Survey of DNN Development Resources

MICRO Tutorial (2016)

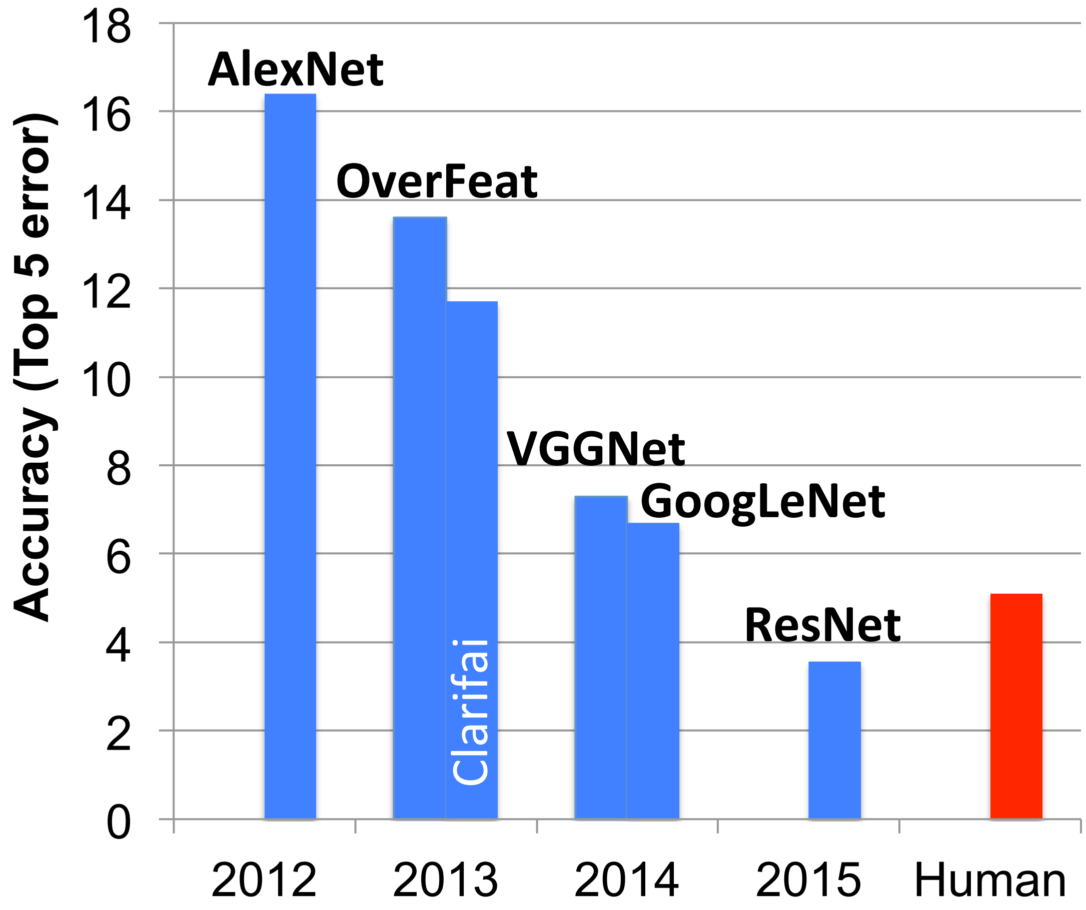
Website: http://eyeriss.mit.edu/tutorial.html

Joel Emer, Vivienne Sze, Yu-Hsin Chen

Popular DNNs

- LeNet (1998)
- AlexNet (2012)
- OverFeat (2013)
- VGGNet (2014)
- GoogleNet (2014)
- ResNet (2015)

ImageNet: Large Scale Visual Recognition Challenge (ILSVRC)

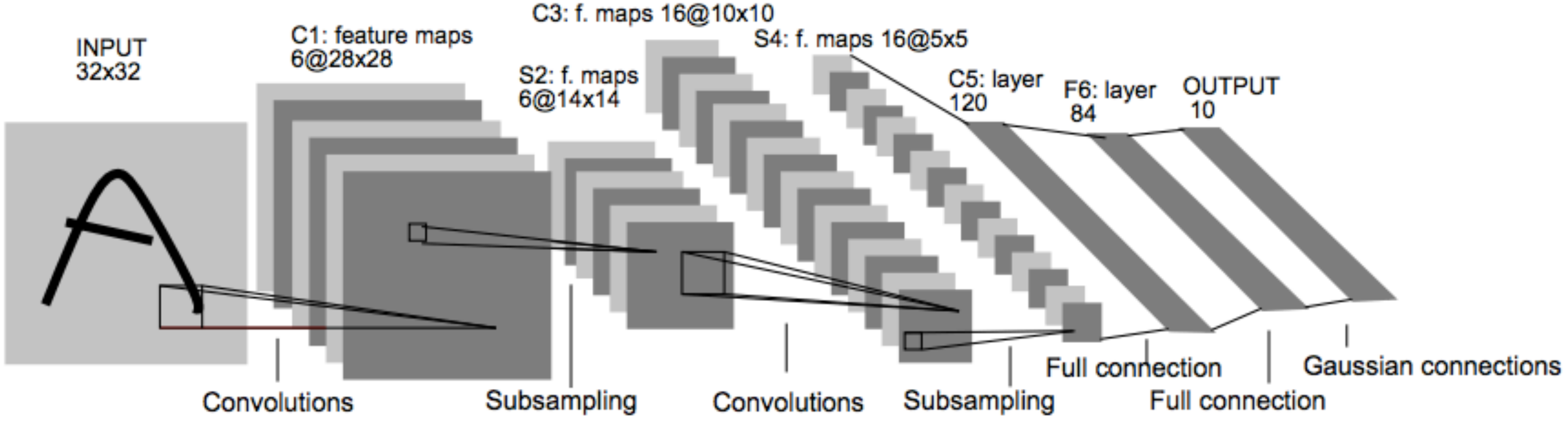


[O. Russakovsky et al., IJCV 2015]

LeNet-5

CONV Layers: 2
Fully Connected Layers: 2
Weights: 431k
MACs: 2.3M

Digit Classification!



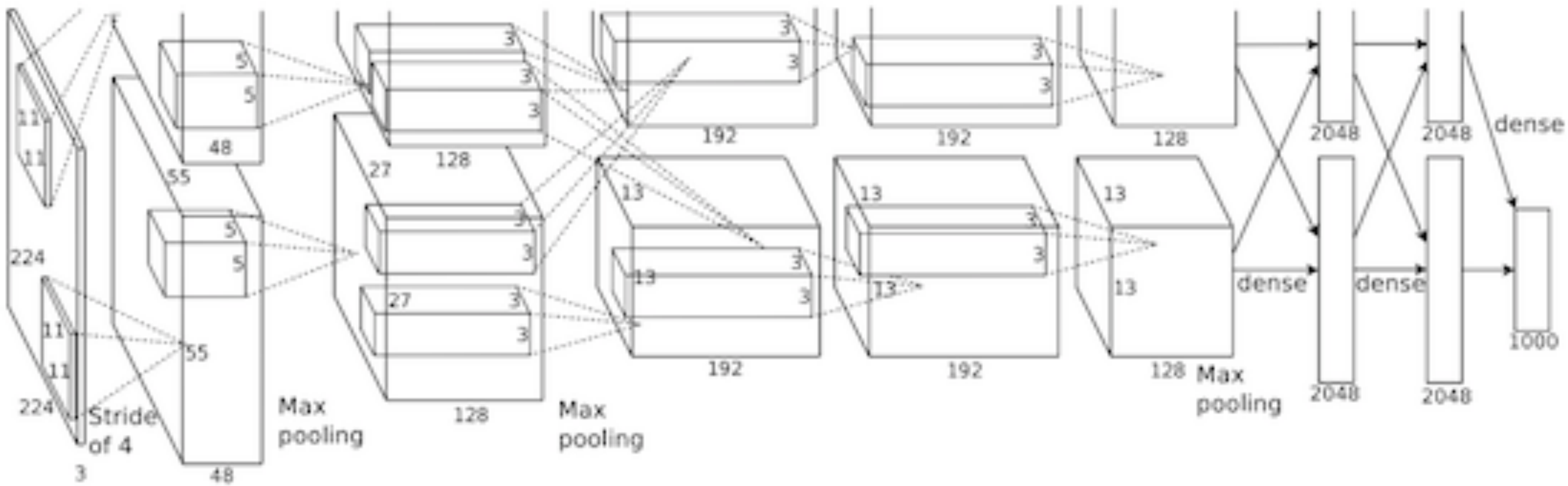
[Y. Lecun et al, Proceedings of the IEEE, 1998]

AlexNet

CONV Layers: 5
Fully Connected Layers: 3
Weights: 61M
MACs: 724M

ILSCVR12 Winner

Uses Local Response Normalization (LRN)



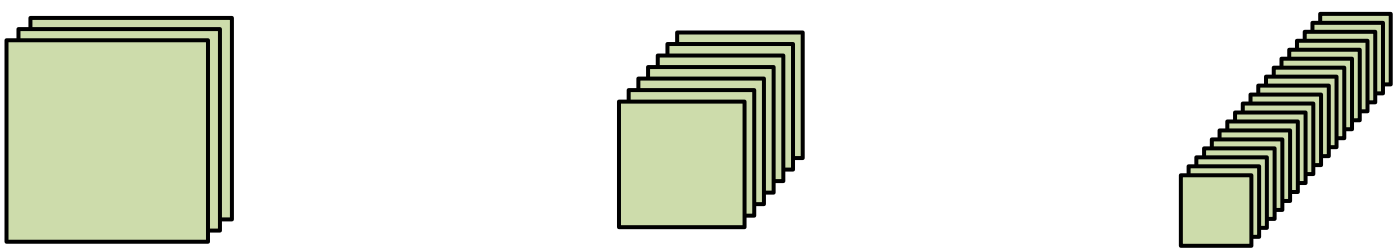
[Krizhevsky et al., NIPS, 2012]

Large Sizes with Varying Shapes

AlexNet Convolutional Layer Configurations

Layer	Filter Size (RxS)	# Filters (M)	# Channels (C)	Stride
1	11x11	96	3	4
2	5x5	256	48	1
3	3x3	384	256	1
4	3x3	384	192	1
5	3x3	256	192	1

Layer 1 **Layer 2** **Layer 3**



34k Params
105M MACs

307k Params
224M MACs

885k Params
150M MACs

[Krizhevsky et al., NIPS, 2012]

OverFeat (fast model)

CONV Layers: 5
Fully Connected Layers: 3
Weights: 144M
MACs: 5.4G

Layer	1	2	3	4	5	6	7	Output 8
Stage	conv + max	conv + max	conv	conv	conv + max	full	full	full
# channels	96	256	512	1024	1024	3072	4096	1000
Filter size	11x11	5x5	3x3	3x3	3x3	-	-	-
Conv. stride	4x4	1x1	1x1	1x1	1x1	-	-	-
Pooling size	2x2	2x2	-	-	2x2	-	-	-
Pooling stride	2x2	2x2	-	-	2x2	-	-	-
Zero-Padding size	-	-	1x1x1x1	1x1x1x1	1x1x1x1	-	-	-
Spatial input size	231x231	24x24	12x12	12x12	12x12	6x6	1x1	1x1

[Sermanet et al., ArXiv 2013, ICLR 2014]

VGG-16

CONV Layers: 16
Fully Connected Layers: 3
Weights: 138M
MACs: 15.5G

Also, 19 layer version

Reduce # of weights

More Layers → Deeper!

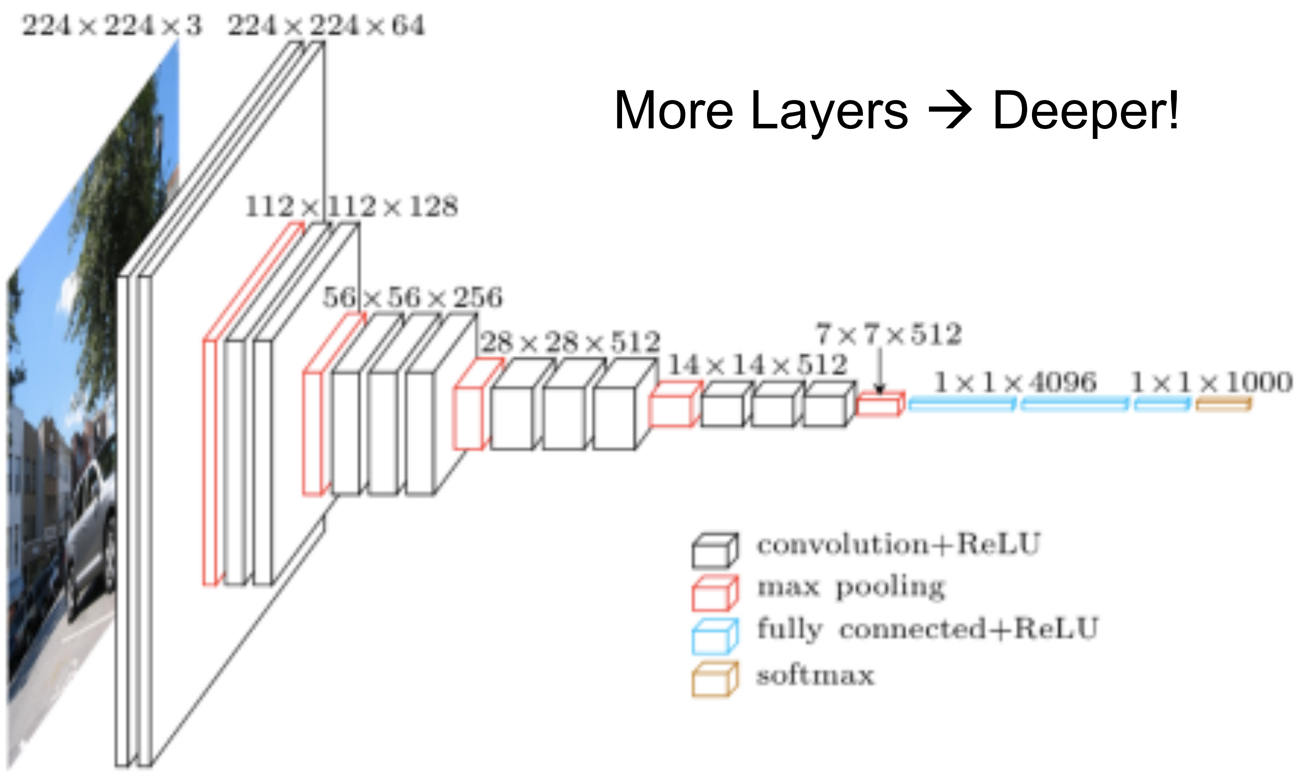
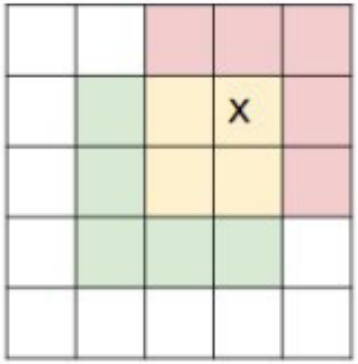


Image Source: http://www.cs.toronto.edu/~frossard/post/vgg16/

stack 2
3x3 conv



for a 5x5
receptive field

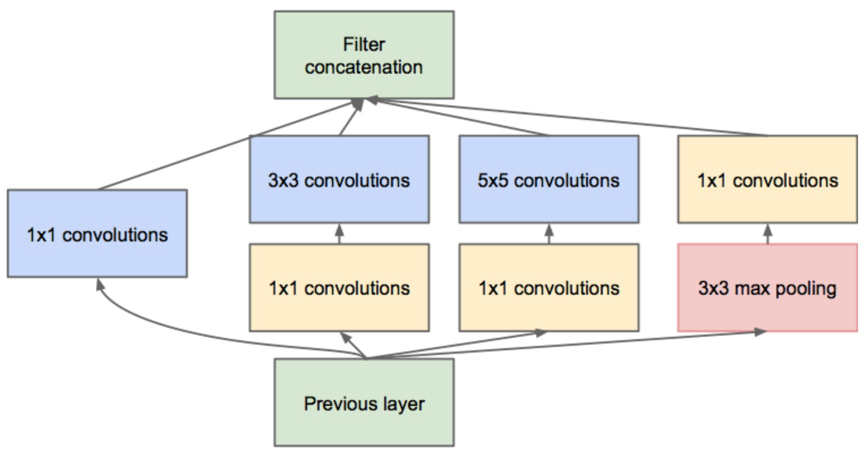
*[figure credit
A. Karpathy]*

[Simonyan et al., ArXiv 2014, ICLR 2015]

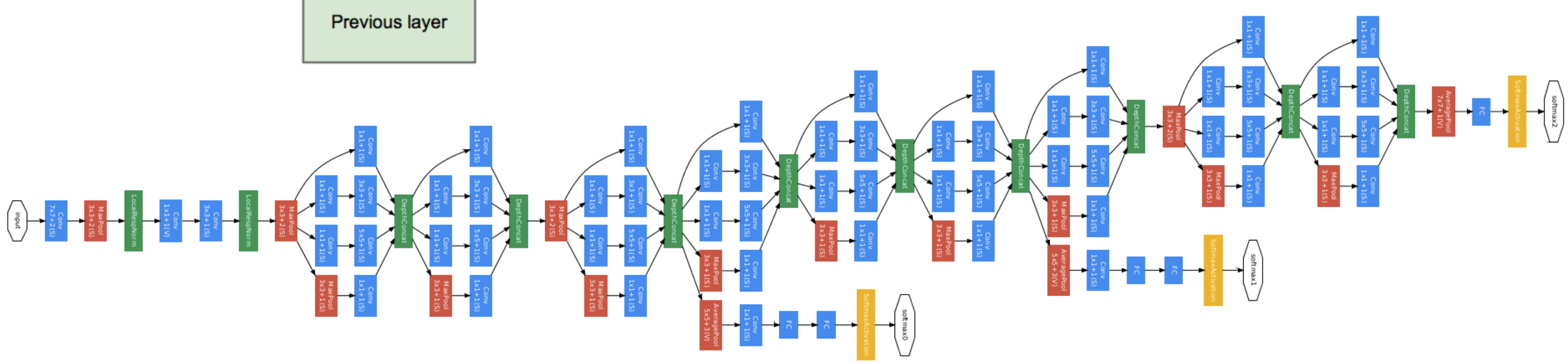
GoogLeNet (v1)

CONV Layers: 21
Fully Connected Layers: 1
Weights: 7.0M
MACs: 1.43G

Also, v2, v3 and v4
ILSVRC14 Winner



Inception module

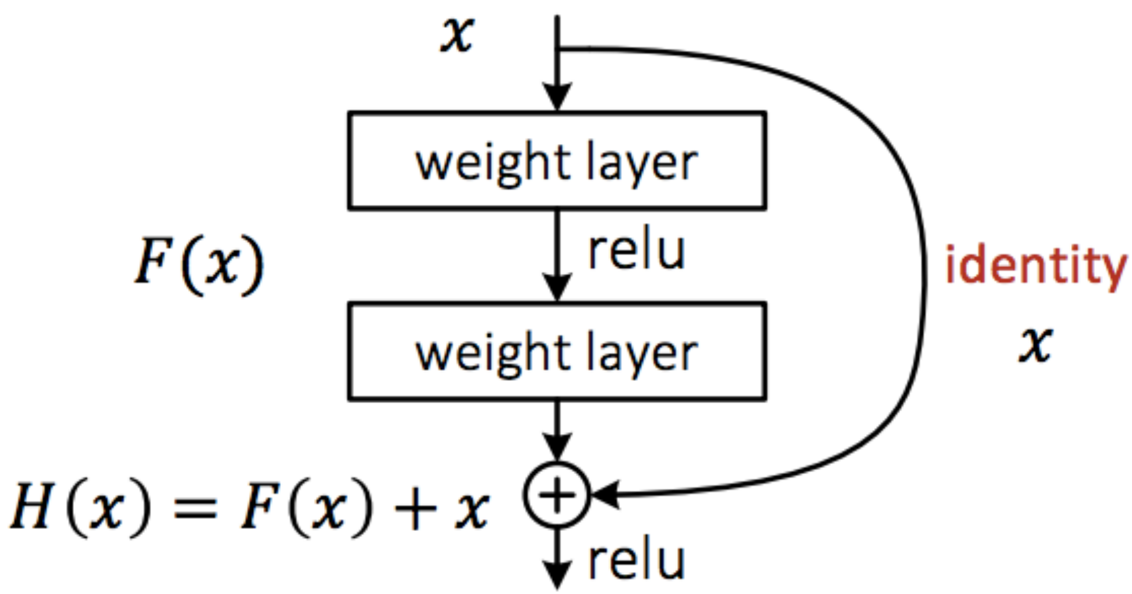


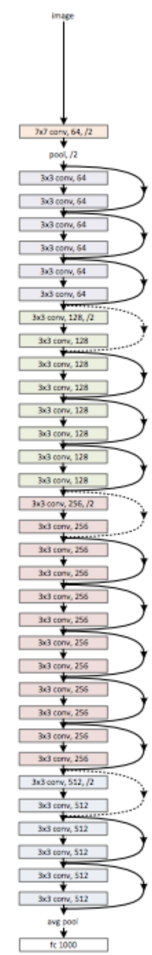
[Szegedy et al., ArXiv 2014, CVPR 2015]

ResNet-50

CONV Layers: 49
Fully Connected Layers: 1
Weights: 25.5M
MACs: 3.9G

Also, 34,152 and 1202 layer versions
ILSVRC15 Winner





[He et al., ArXiv 2015, CVPR 2016]

Revolution of Depth

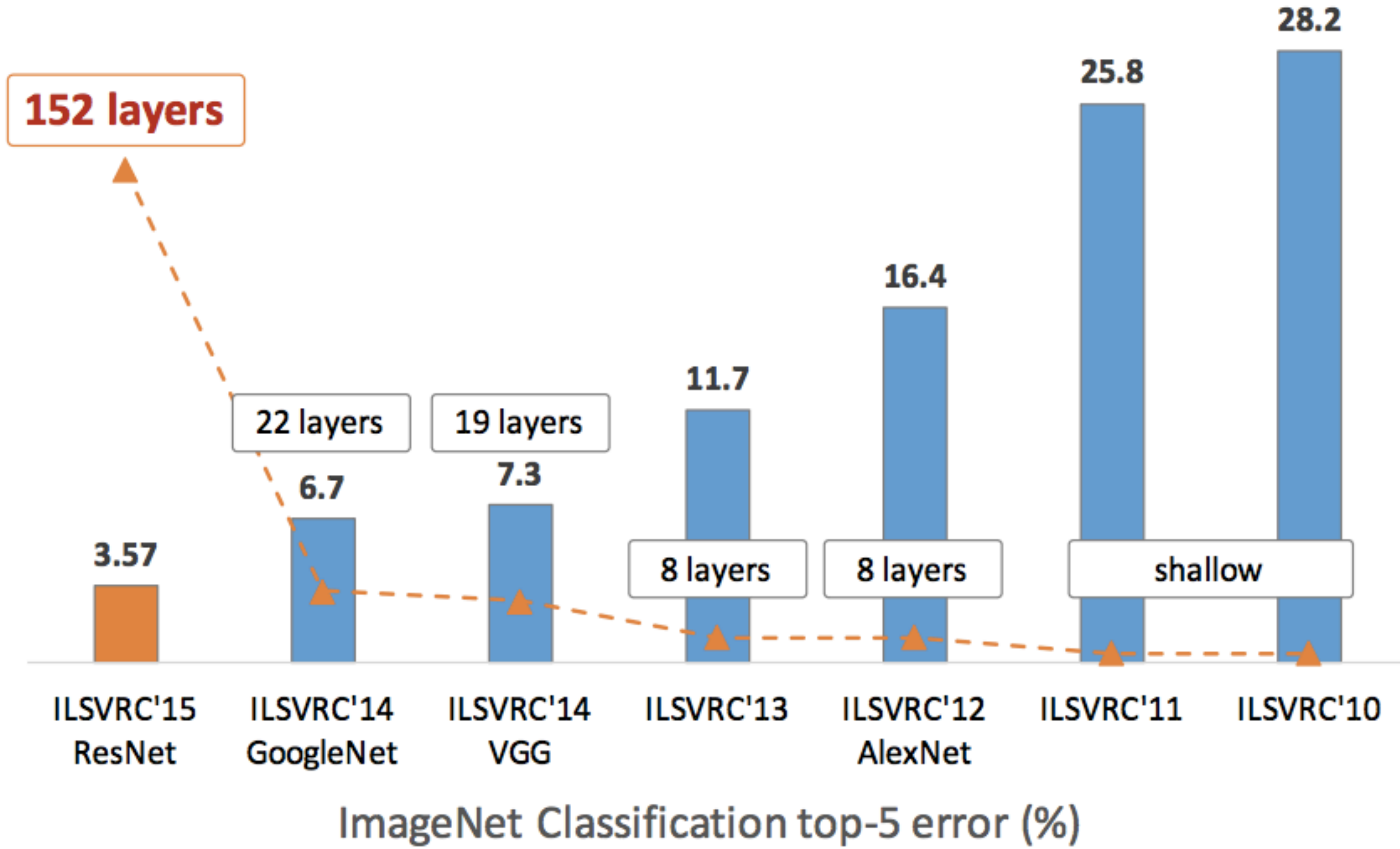


Image Source: http://icml.cc/2016/tutorials/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

Summary of Popular DNNs

Metrics	LeNet-5	AlexNet	OverFeat (fast)	VGG-16	GoogLeNet (v1)	ResNet-50
Top-5 error	n/a	16.4	14.2	7.4	6.7	5.3
Input Size	28x28	227x227	231x231	224x224	224x224	224x224
# of CONV Layers	**2**	**5**	**5**	**16**	**21**	**49**
Filter Sizes	5	3, 5,11	3, 7	3	1, 3 , 5, 7	1, 3, 7
# of Channels	1, 6	3 - 256	3 - 1024	3 - 512	3 - 1024	3 - 2048
# of Filters	6, 16	96 - 384	96 - 1024	64 - 512	64 - 384	64 - 2048
Stride	1	1, 4	1, 4	1	1, 2	1, 2
# of Weights	26k	2.3M	16M	14.7M	6.0M	23.5M
# of MACs	1.9M	666M	2.67G	15.3G	1.43G	3.86G
# of FC layers	**2**	**3**	**3**	**3**	**1**	**1**
# of Weights	406k	58.6M	130M	124M	1M	2M
# of MACs	405k	58.6M	130M	124M	1M	2M
Total Weights	**431k**	**61M**	**146M**	**138M**	**7M**	**25.5M**
Total MACs	**2.3M**	**724M**	**2.8G**	**15.5G**	**1.43G**	**3.9G**

CONV Layers increasingly important!

Summary of Popular DNNs

- **AlexNet**
  - First CNN Winner of ILSVRC
  - Uses LRN (deprecated after this)
- **VGG-16**
  - Goes Deeper (16+ layers)
  - Uses only 3x3 filters (stack for larger filters)
- **GoogLeNet (v1)**
  - Reduces weights with Inception and only one FC layer
  - Inception: 1x1 and DAG (parallel connections)
  - Batch Normalization
- **ResNet**
  - Goes Deeper (24+ layers)
  - Shortcut connections

Frameworks

Berkeley / BVLC
(C, C++, Python, MATLAB)

Google
(C++, Python)

U. Montreal
(Python)

Facebook / NYU
(C, C++, Lua)

Also, CNTK, MXNet, etc.
More at: https://developer.nvidia.com/deep-learning-frameworks

Example: Layers in Caffe

Convolution Layer

```
layer {
  name: "conv1"
  type: "Convolution"
  bottom: "data"
  top: "conv1"
 ...
  convolution_param {
    num_output: 20
    kernel_size: 5
    stride: 1
...
```

Non-Linearity

```
layer {
  name: "relu1"
  type: "ReLU"
  bottom: "conv1"
  top: "conv1"
}
```

Pooling Layer

```
layer {
  name: "pool1"
  type: "Pooling"
  bottom: "conv1"
  top: "pool1"
  pooling_param {
    pool: MAX
    kernel_size: 2
    stride: 2 ...
```

http://caffe.berkeleyvision.org/tutorial/layers.html

Image Classification Datasets

- **Image Classification/Recognition**
  - Given an entire image → Select 1 of N classes
  - No localization (detection)

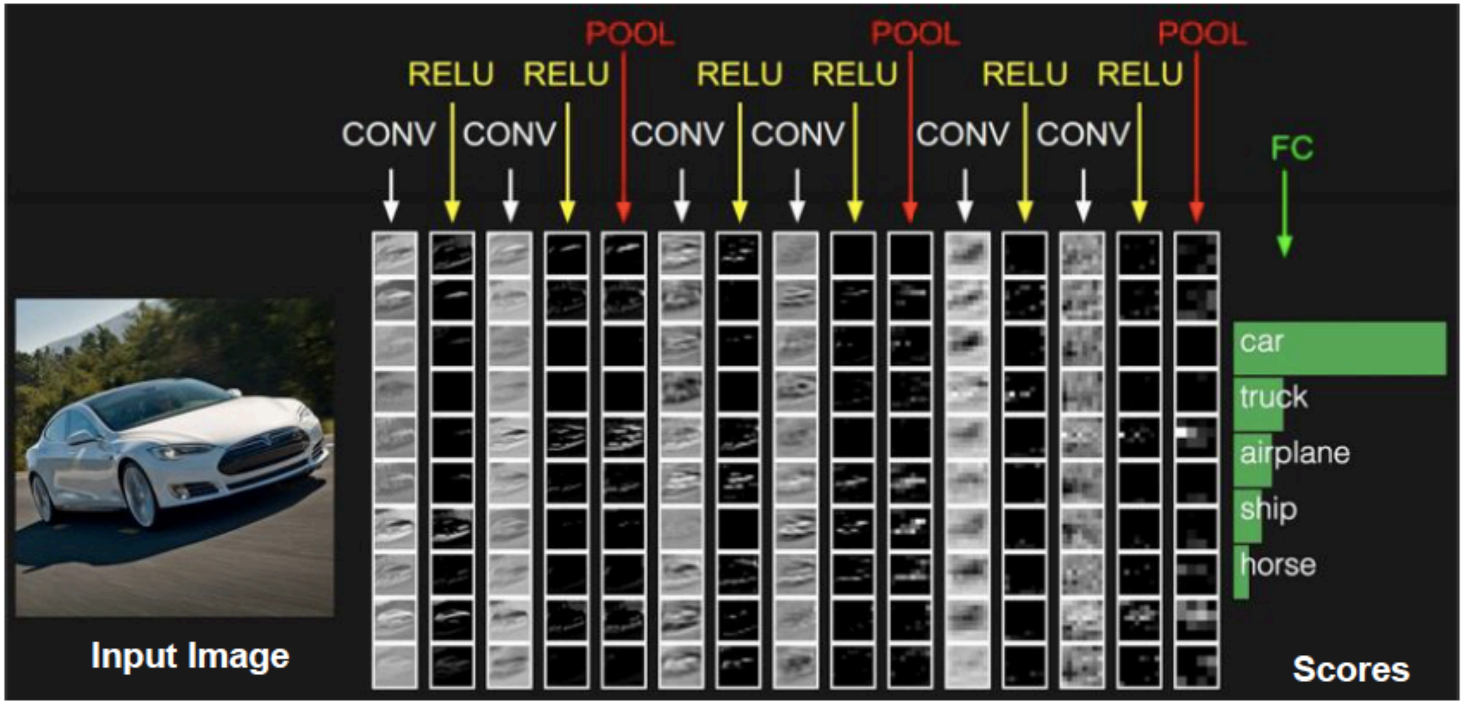


Image Source: Stanford cs231n

Datasets affect difficulty of task

MNIST

Digit Classification
28x28 pixels (B&W)
10 Classes
60,000 Training
10,000 Testing

LeNet in 1998
(0.95% error)

ICML 2013
(0.21% error)

http://yann.lecun.com/exdb/mnist/

CIFAR-10/CIFAR-100

Object Classification
32x32 pixels (color)
10 or 100 Classes
50,000 Training
10,000 Testing

CIFAR-10
RBM+finetuning in 2009
(35.16% error)

↓

ArXiv 2015
(3.47% error)

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

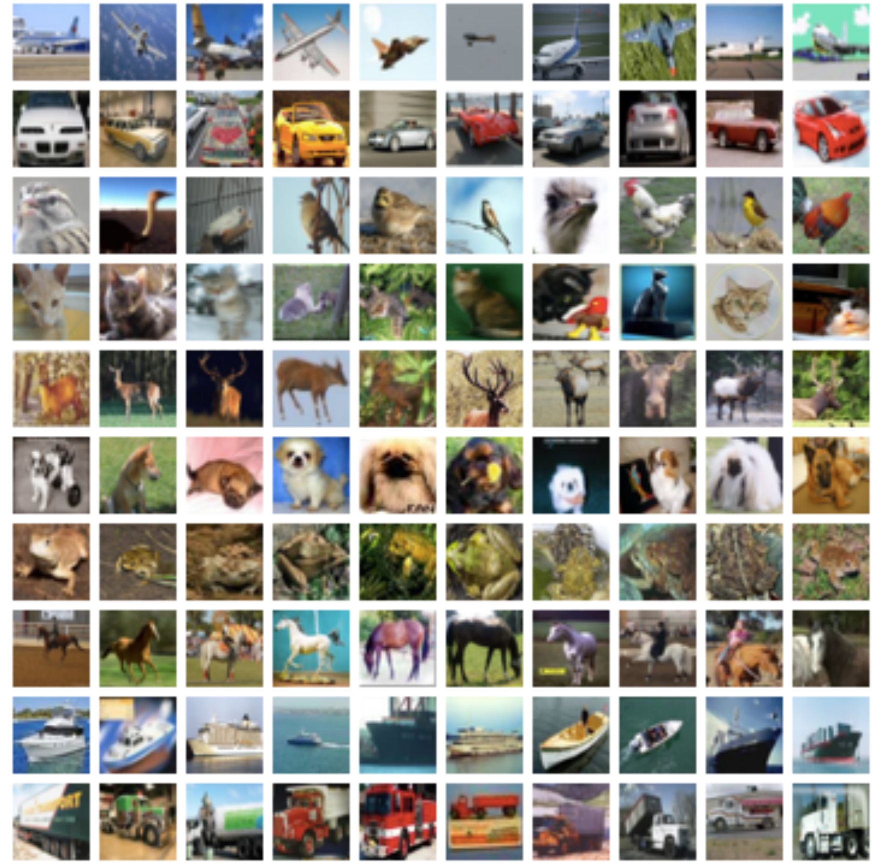


Image Source: http://karpathy.github.io/

Subset of 80 [Tiny Images Dataset](#) (Torrabla)

https://www.cs.toronto.edu/~kriz/cifar.html

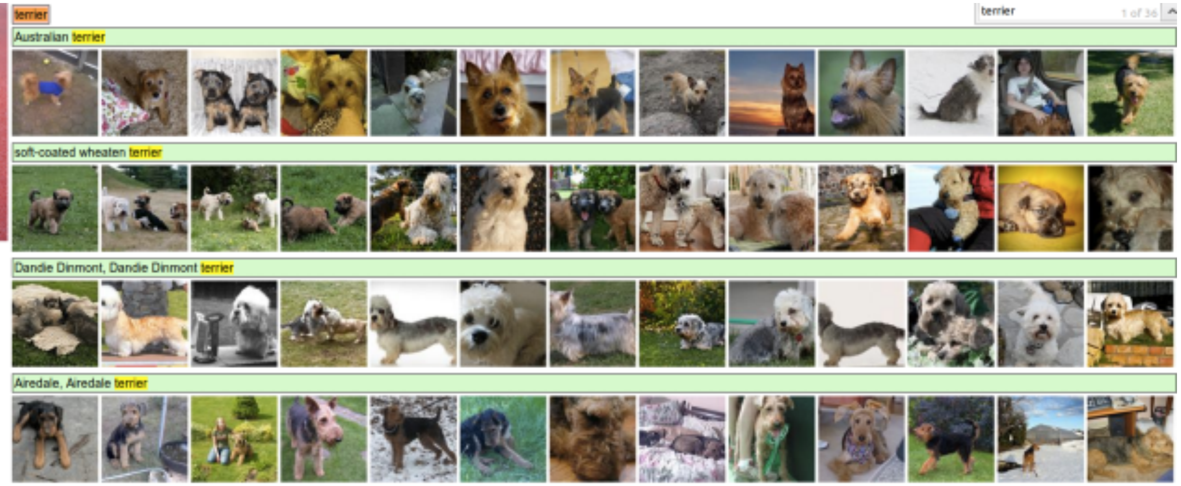
IMAGENET

Object Classification
~256x256 pixels (color)
1000 Classes
1.3M Training
100,000 Testing (50,000 Validation)

Image Source: http://karpathy.github.io/

http://www.image-net.org/challenges/LSVRC/

IMAGENET



Fine grained Classes
(120 breeds)

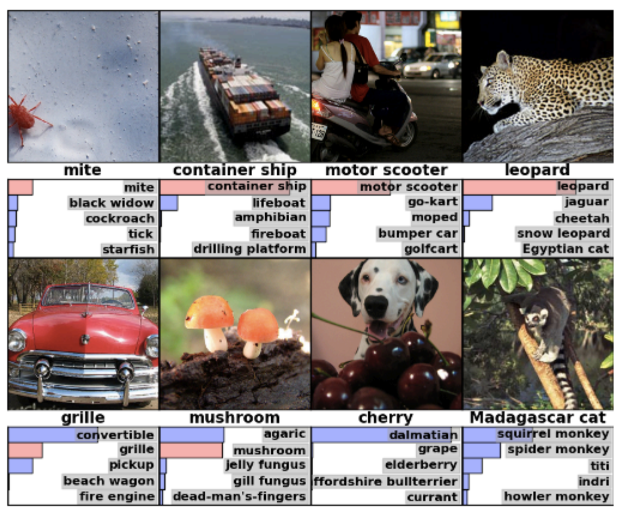
Image Source: http://karpathy.github.io/

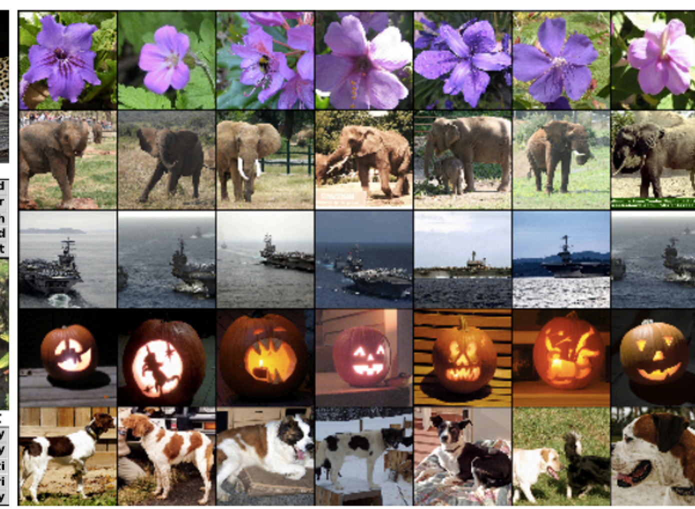
Image Source: Krizhevsky et al., NIPS 2012

Top-5 Error

Winner 2012
(16.42% error)

Winner 2016
(2.99% error)





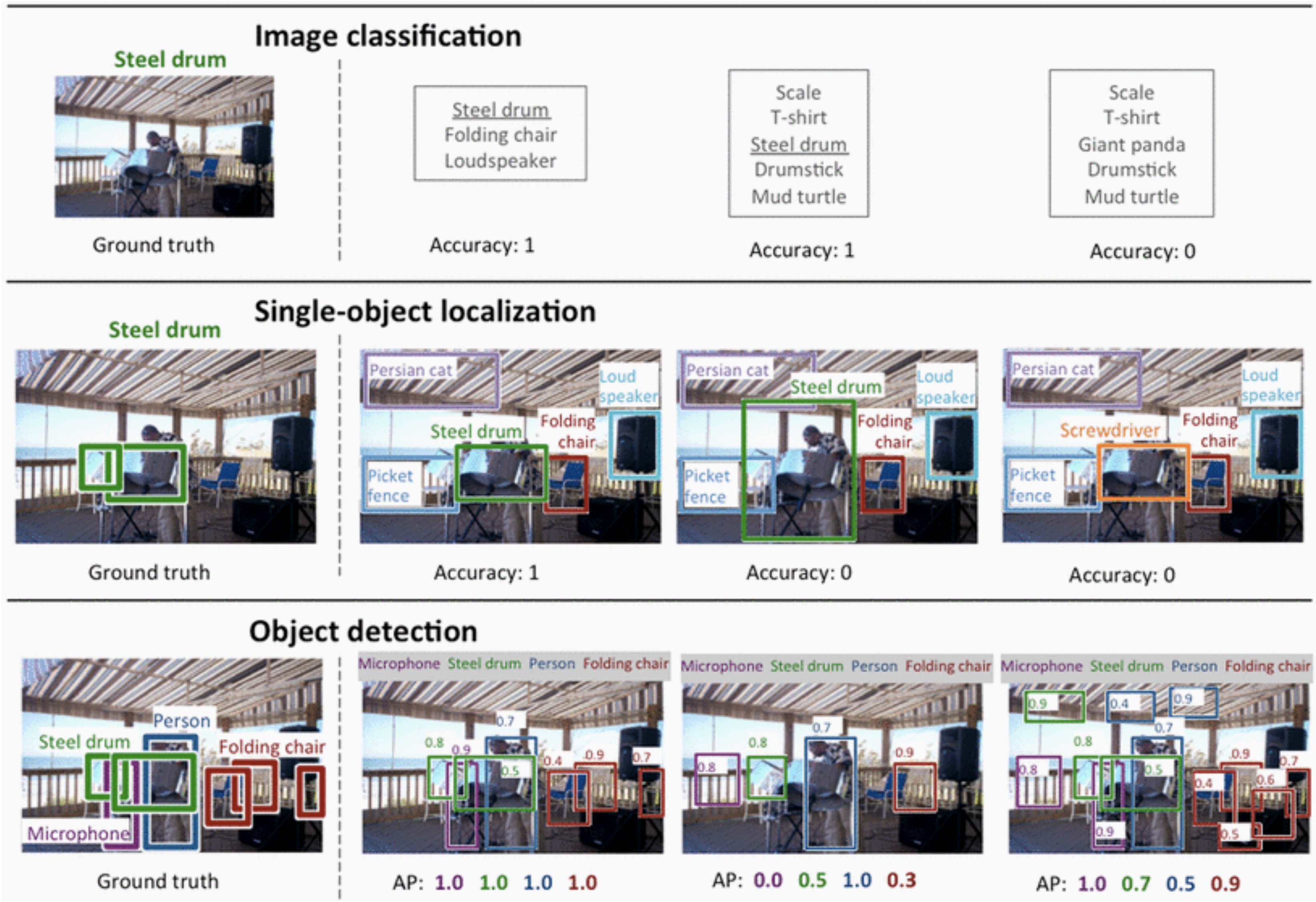
http://www.image-net.org/challenges/LSVRC/

Image Classification Summary

	MNIST	CIFAR-10	CIFAR-100	IMAGENET
Year	1998	2009	2009	2012
Resolution	28x28	32x32	32x32	256x256
Classes	10	10	100	1000
Training	60k	50k	50k	1.3M
Testing	10k	10k	10k	100k
Accuracy	0.21% error (ICML 2013)	3.47% error (arXiv 2015)	24.28% error (arXiv 2015)	2.99% top-5 error (2016 winner)

http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html

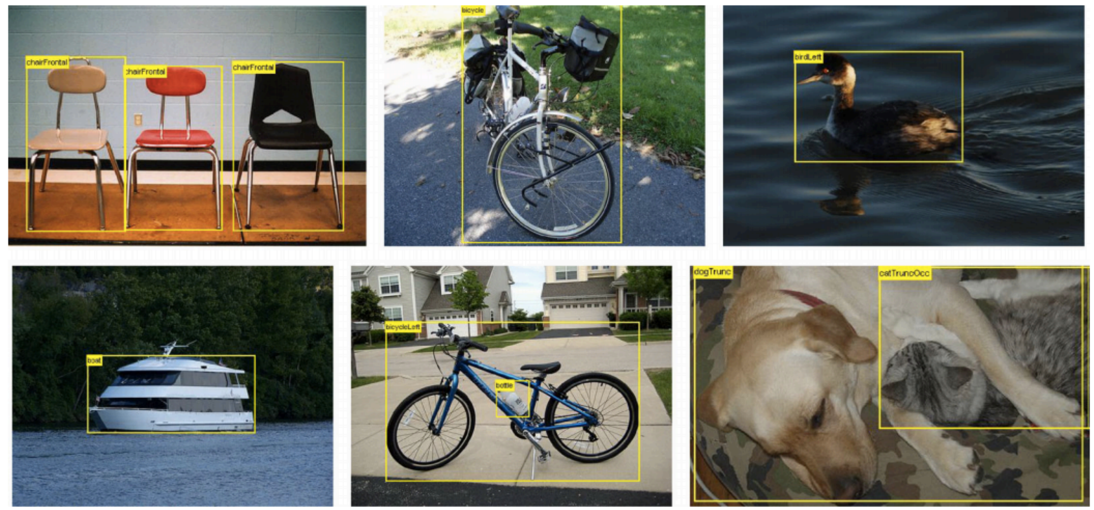
Next Tasks: Localization and Detection



[Russakovsky et al., IJCV, 2015]

Others Popular Datasets

- **Pascal VOC**
  - **11k images**
  - **Object Detection**
  - **20 classes**



- **MS COCO**
  - **300k images**
  - **Detection, Segmentation**
  - **Recognition in context**

http://host.robots.ox.ac.uk/pascal/VOC/ http://mscoco.org/

Recently Introduced Datasets

- **Announced Sept 2016:**
- **Google Open Images (~9M images)**
  - **https://github.com/openimages/dataset**
- **Youtube-8M (8M videos)**
  - **https://research.google.com/youtube8m/**