ILSVRC(The ImageNet Large Scale Visual Recognition Challenge)'s History

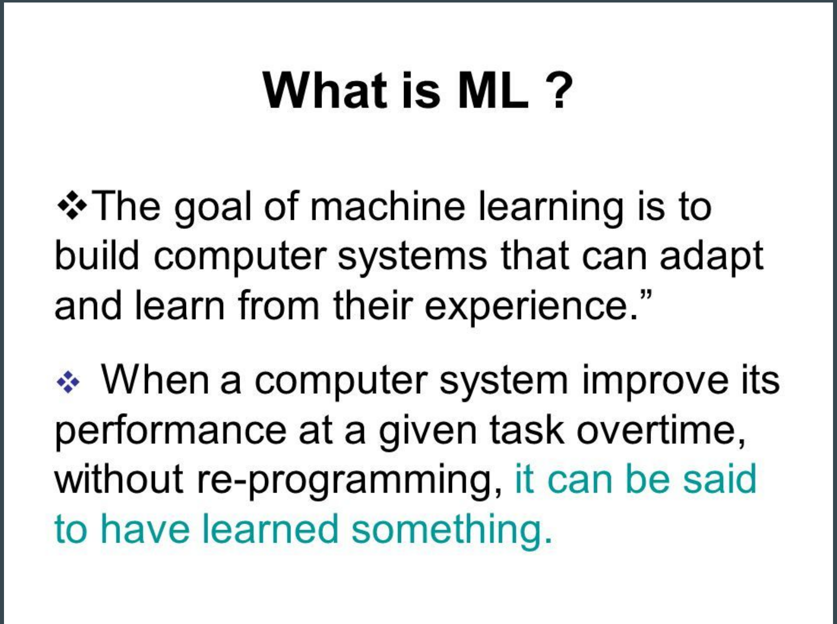
...

Deep Learning Reading Group - Session 1

Sepehr Sameni (@Separius)

Background

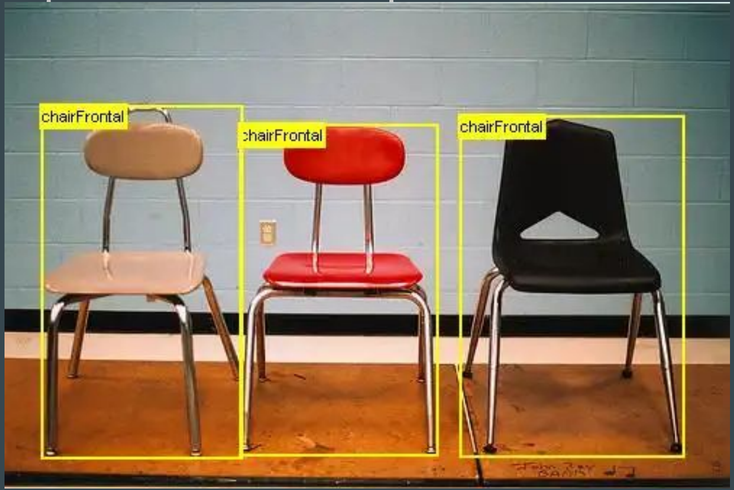
ML = Data + Prior



Contest

Standard evaluation method(Train/Test)

Supervised vs Unsupervised

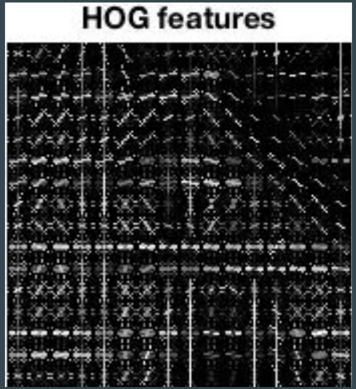


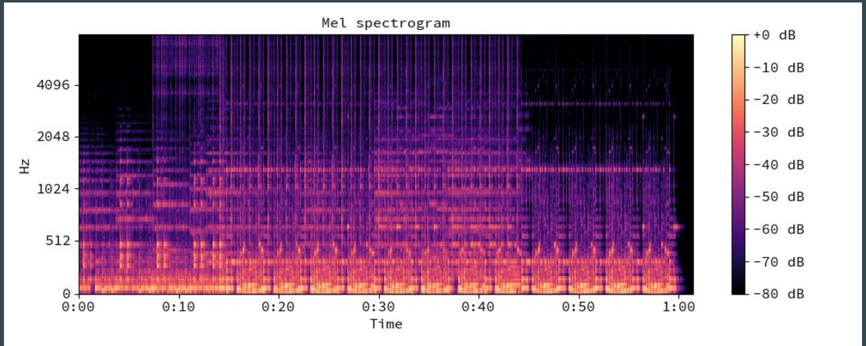
PASCAL visual object classes challenge

kaggle

Old Recipe

Feature Engineering + SVM/Linear classifier

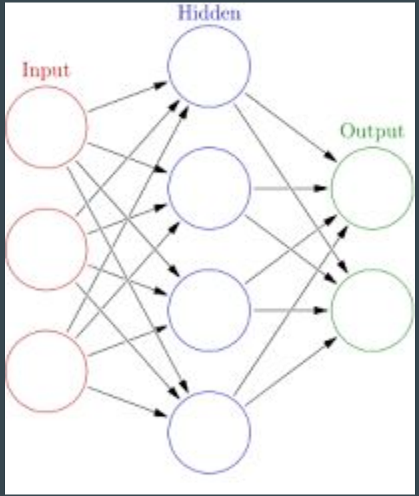


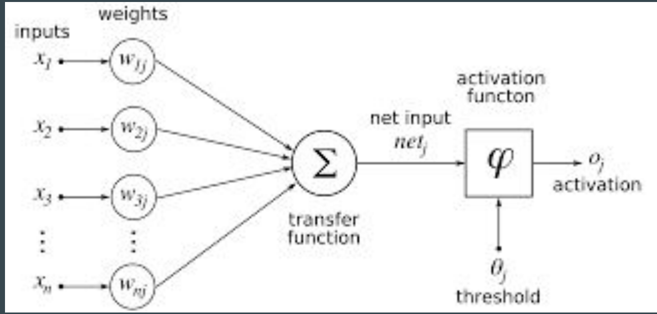


ANN

Diff with real?

- Diff learning
- Synapse computation
- Feedback
- More!

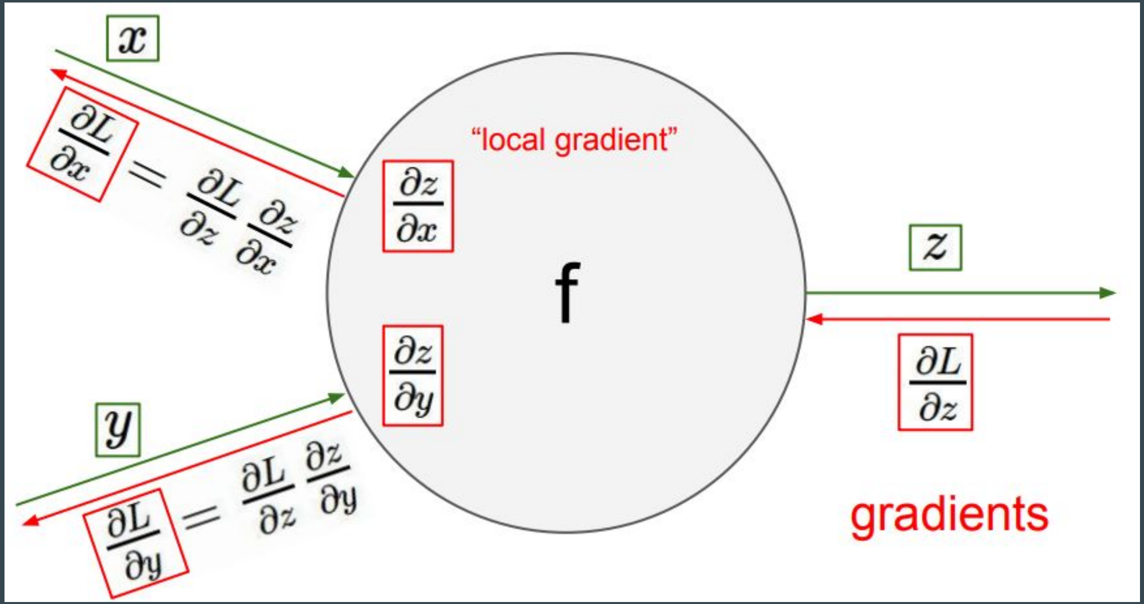




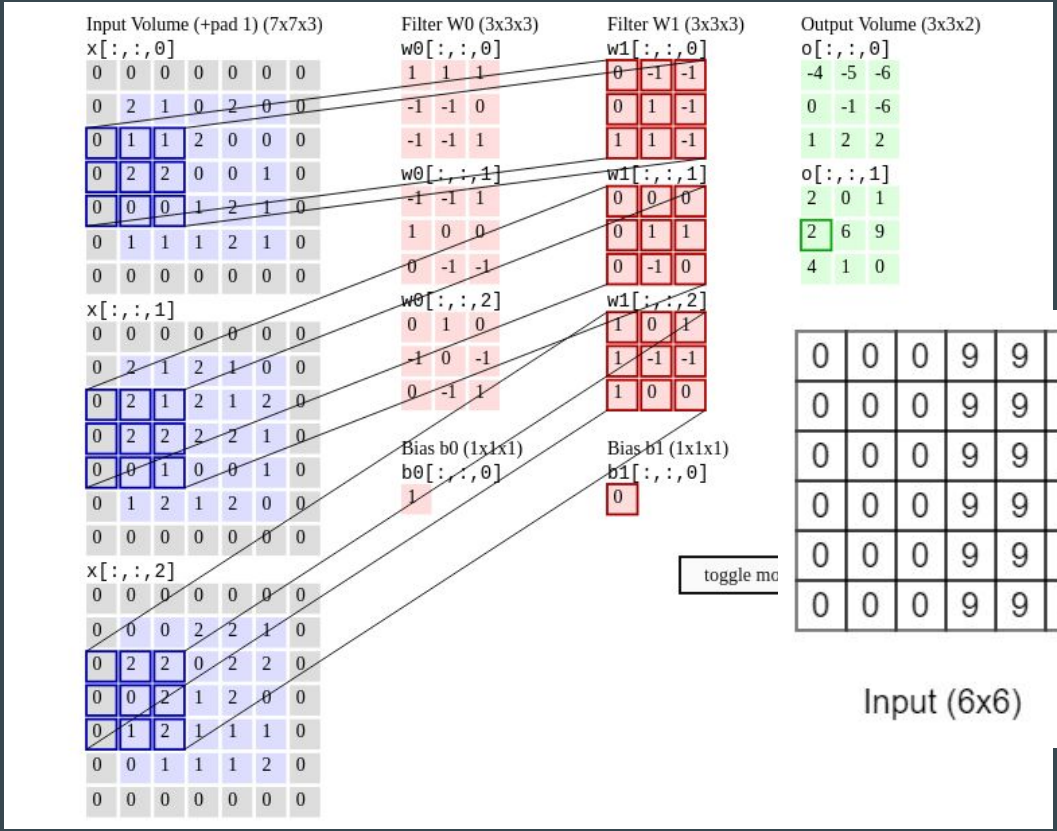
Universal Function Approximator(sigmoid)

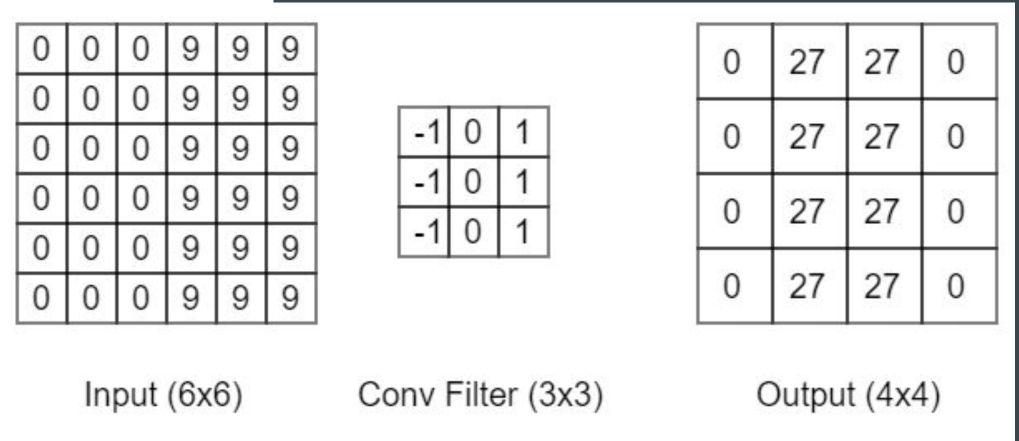
Backpropagation

Also talk about loss functions(mse, cross entropy, triplet)



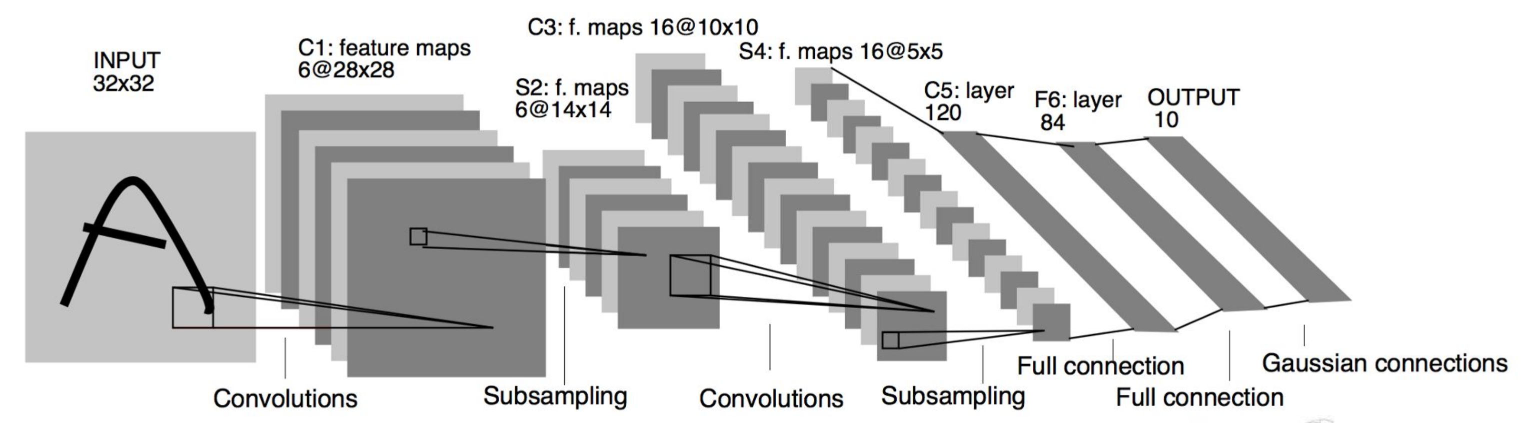
Convolution

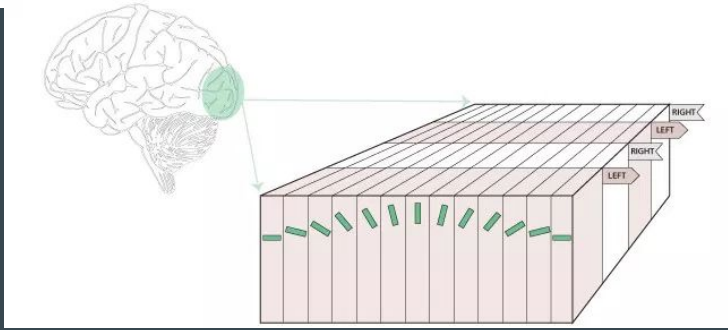




CNN [1]

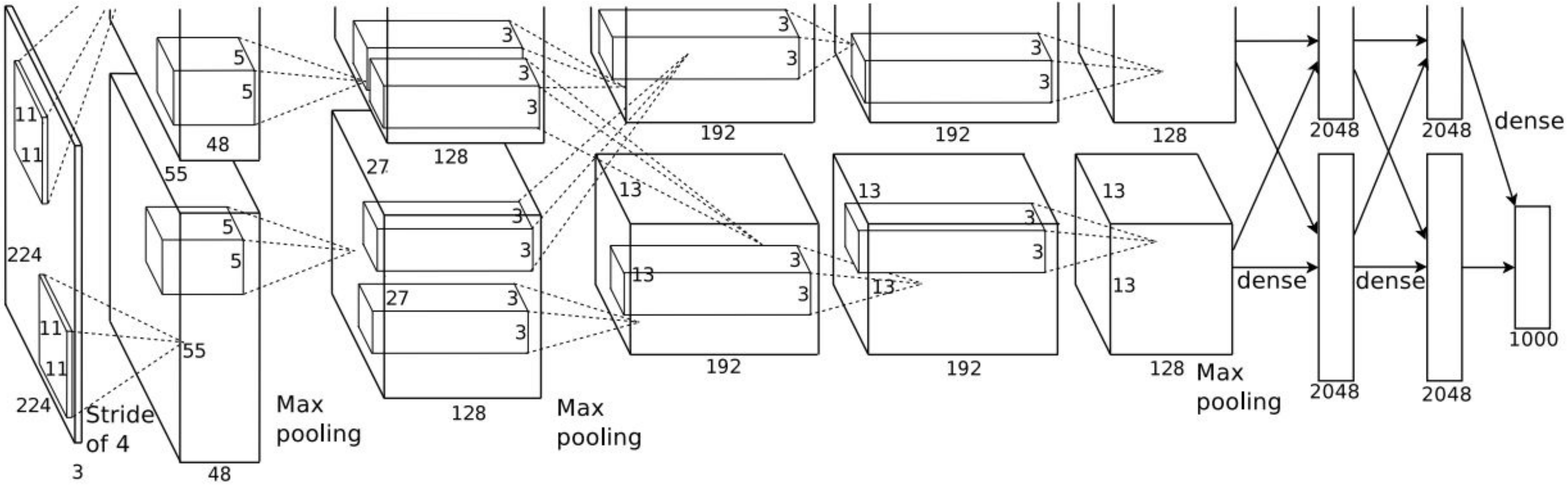
Structural Bias(bias/variance), Invariant vs Covariant, Pooling, Brain Inspired!





AlexNet(finally! :D) [2]

Hinton, GPU, Pooling; ReLU, Local Response Normalization(lateral inhibition)



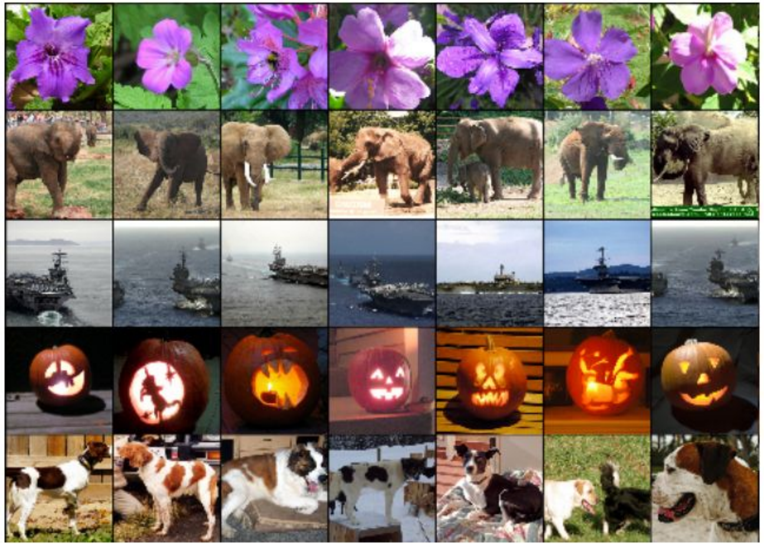
AlexNet(continued)

conv(3.7M), FC(58.6M) => over-fitting => flip, crop, color, dropout, weight decay

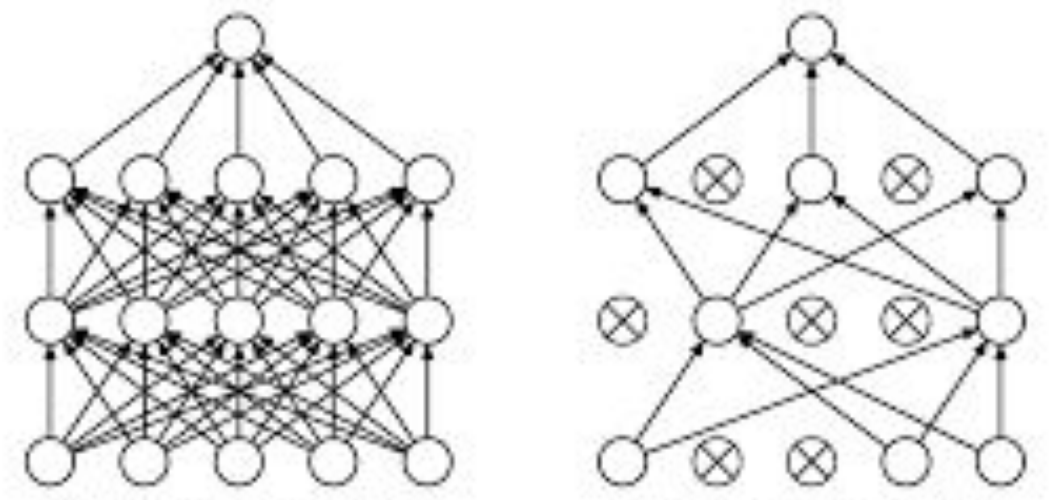
Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

Figure 3: 96 convolutional kernels of size $11\times11\times3$ learned by the first convolutional layer on the $224\times224\times3$ input images. The top 48 kernels were learned on GPU 1 while the bottom 48 kernels were learned on GPU 2. See Section 6.1 for details.



Dropout [3]



(a) Standard Neural Net

(b) After applying dropout.

Inception(idea) [4]

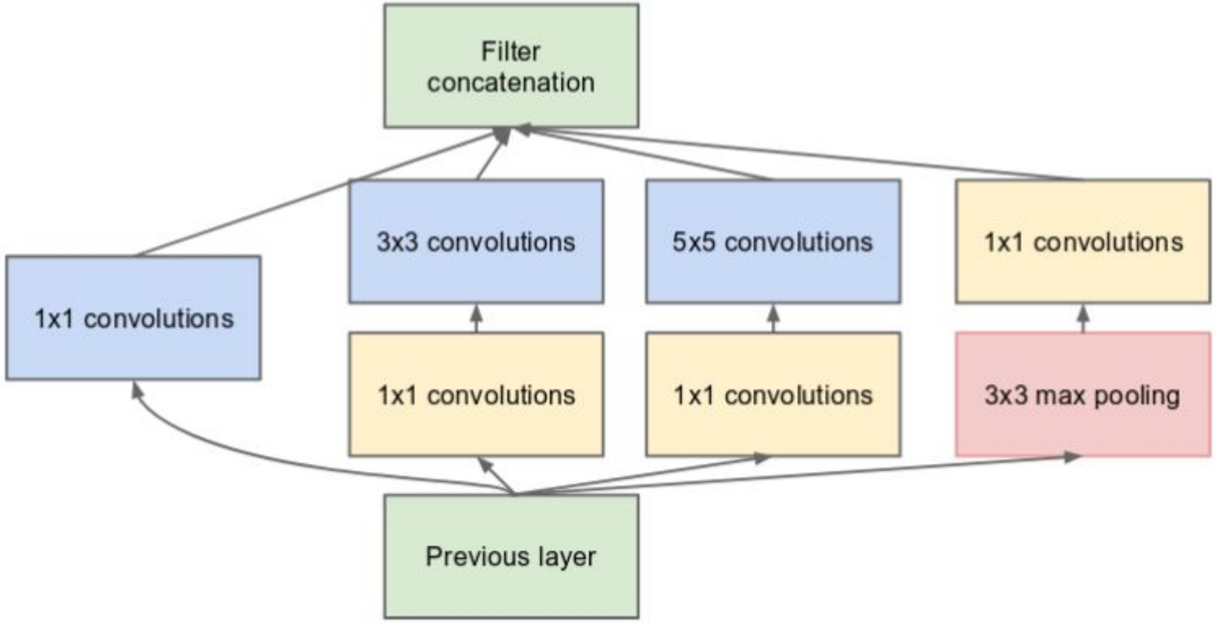
1.increase depth

 => over-fitting

Salient parts in the image can have extremely large variation in size. => 2. diff K sizes

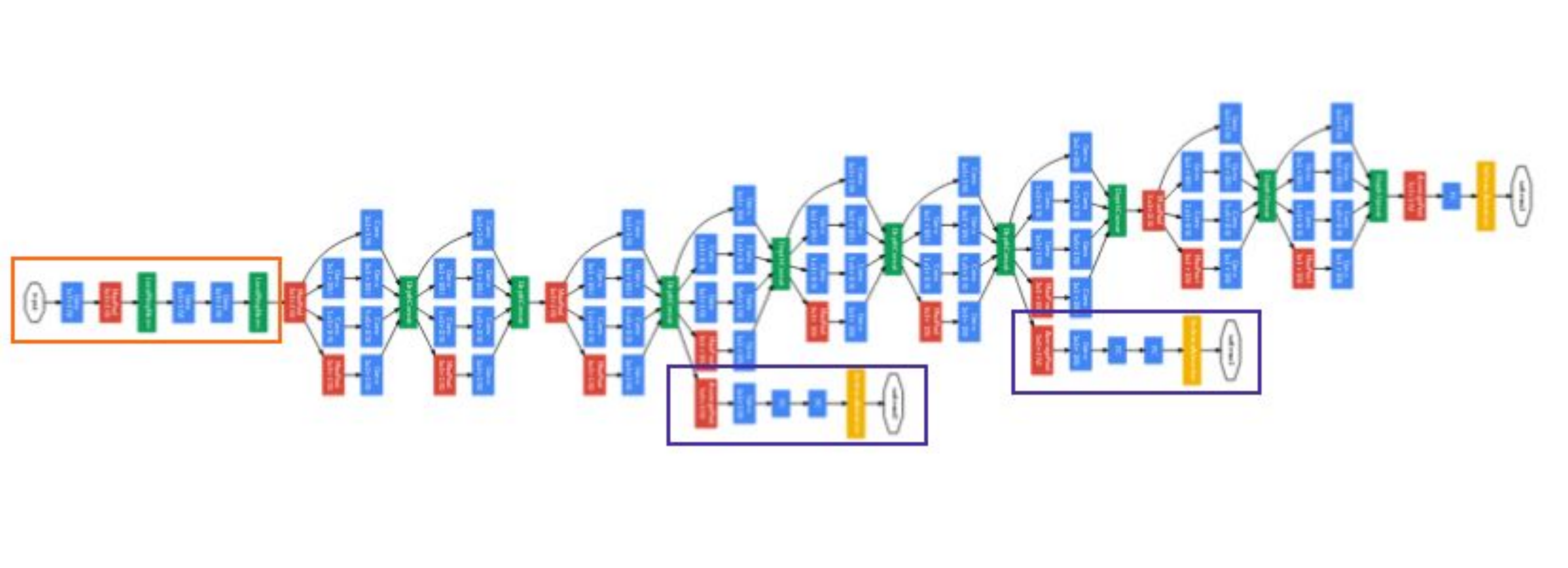
1,2 => make it wider instead!

Inception Module



(b) Inception module with dimension reductions

GoogLeNet



Auxiliary loss(0.3), global average pooling

Inception Results

Number of models	Number of Crops	Cost	Top-5 error	compared to base
1	1	1	10.07%	base
1	10	10	9.15%	-0.92%
1	144	144	7.89%	-2.18%
7	1	7	8.09%	-1.98%
7	10	70	7.62%	-2.45%
7	144	1008	6.67%	-3.45%

7 Models + 144 Crops

ResNet [5]

Problem(prev solutions: Init, Highway[6])

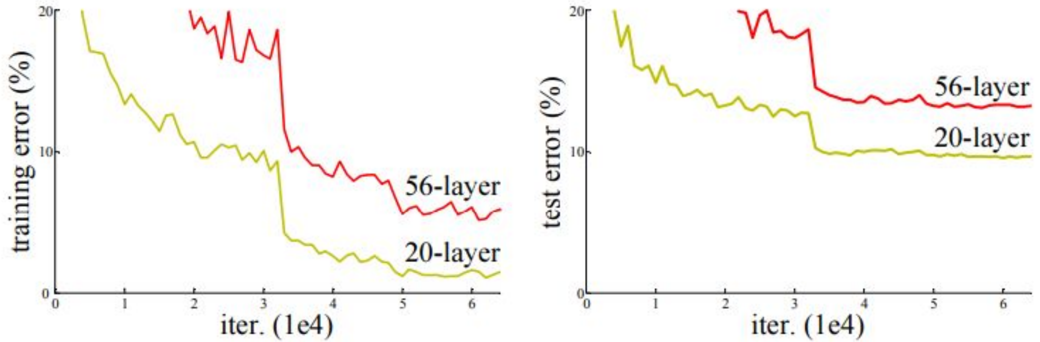


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

Residual Connection

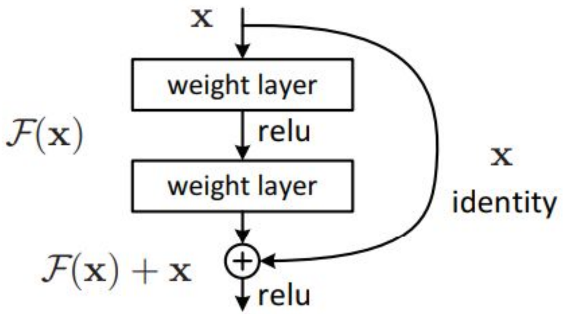
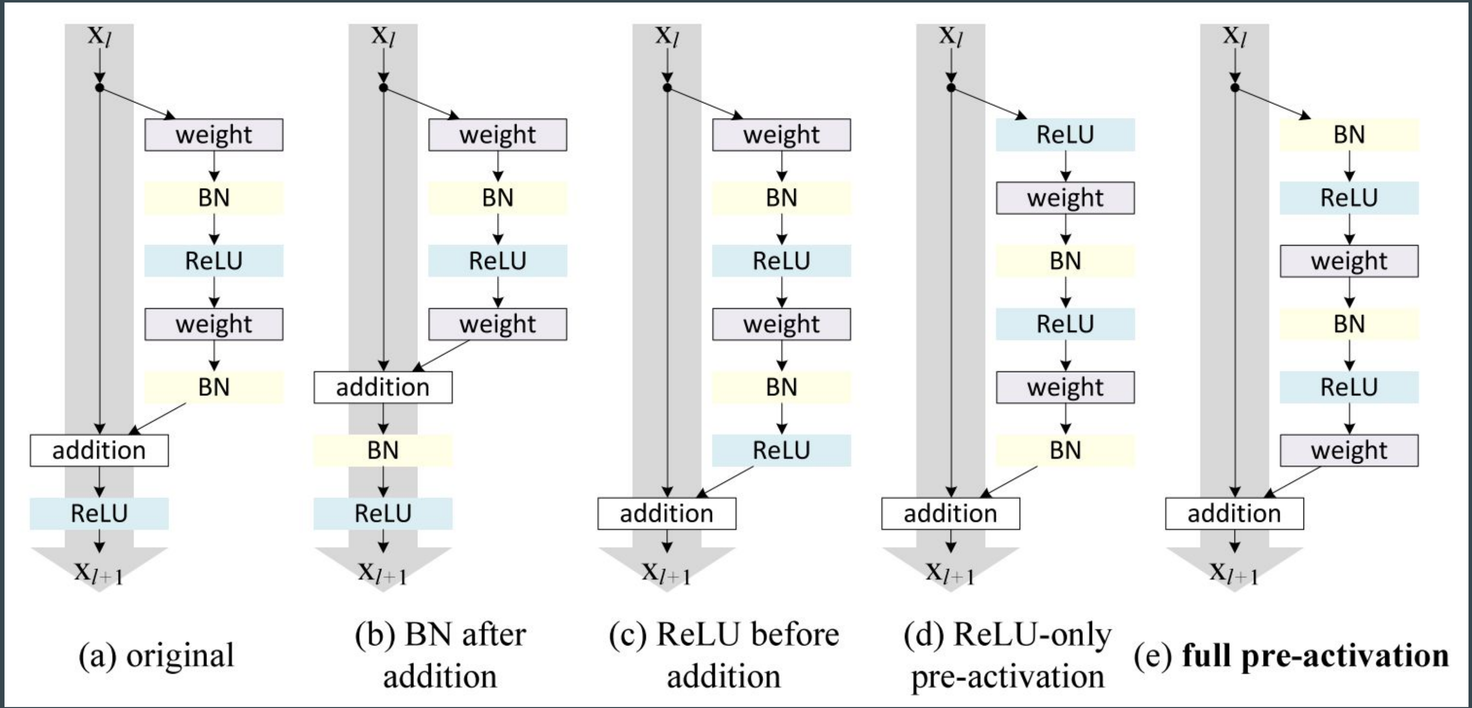


Figure 2. Residual learning: a building block.

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except † reported on the test set).

ResNet variants

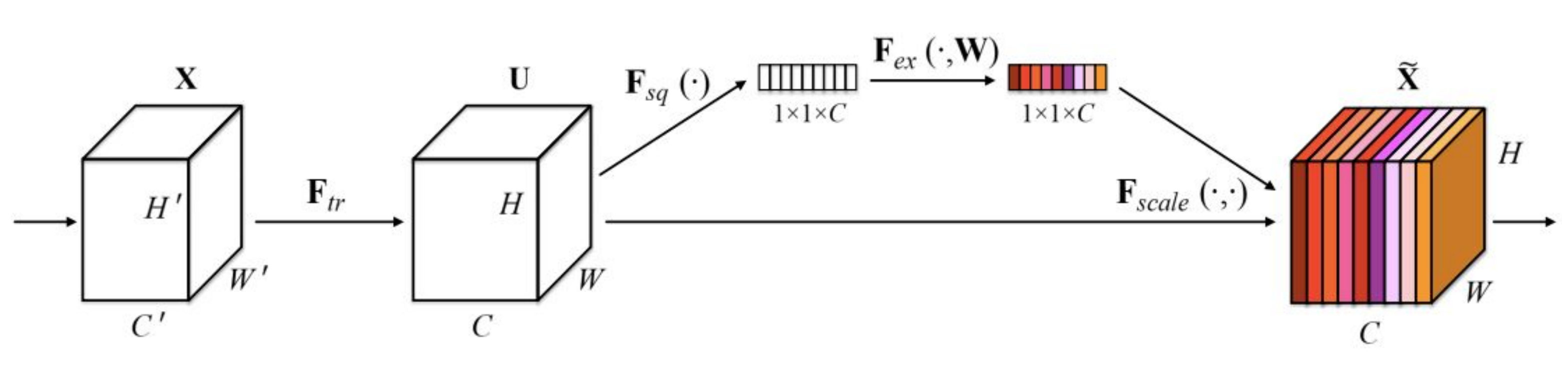


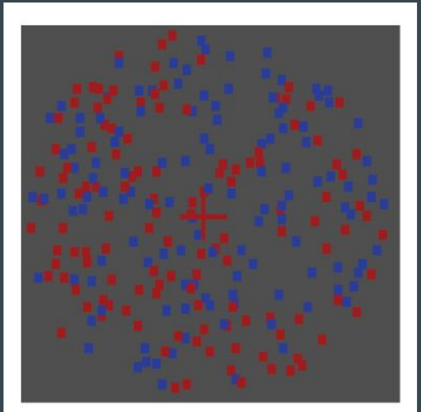
Squeeze & Excitation [7,8]

Global view

(like human

ch attention)





	original		re-implementation			SENet		
	top-1 err.	top-5 err.	top-1 err.	top-5 err.	GFLOPs	top-1 err.	top-5 err.	GFLOPs
ResNet-50 [9]	24.7	7.8	24.80	7.48	3.86	23.29$_{(1.51)}$	6.62$_{(0.86)}$	3.87
ResNet-101 [9]	23.6	7.1	23.17	6.52	7.58	22.38$_{(0.79)}$	6.07$_{(0.45)}$	7.60
ResNet-152 [9]	23.0	6.7	22.42	6.34	11.30	21.57$_{(0.85)}$	5.73$_{(0.61)}$	11.32
ResNeXt-50 [43]	22.2	-	22.11	5.90	4.24	21.10$_{(1.01)}$	5.49$_{(0.41)}$	4.25
ResNeXt-101 [43]	21.2	5.6	21.18	5.57	7.99	20.70$_{(0.48)}$	5.01$_{(0.56)}$	8.00
BN-Inception [14]	25.2	7.82	25.38	7.89	2.03	24.23$_{(1.15)}$	7.14$_{(0.75)}$	2.04
Inception-ResNet-v2 [38]	19.9†	4.9†	20.37	5.21	11.75	19.80$_{(0.57)}$	4.79$_{(0.42)}$	11.76

Future Directions

Generalizability is important => weakly supervised methods[9]

Training time is important => 224 seconds[10]

Computational budget is important => mobile net[11]

References

1. Gradient-Based Learning Applied to Document Recognition
2. ImageNet Classification with Deep Convolutional Neural Networks
3. Dropout: A Simple Way to Prevent Neural Networks from Overfitting
4. Going Deeper with Convolutions
5. Deep Residual Learning for Image Recognition
6. Highway Networks
7. Squeeze-and-Excitation Networks
8. Feature-selective attention enhances color signals in early visual areas of the human brain
9. Exploring the Limits of Weakly Supervised Pretraining
10. ImageNet/ResNet-50 Training in 224 Seconds
11. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications