.wikipedia.org/wiki/Tesla_Model_S

More SoC Hardwares Are Available – cont.

DESAY SV automotive





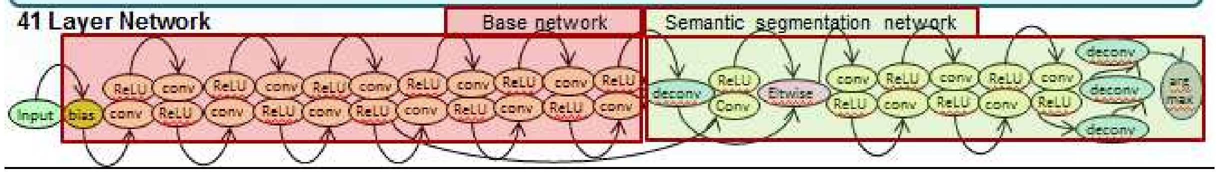
* TI Deep Learning Library And Semantic Segmentation Demo

More SoC Hardwares Are Available – cont.

DESAY SV automotive

Semantic Segmentation

- **Semantic Segmentation (or pixel classification)** partitions the image into semantically meaningful parts, and labels each part.
- **TI Deep learning Library (TIDL) & Deep learning network (JSegNet)** accelerates Semantic Segmentation using variety of public/custom networks trained with open source frameworks.
  - Semantic Segmentation Network with 41 Layer. Designed for 8 classes, trained for 5 classes (pedestrian, vehicles, roads, road sign, background)
- **High Efficiency & Low Power implementation**
  - **270 Giga** DLOPs/sec in less than **2.5W** with **2*C66x DSP** and **2*A15** completely free

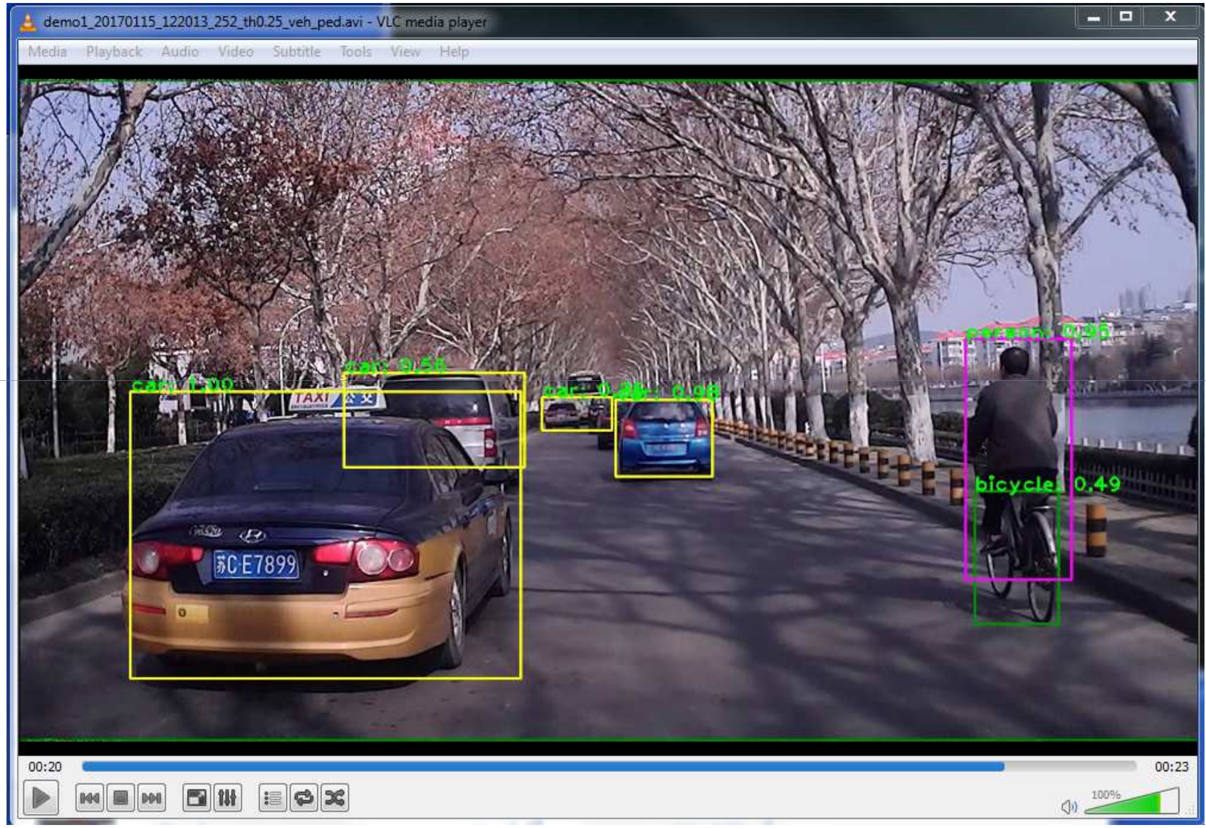


* TI Deep Learning Library And Semantic Segmentation Demo

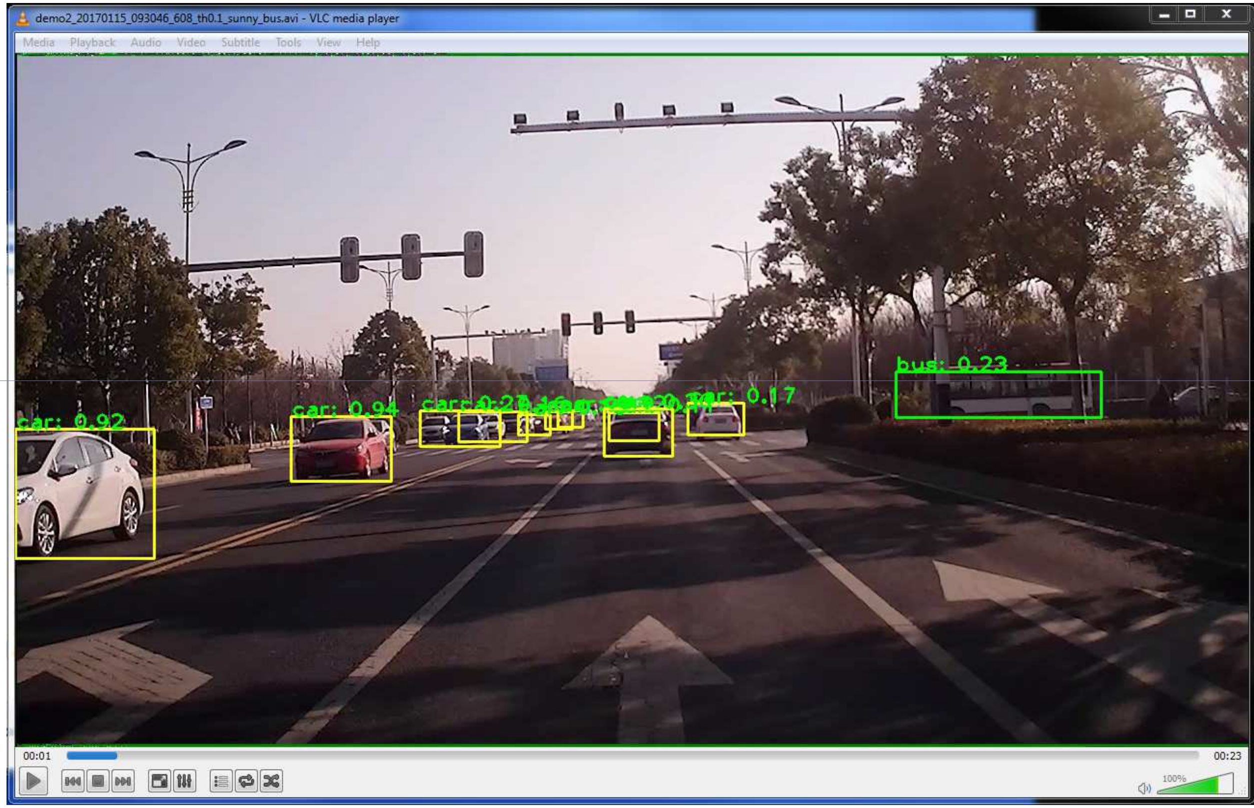
Content

- ❑ Overview
- ❑ Machine Learning
- ❑ Neural Network
  - ❑ Shallow Architecture
  - ❑ Deep Architecture
  - ❑ Convolutional Neural Network (CNN)
- ❑ Big Data and Deep Neural Network
- ❑ Deep NN (Deep CNN) for ADAS
- ❑ Demo

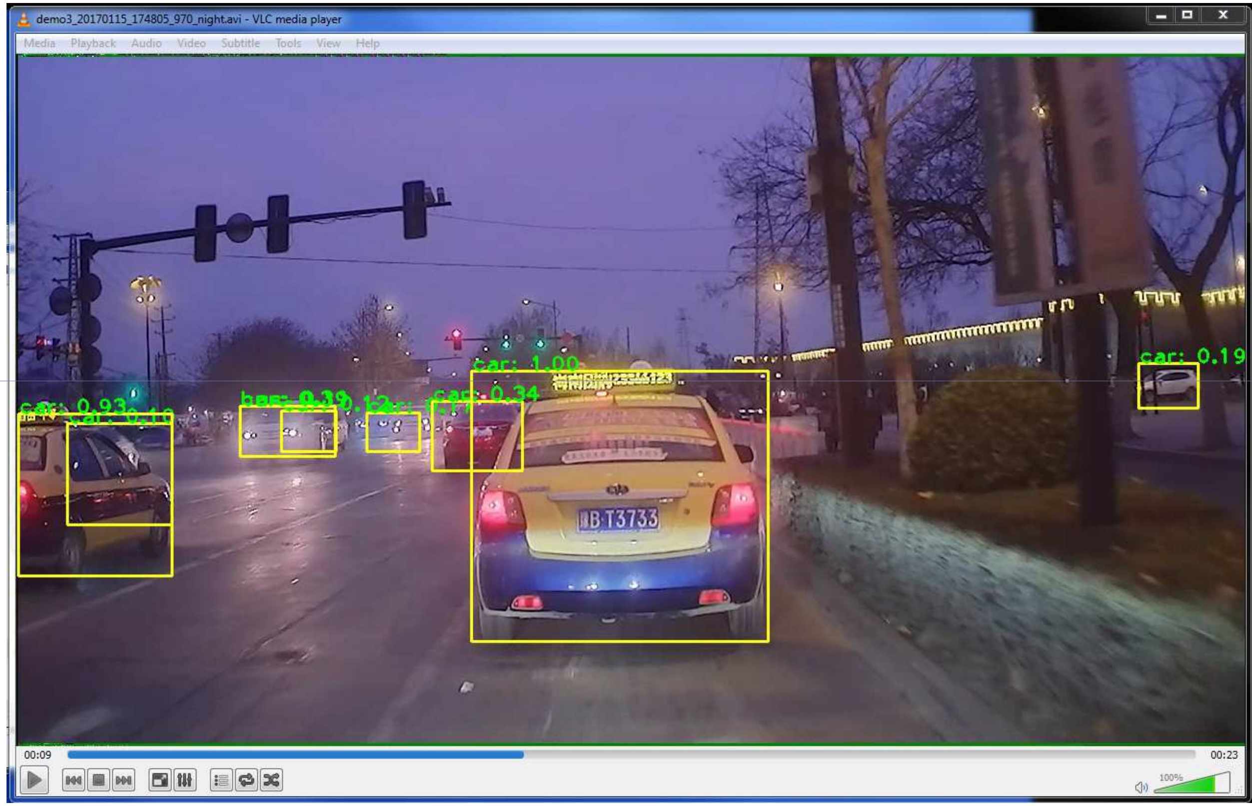
Examples of ADAS on Object Detection



Examples of ADAS on Object Detection – cont.



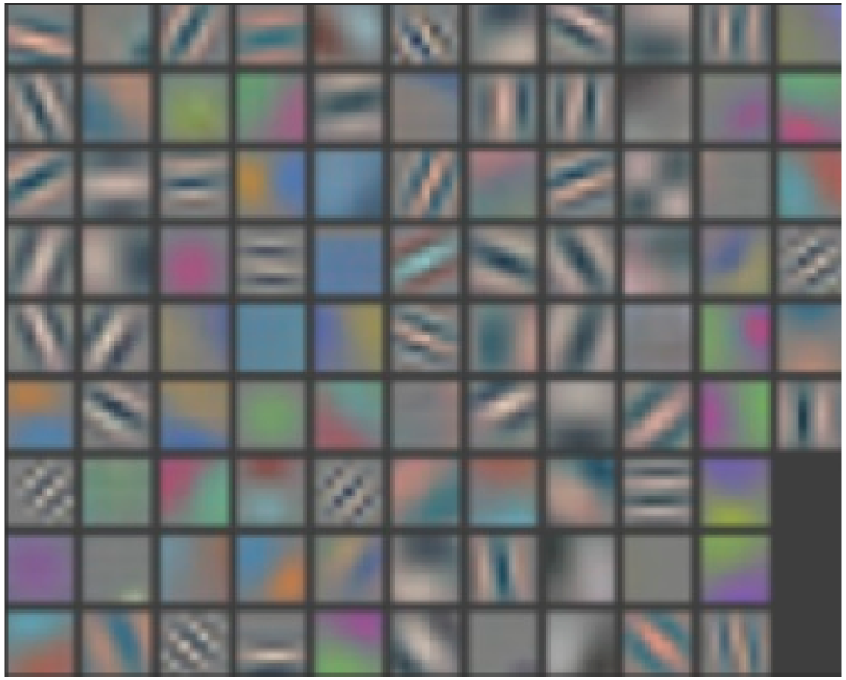
Examples of ADAS on Object Detection – cont.



Thank You !

DESAY SV
automotive

Why Deep Learning – First Layer



* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, "DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe", Univ. of Berkeley, BerkeleyVision

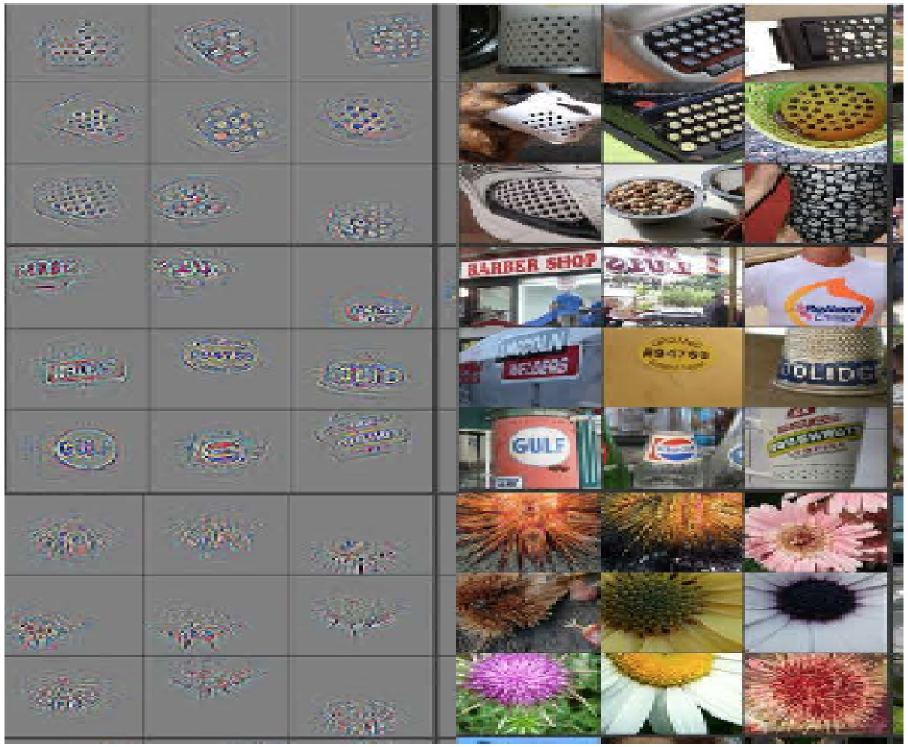
Why Deep Learning – Going deeper

* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, "DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe", Univ. of Berkeley, BerkeleyVision

Why Deep Learning – and deeper

* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, "DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe", Univ. of Berkeley, BerkeleyVision

Why Deep Learning – Global Feature Extraction



conv$_5$ DeConv visualization

* Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick, “DIY Deep Learning for Vision: a Hands-On Tutorial with Caffe”, Univ. of Berkeley, BerkeleyVision

Problem of Deep Learning

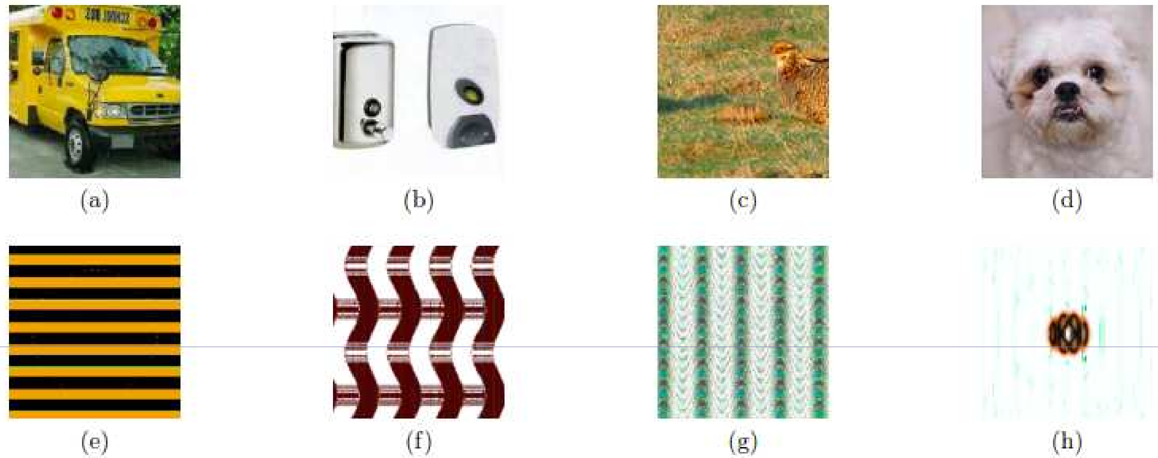


Figure 20: Illustrations of some mistakes of neural networks. (a)-(d) (from (Szegedy et al., 2013)) are adversarial images that are generated based on original images. The differences between these and the original ones are un-observable by naked eye, but the neural network can successfully classify original ones but fail adversarial ones. (e)-(h) (from (Nguyen et al., 2015)) are patterns that are generated. A neural network classify them into (e) school bus, (f) guitar, (g) peacock and (h) Pekinese respectively.

* Haohan Wang, Bhiksha Raj, "On the Origin of Deep Learning", School of Computer Science, Carnegie Mellon University, arXiv:1702.07800v3 [cs.LG] 2 Mar 2017