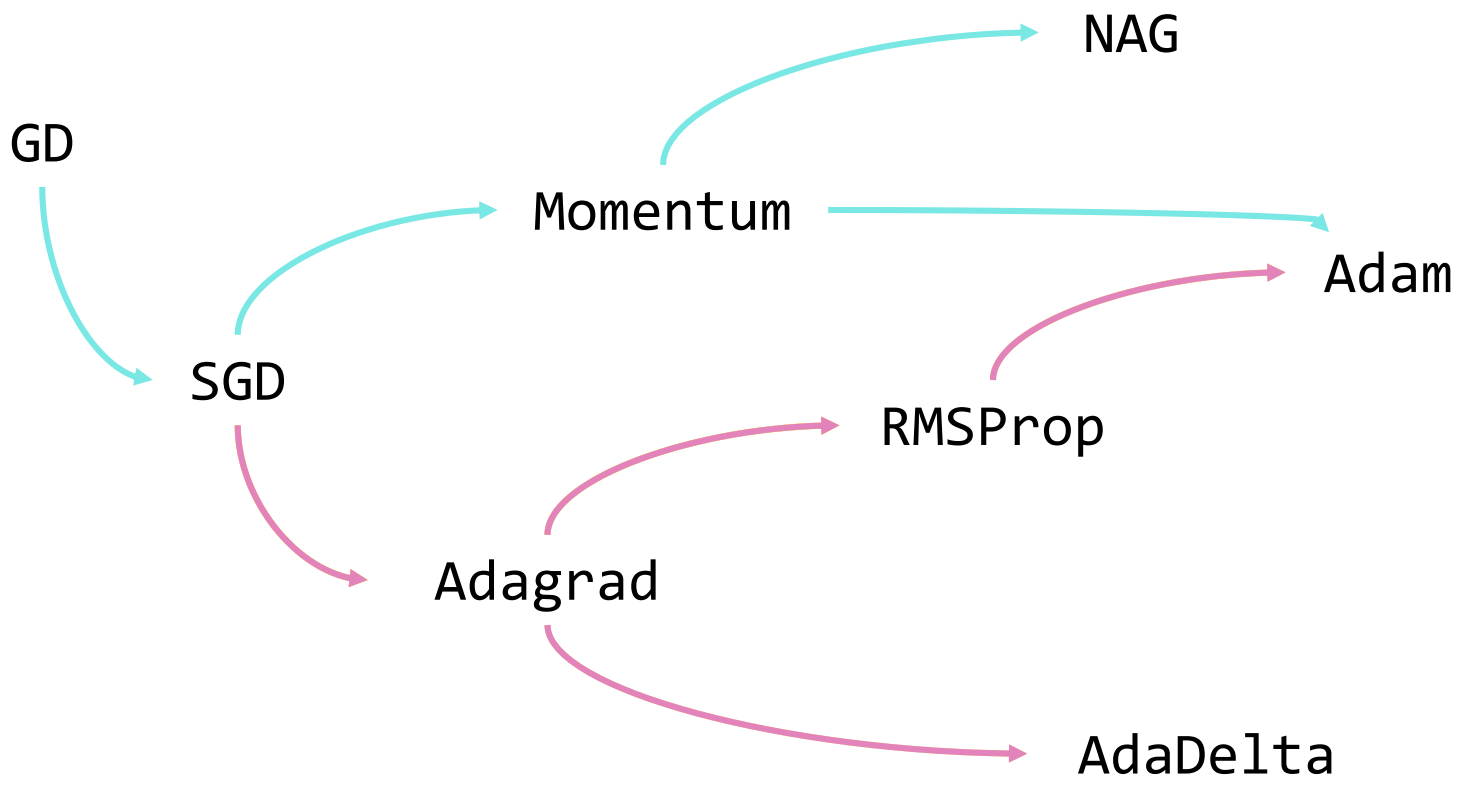
Lecture 9.
Various CNN Architectures

Review

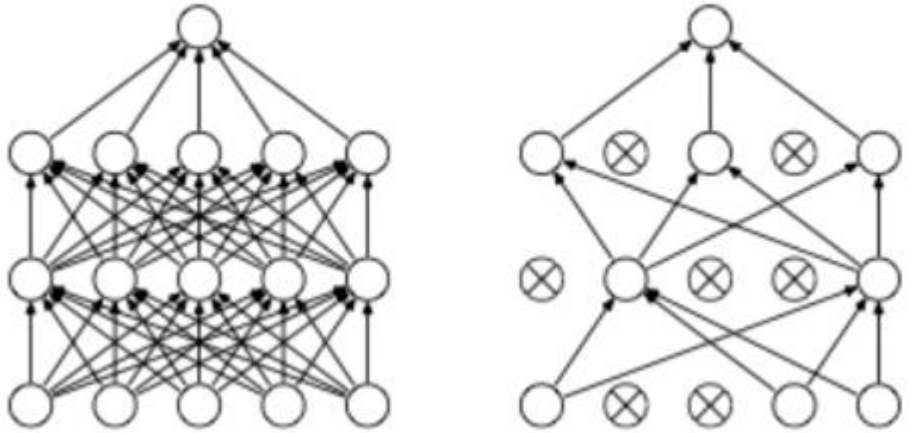
1. Gradient Descent
2. Stochastic Gradient Descent (SGD)
3. Momentum
4. Nesterov Accelerated Gradient (NAG)
5. Adagrad
6. RMSProp
7. AdaDelta
8. Adam

Review



Review

Dropout: A Simple Way to Prevent Neural Networks from Overfitting [Srivastava et al. 2014]

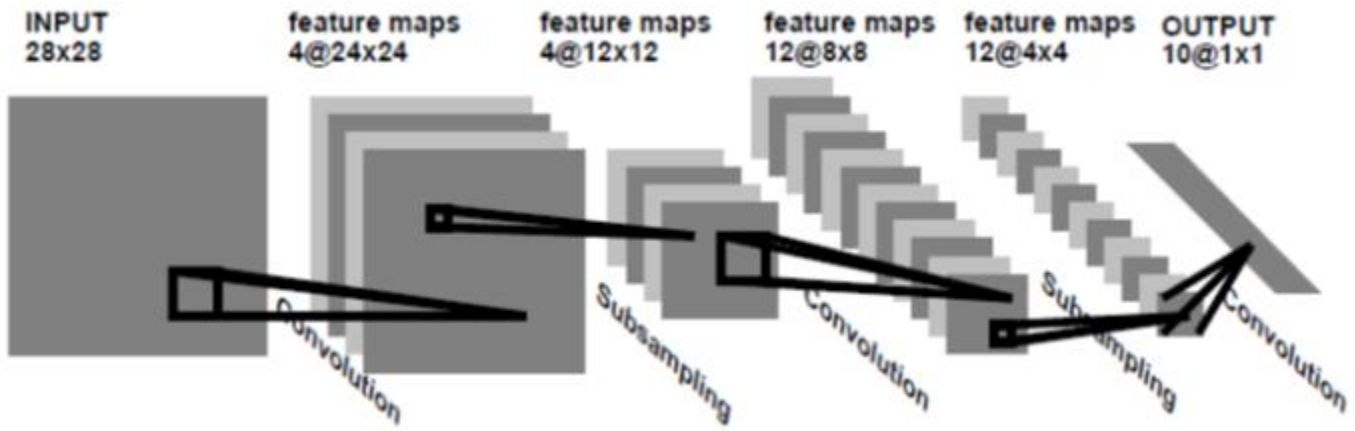


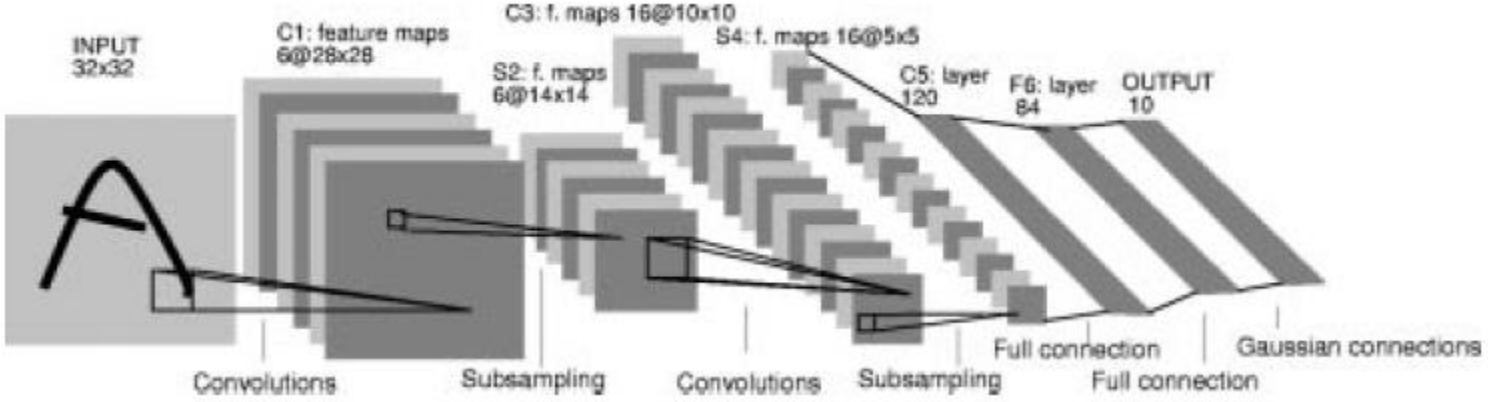
(a) Standard Neural Net (b) After applying dropout.

CNN Architectures

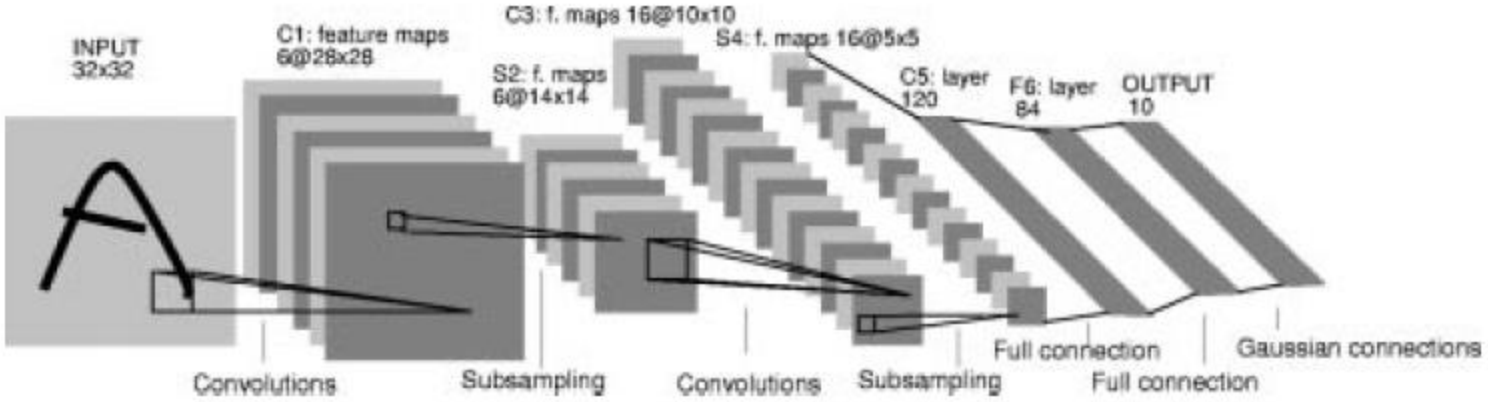
1. LeNet (AlexNet)
2. VGGNet
3. GoogLeNet
4. ResNetb

LeNet(1990 – 1998)

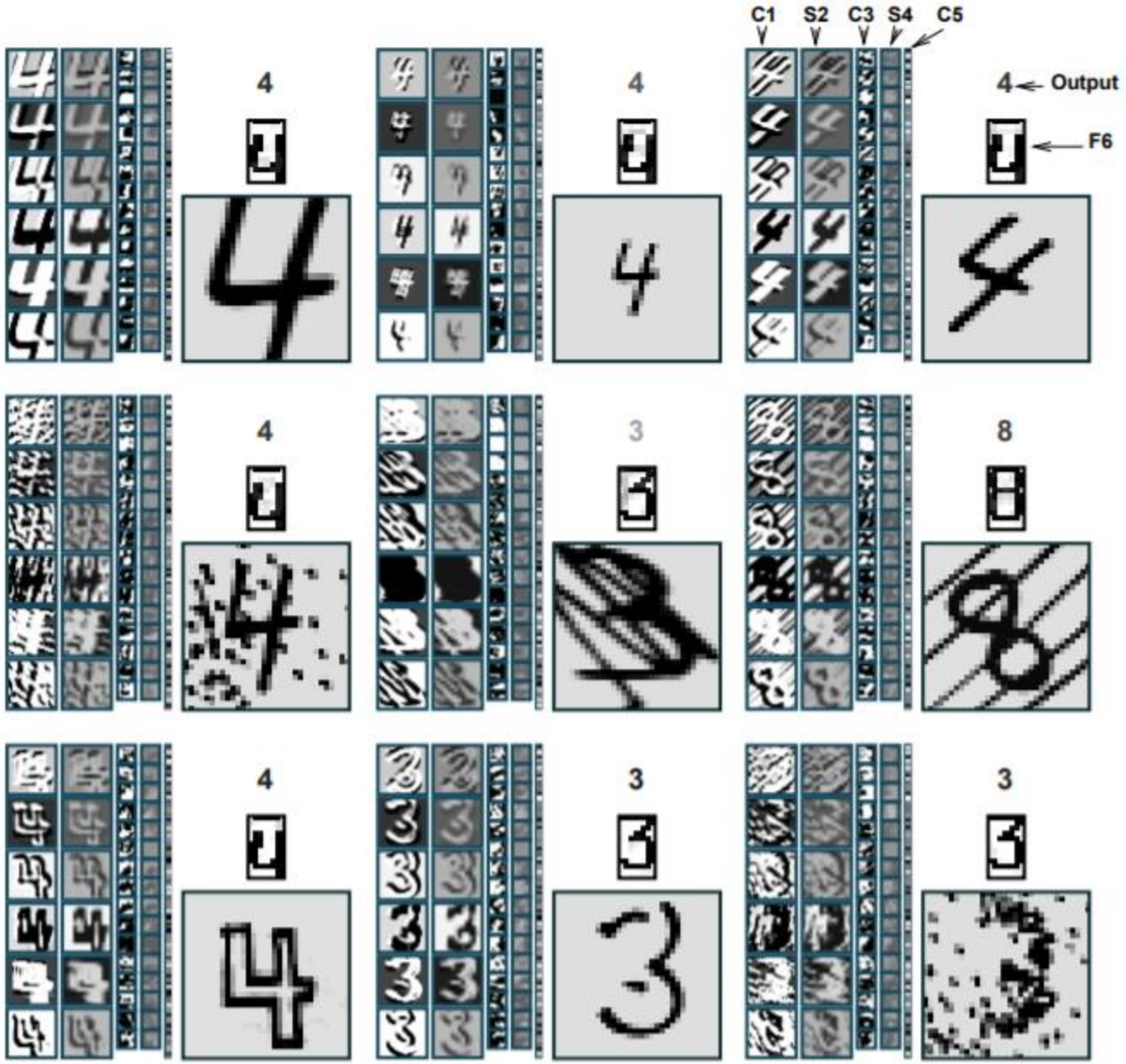




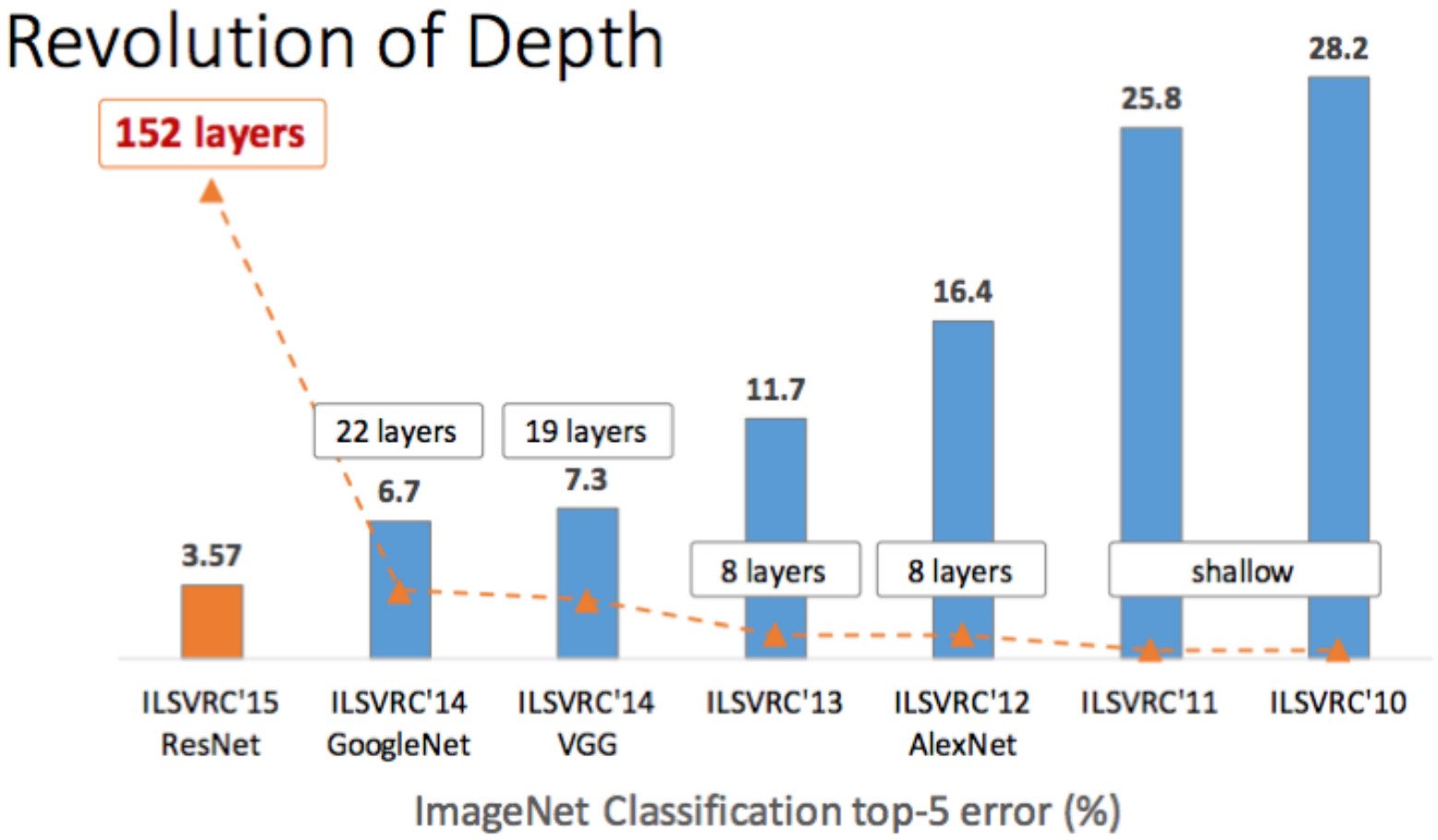
LeNet 5(1998)



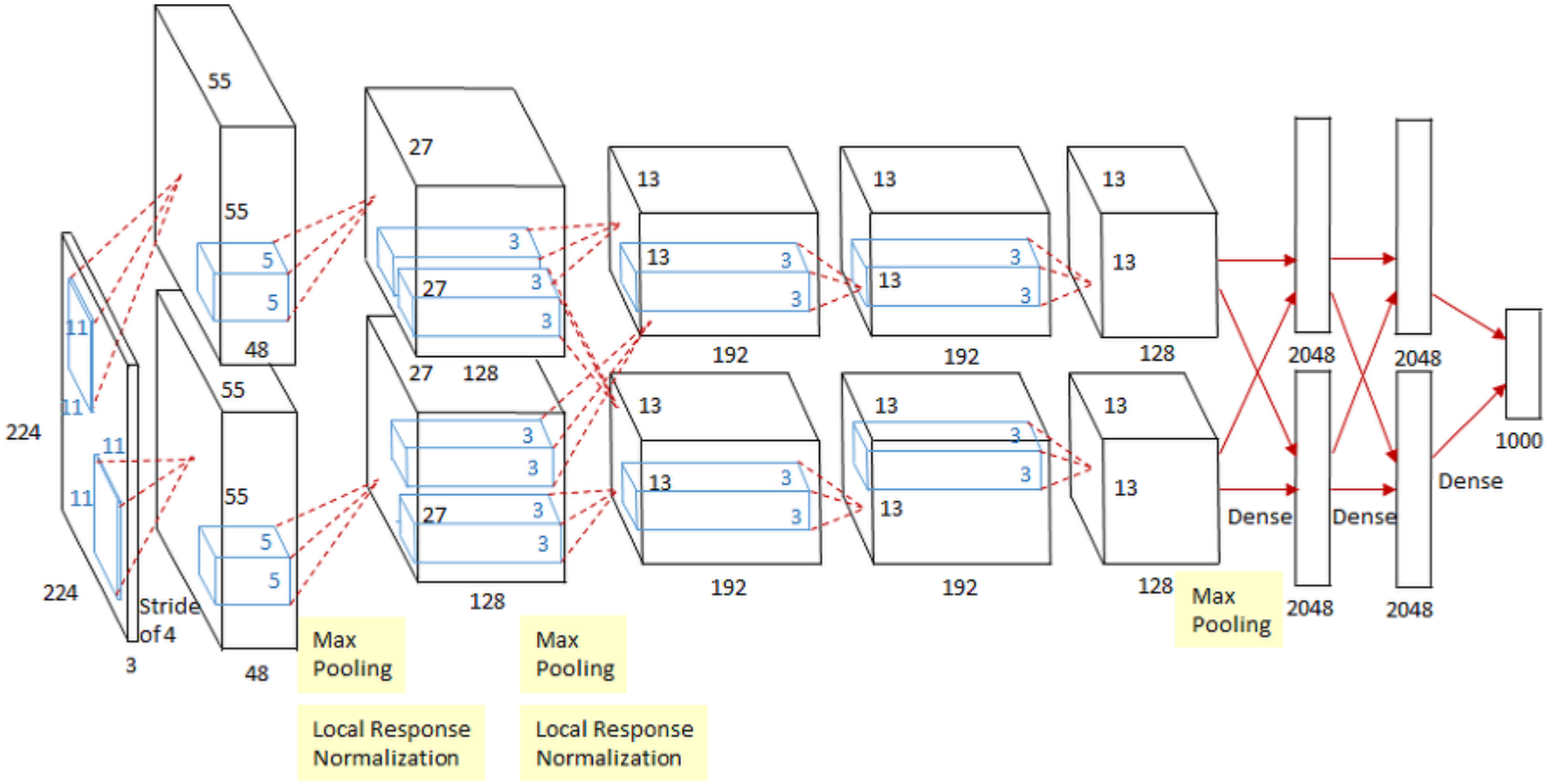
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X


C1
S2
C3
S4
C5
4
4
4
4 Output
F6
4
3
8
4
3
3

CNN Architectures

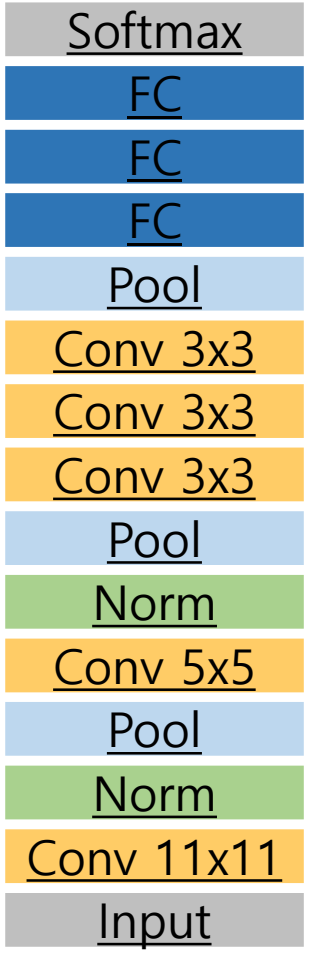


AlexNet(2012)

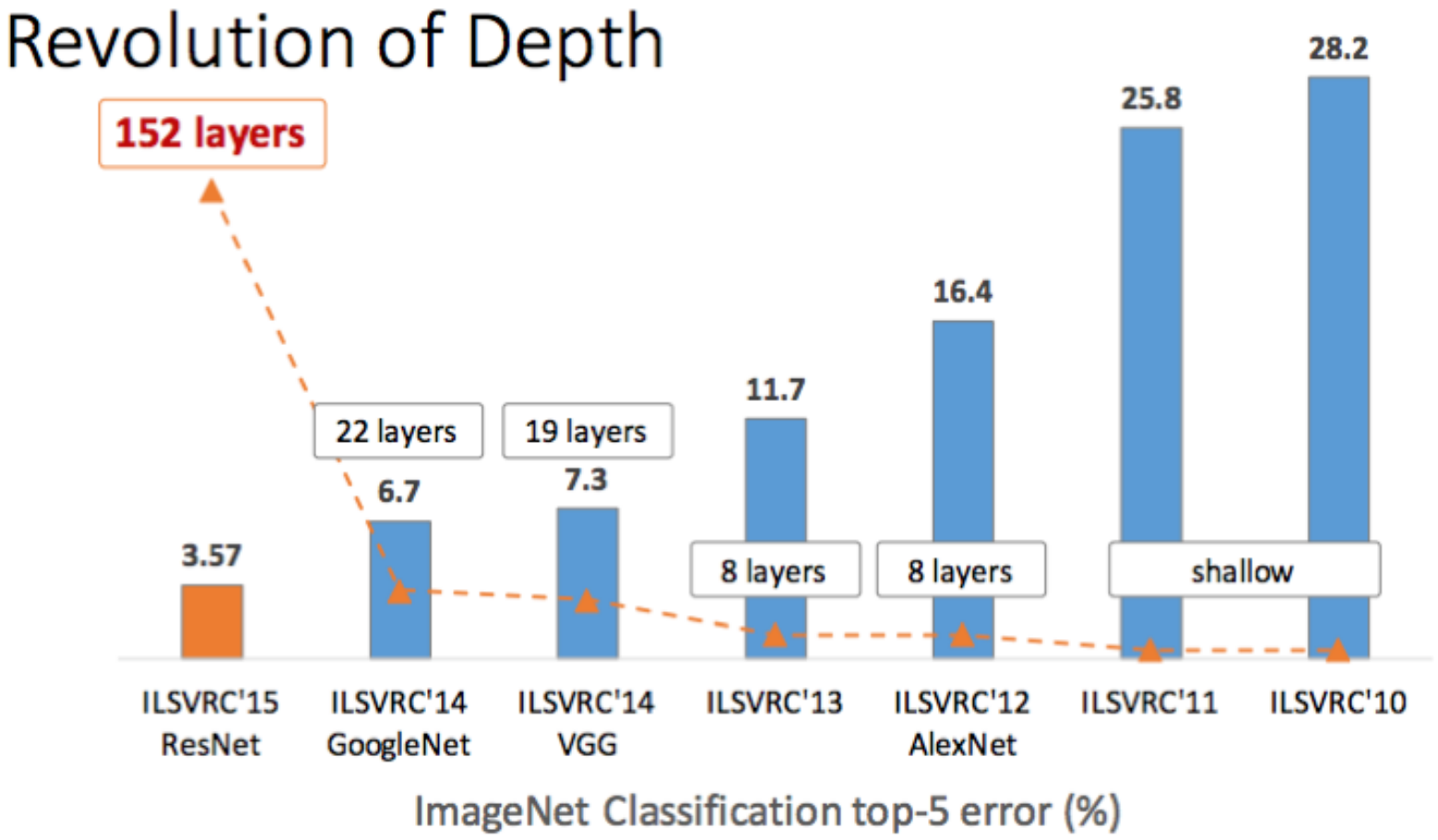


AlexNet(2012)

1. GPU 2개를 이용한 특이한 구조
2. ReLU 처음으로 사용한 모델
3. 현재는 사용하지 않는 Normalization layer를 사용
4. Data Augmentation / Dropout 으로 overfitting 방지
5. SGD + Momentum (0.9)



CNN Architectures

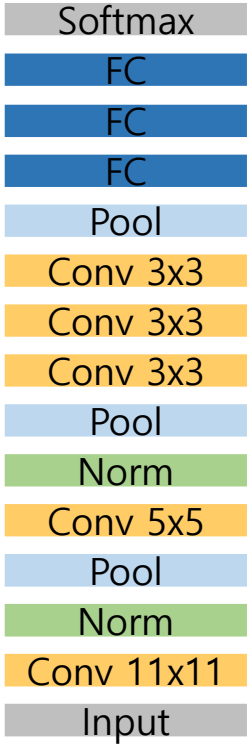


VGGNet

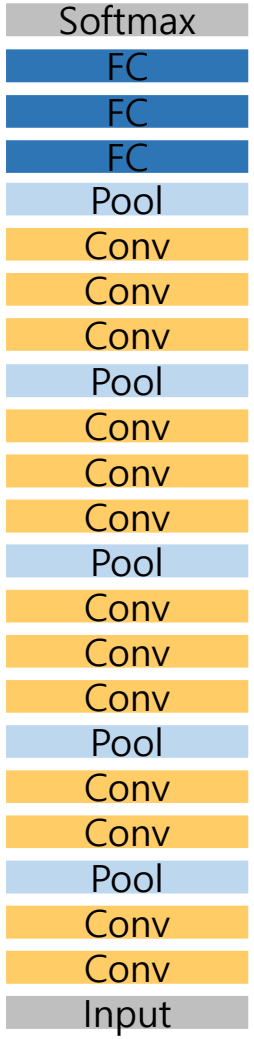
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

VGGNet

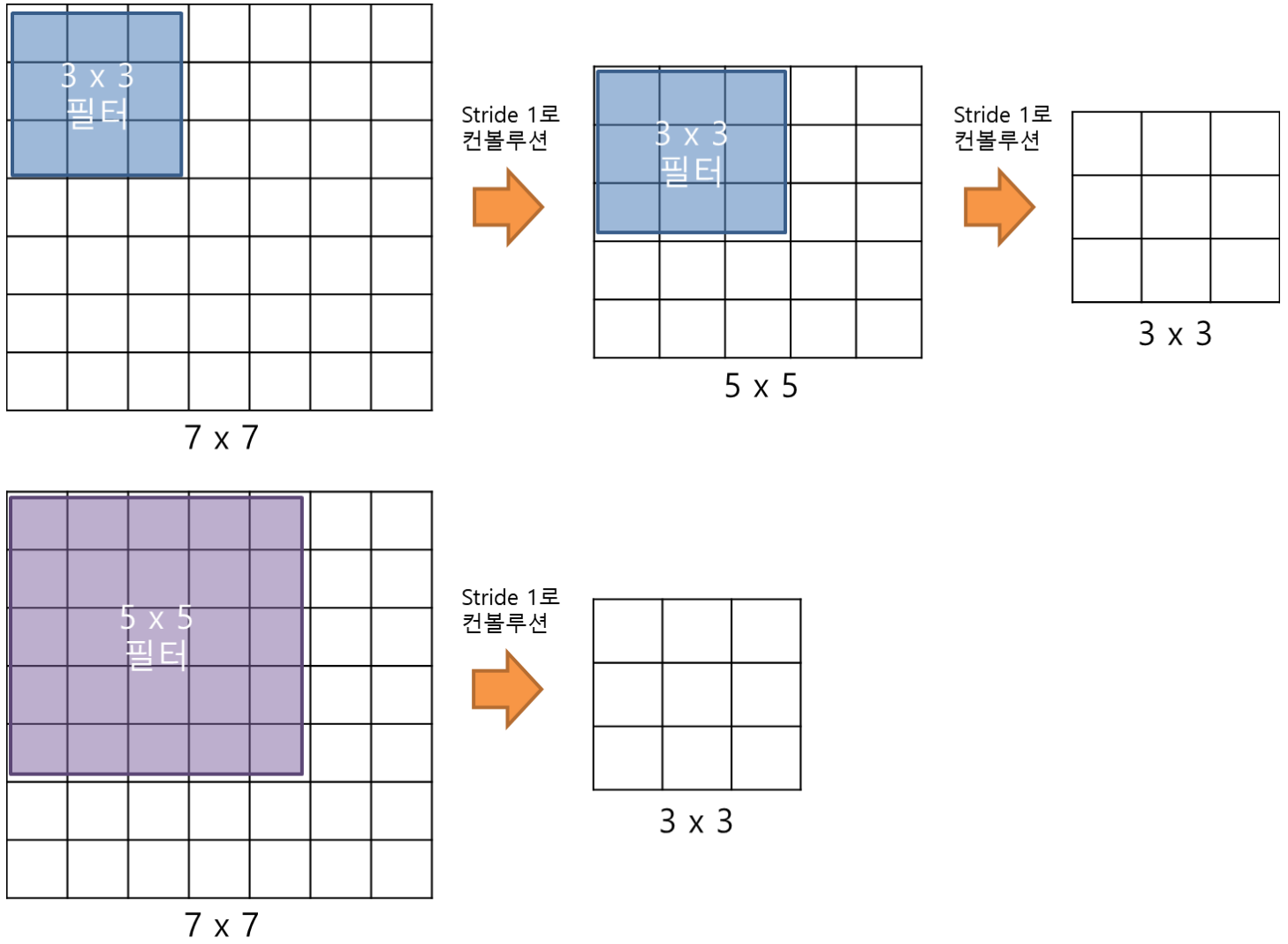
AlexNet



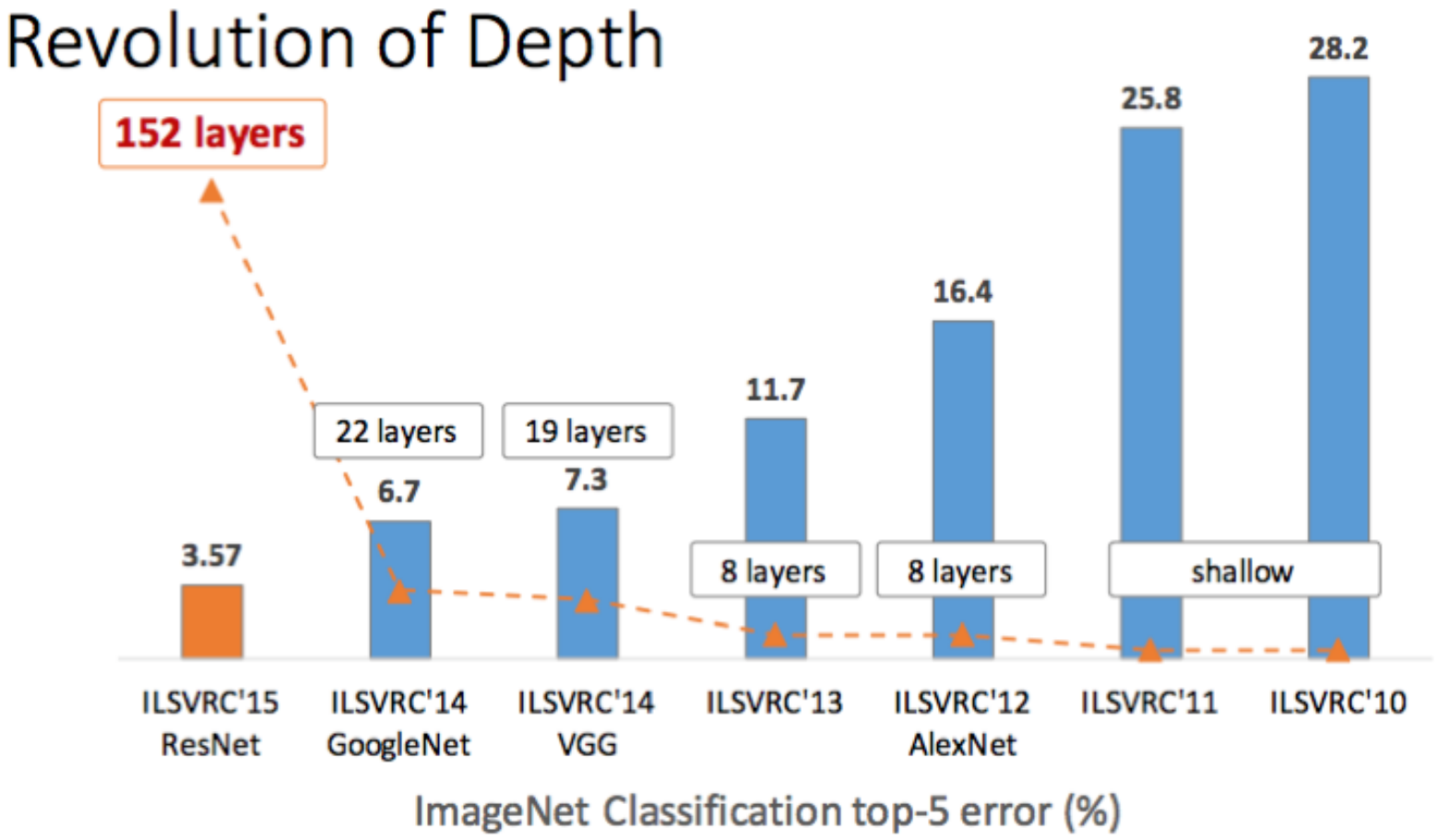
VGGNet16



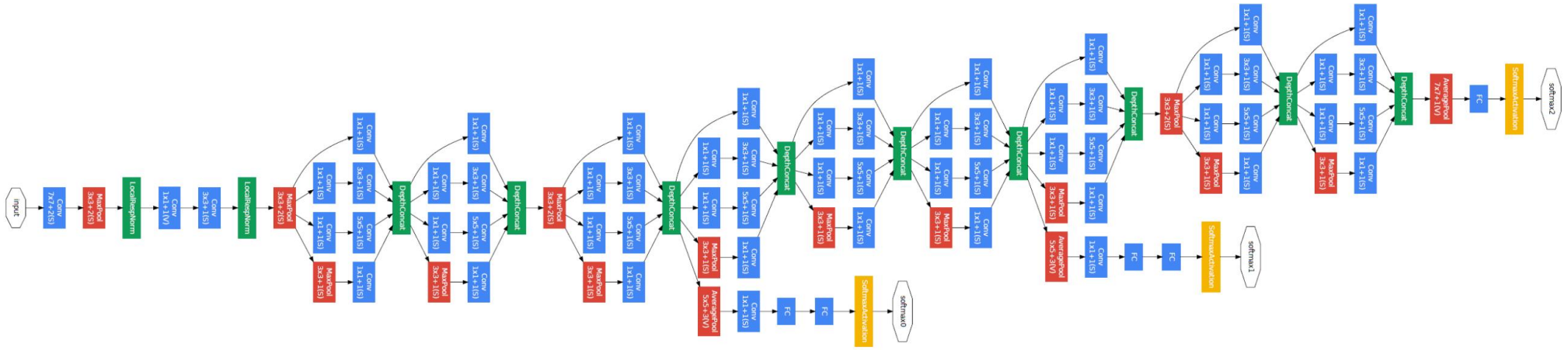
VGGNet


3 x 3
필터
7 x 7
Stride 1로
컨볼루션
3 x 3
필터
5 x 5
Stride 1로
컨볼루션
3 x 3
5 x 5
필터
7 x 7
Stride 1로
컨볼루션
3 x 3

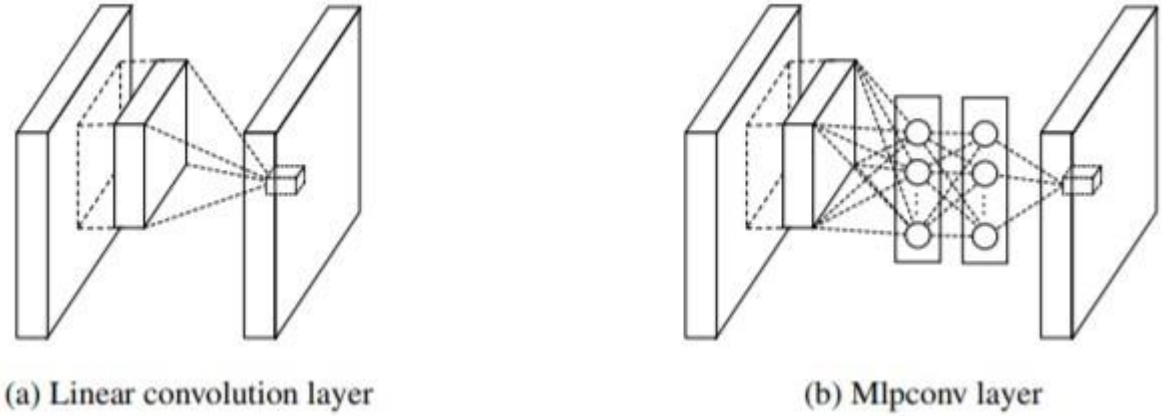
CNN Architectures

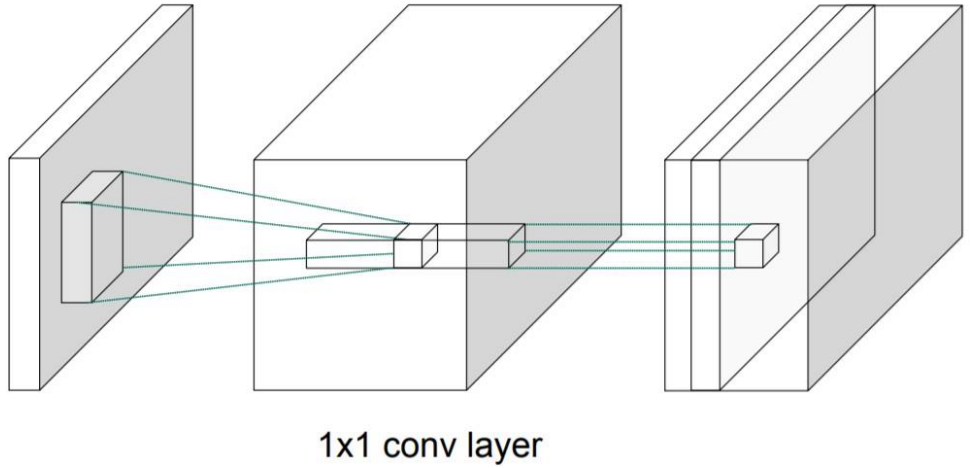


GoogLeNet

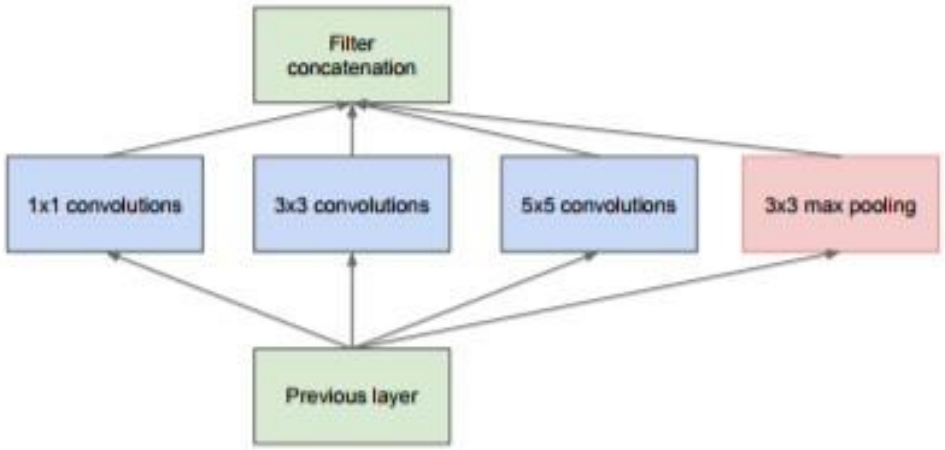

input
Conv 7x7+2(S)
MaxPool 3x3+2(S)
LocalRespNorm
Conv 1x1+1(V)
Conv 3x3+1(S)
LocalRespNorm
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
AveragePool 5x5+3(V)
FC
SoftmaxActivation
softmax0
softmax1
softmax2
AveragePool 7x7+1(V)

* Network in Network (NIN)

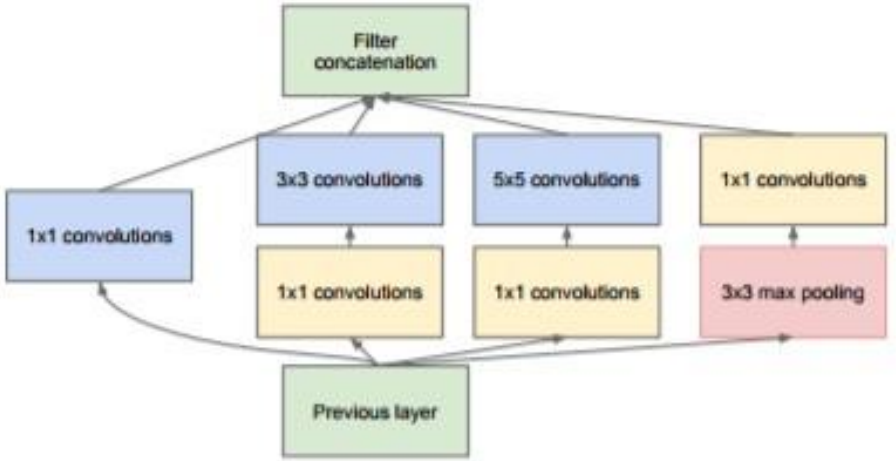




GoogLeNet

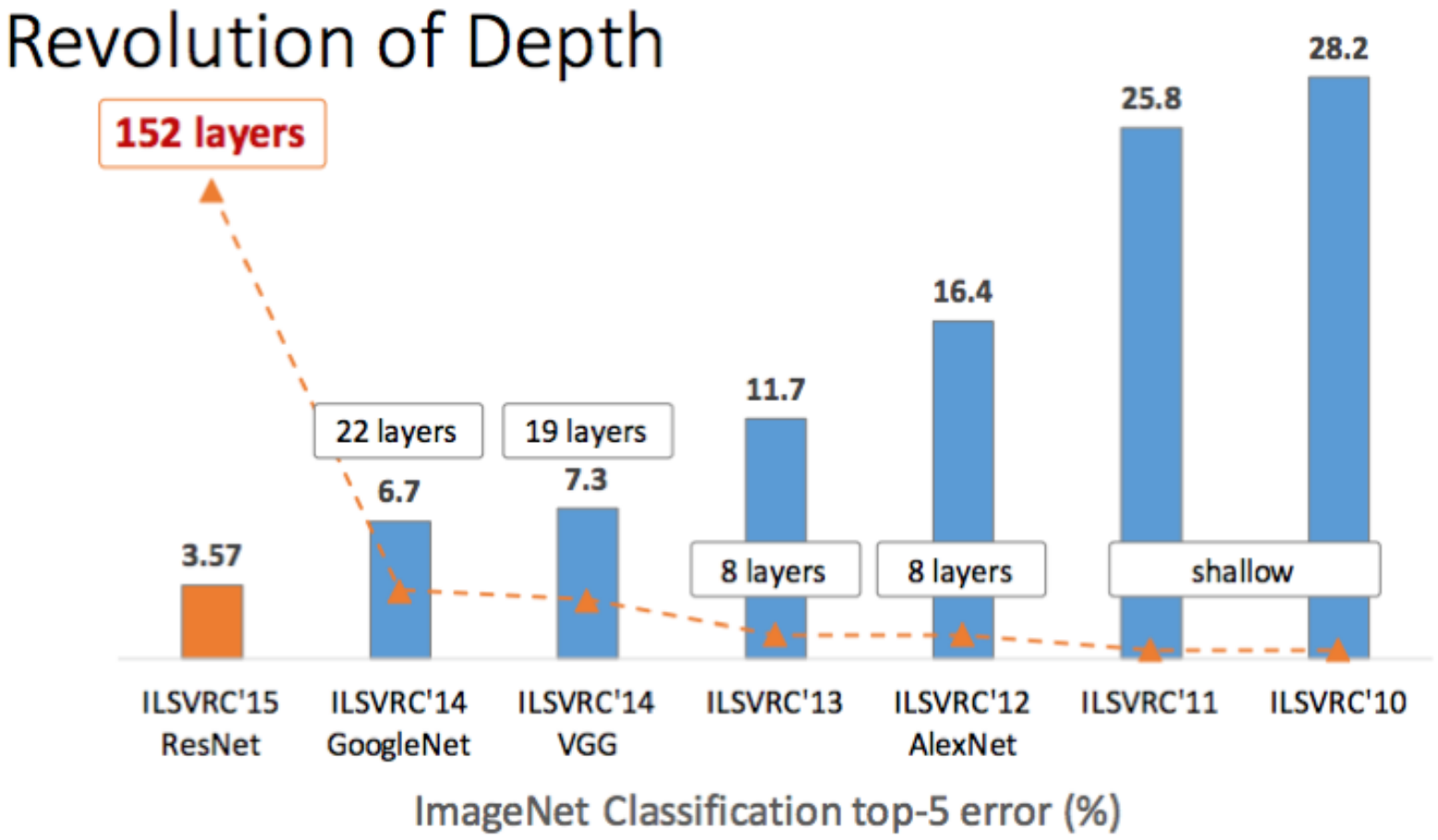


(a) Inception module, naïve version

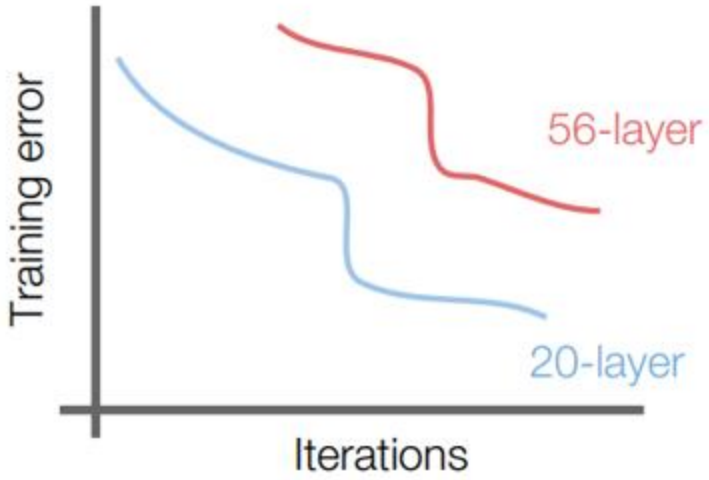


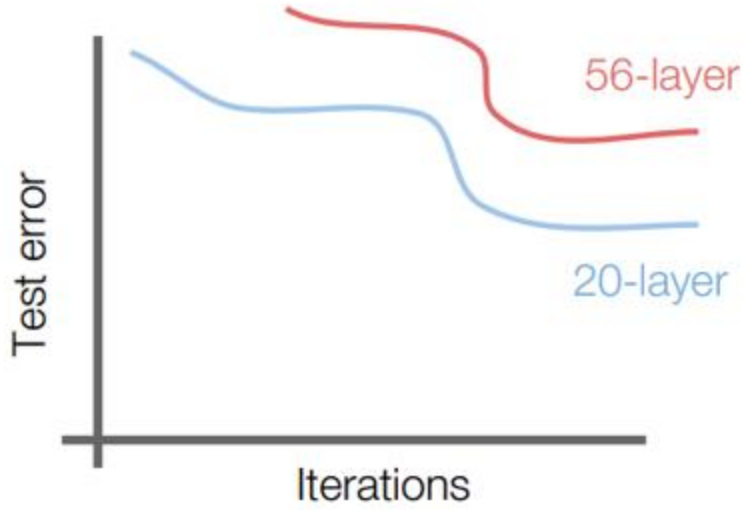
(b) Inception module with dimension reductions

CNN Architectures

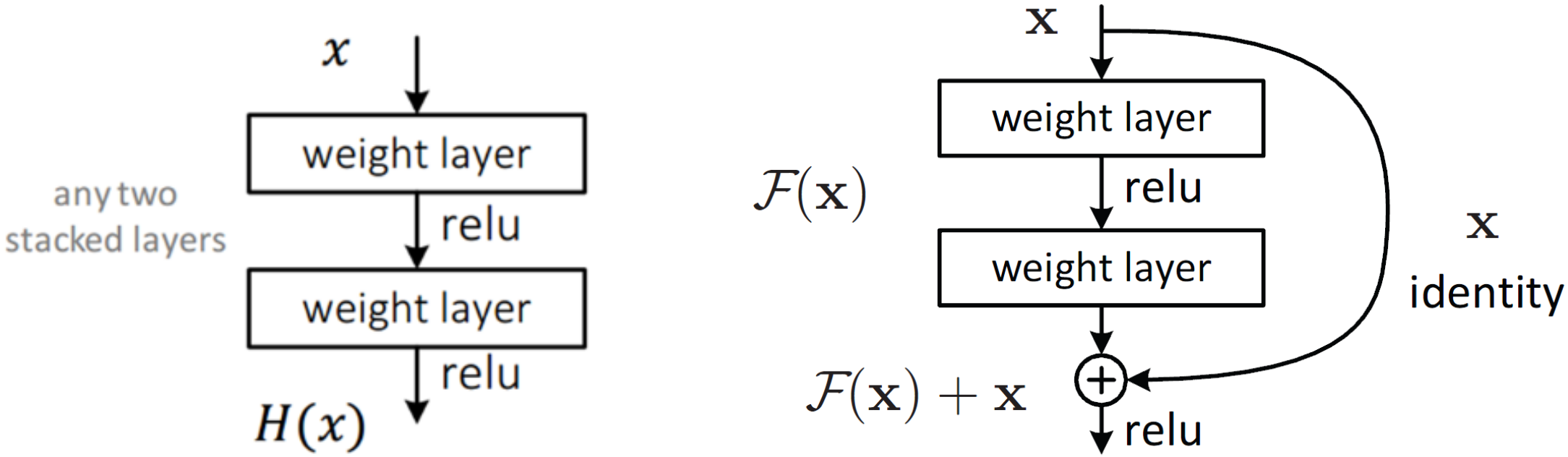


ResNet


Training error
56-layer
20-layer
Iterations

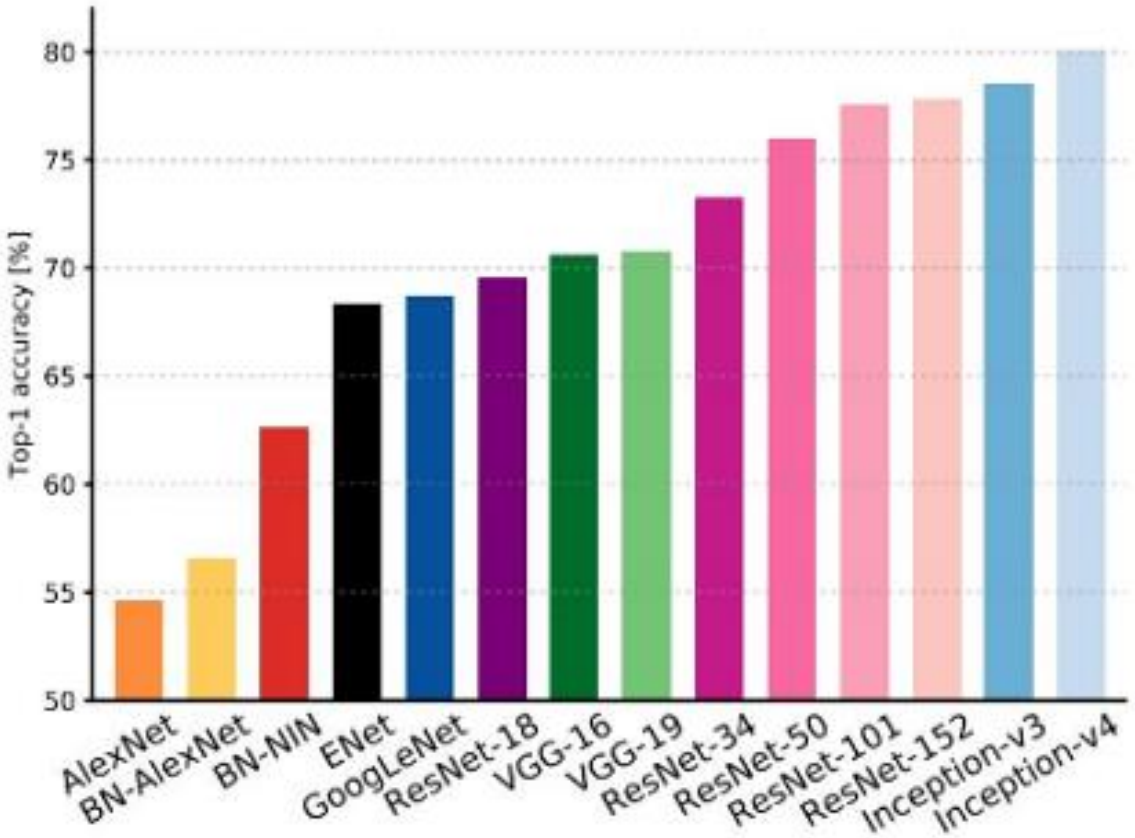

Test error
56-layer
20-layer
Iterations

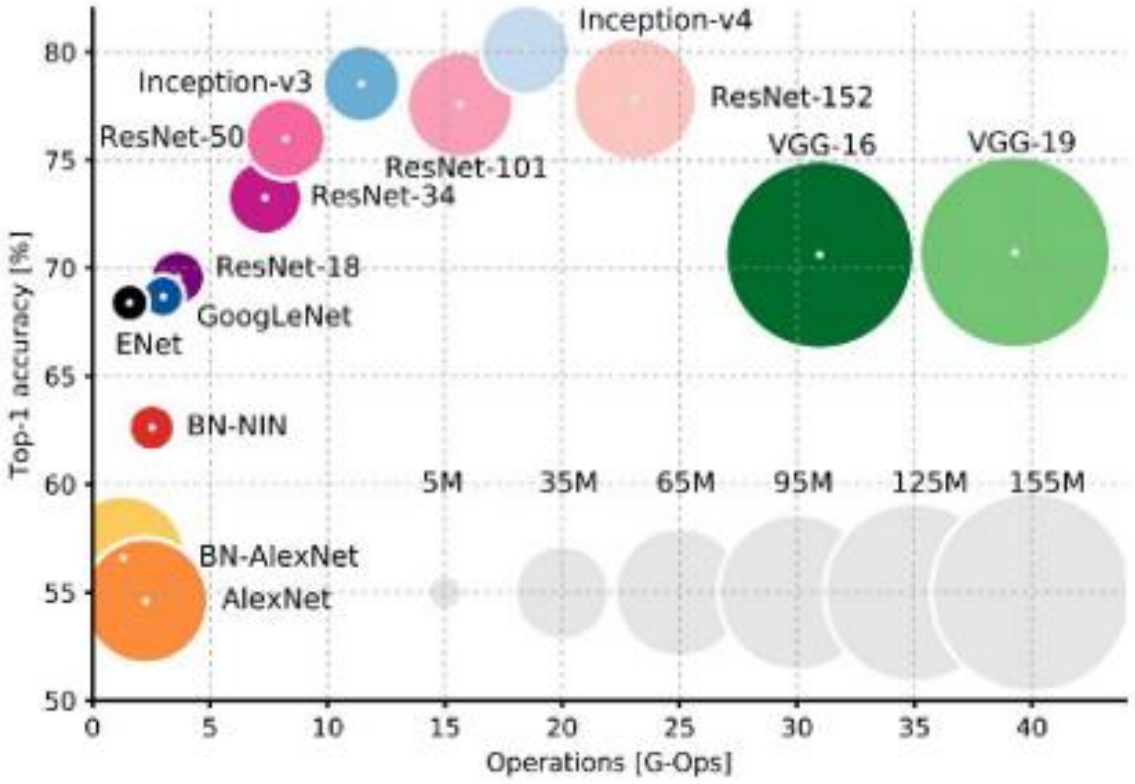
ResNet


any two
stacked layers
x
weight layer
relu
weight layer
relu
H(x)
x
weight layer
F(x)
relu
weight layer
x
identity
F(x) + x
relu

ResNet

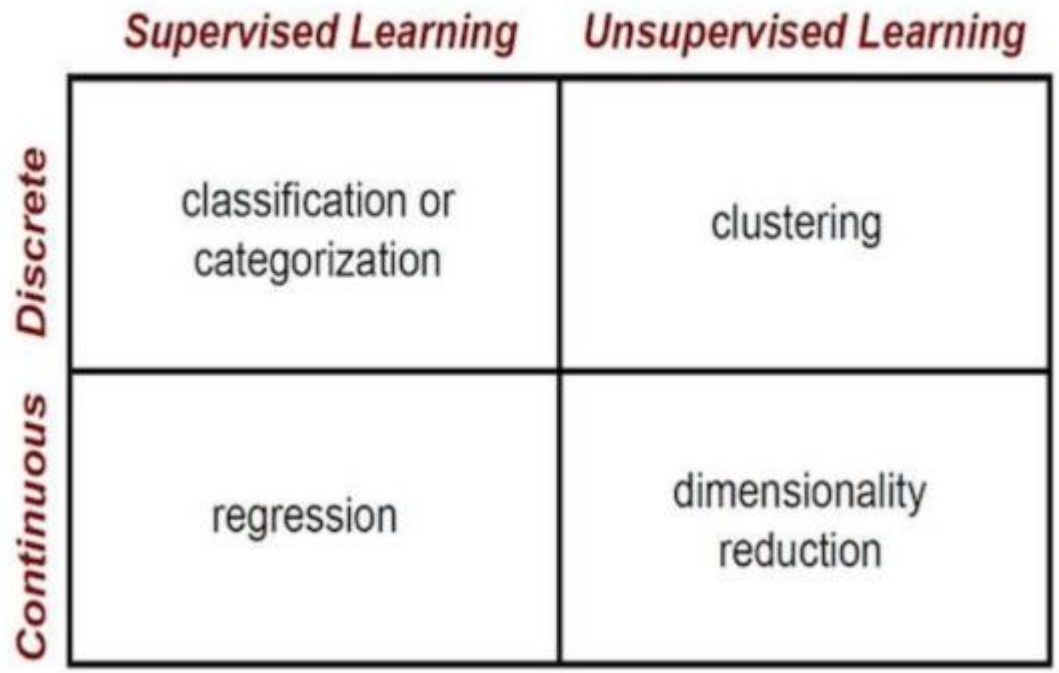
method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**



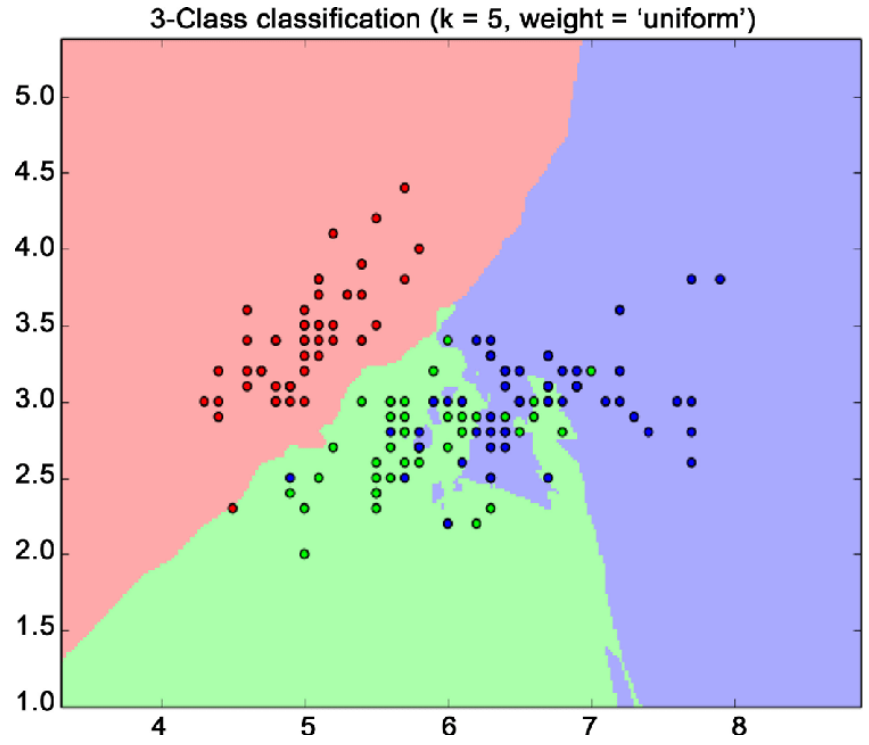


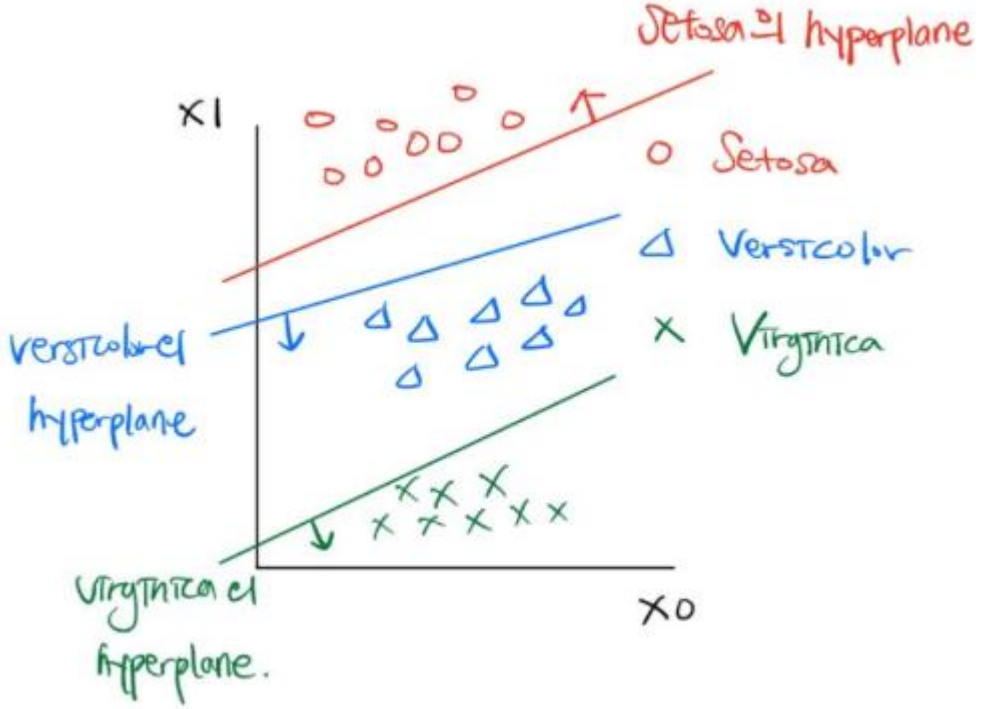
An Analysis of Deep Neural Network Models for Practical Applications, 2017.

9회차 동안...

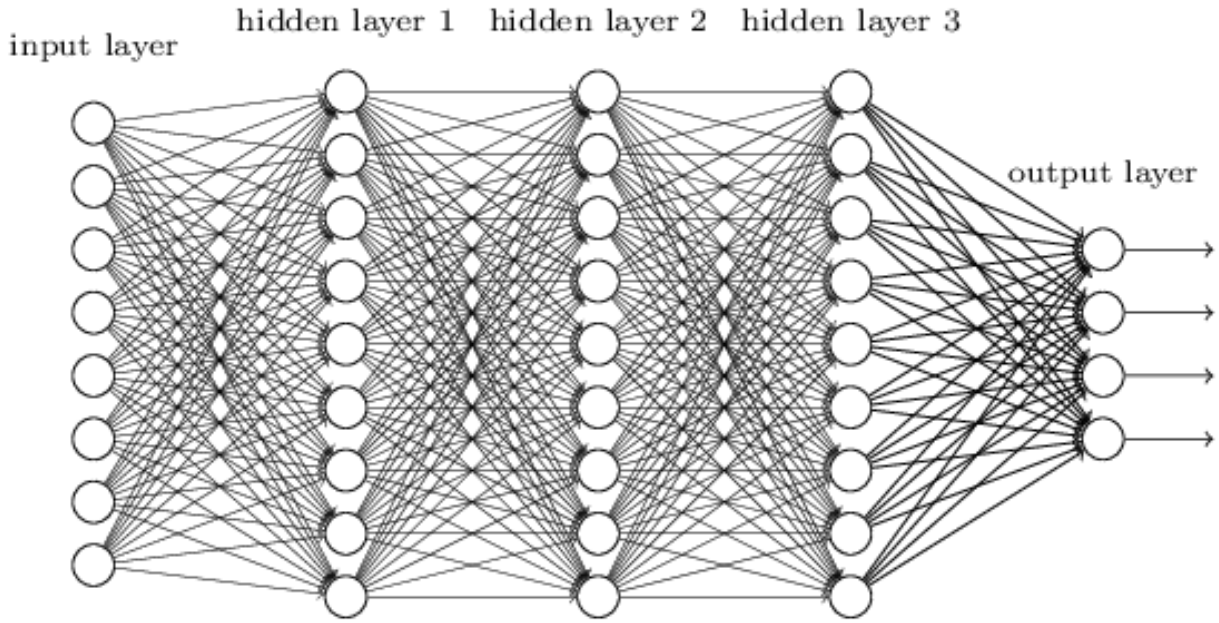

Supervised Learning
Unsupervised Learning
Discrete
Continuous
classification or categorization
clustering
regression
dimensionality reduction

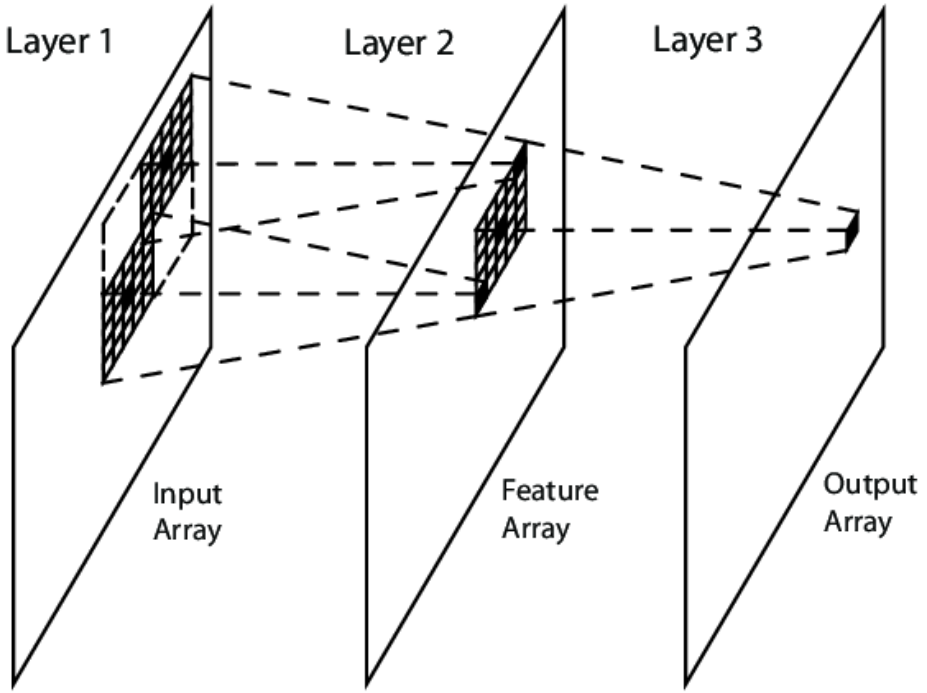
9회차 동안...


3-Class classification (k = 5, weight = 'uniform')
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
4
5
6
7
8


Setosa의 hyperplane
x1
Setosa
Versicolor
Virginica
Versicolor의
hyperplane
Virginica의
hyperplane.
x0

9회차 동안...


input layer
hidden layer 1
hidden layer 2
hidden layer 3
output layer

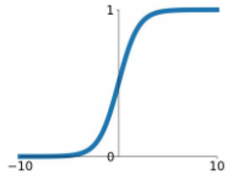

Layer 1
Layer 2
Layer 3
Input Array
Feature Array
Output Array

9회차 동안...

Activation Function

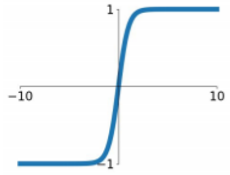
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



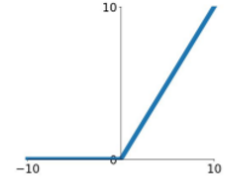
tanh

$\tanh(x)$



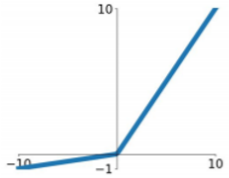
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

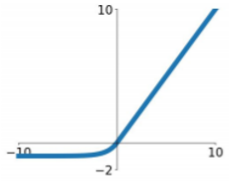


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

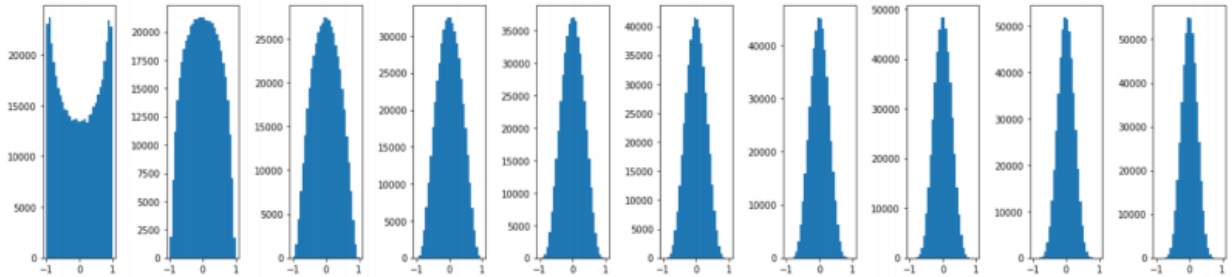
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Xavier initialize

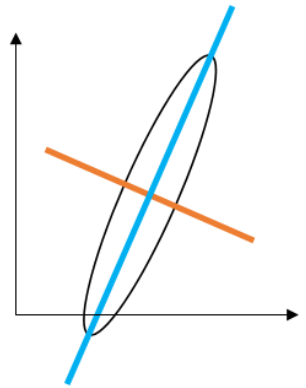
$np.random.randn(size = (D, H))/np.sqrt(D)$

```
W = np. random.randn(fan_in, fan_out) / np.sqrt(fan_in)
```

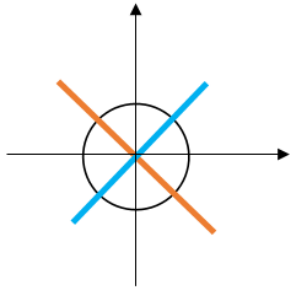


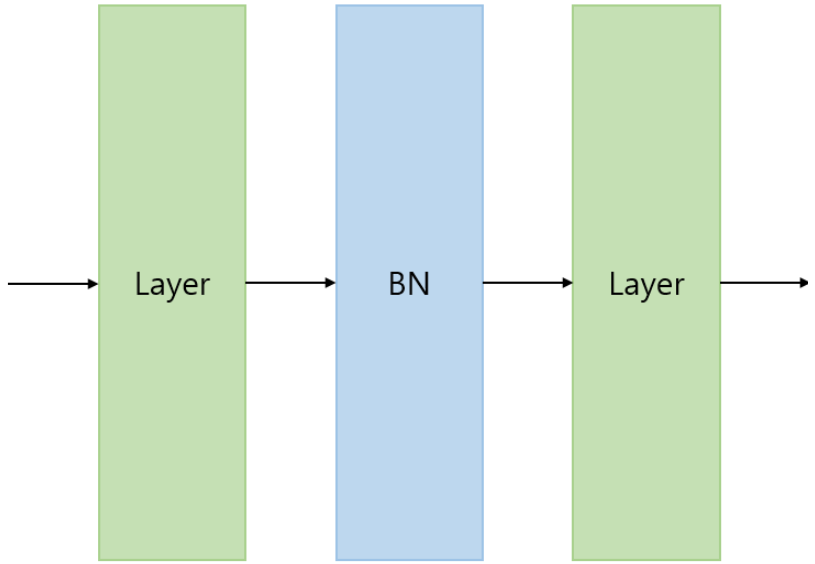
9회차 동안...

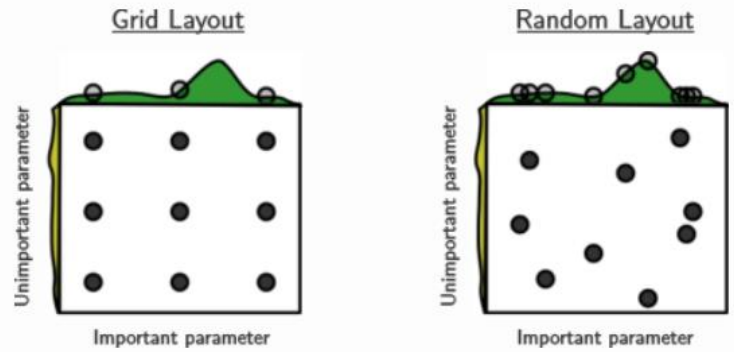
Before normalization



After normalization







Grid Search 결과와 Random Search 결과 비교 예시

[Bergstra and Bengio(2012)]

9회차 동안...

Crop

Flip

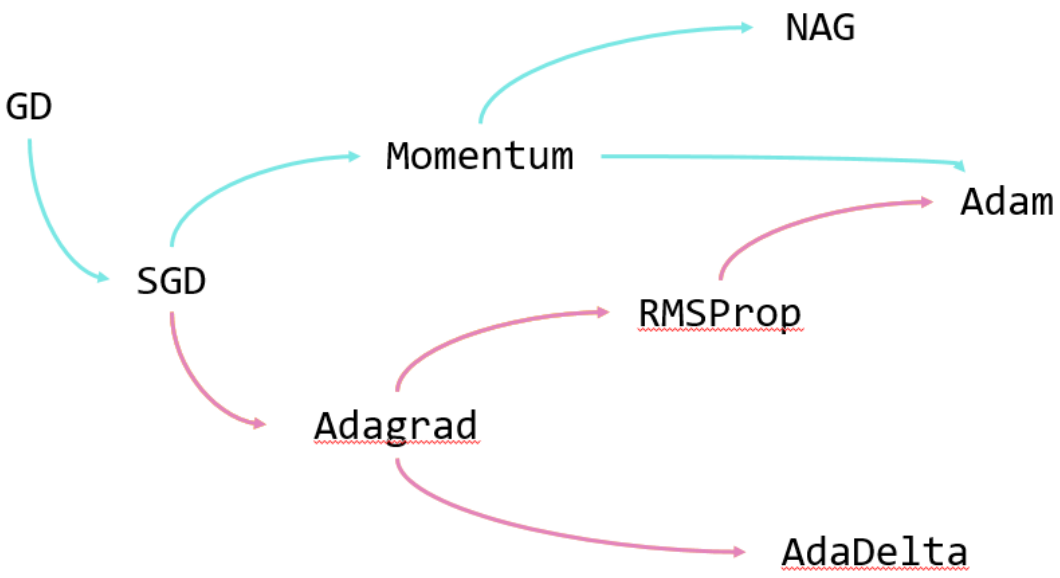
Noise

Scale

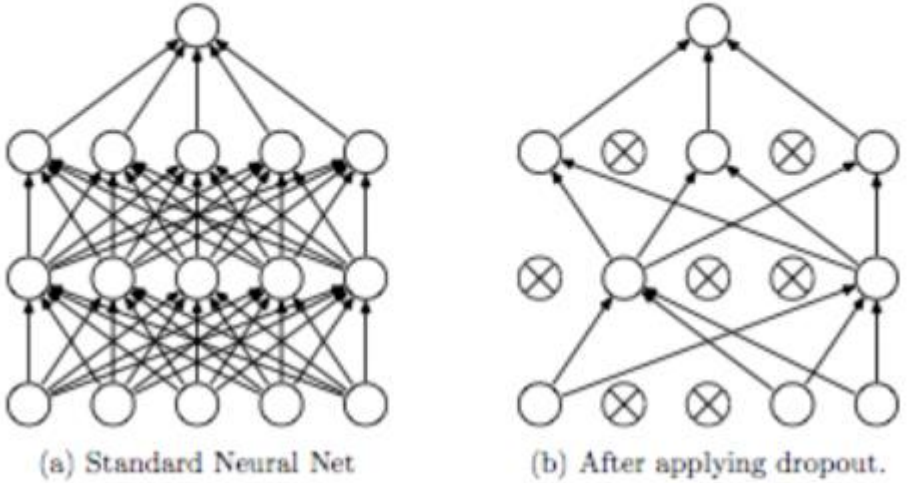
Rotate

Translation

9회차 동안...



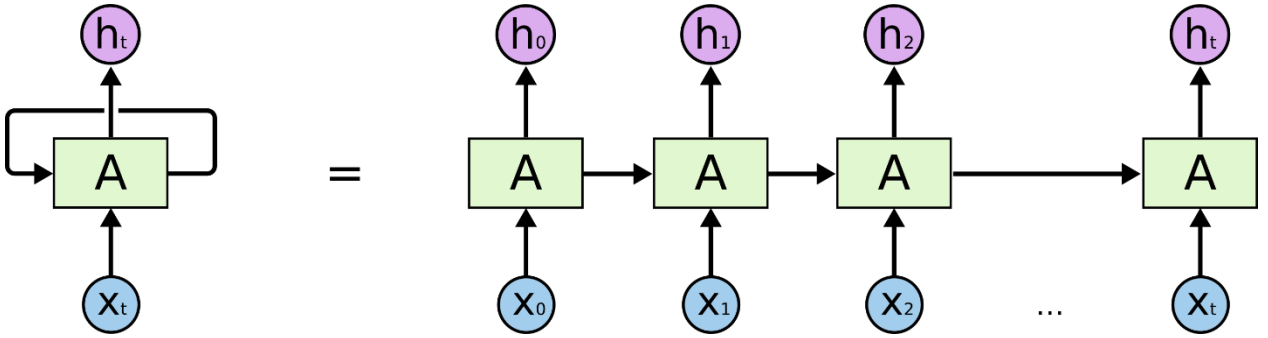
Dropout: A Simple Way to Prevent Neural Networks from Overfitting [Srivastava et al. 2014]



(a) Standard Neural Net (b) After applying dropout.

Preview on Next Ways

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction



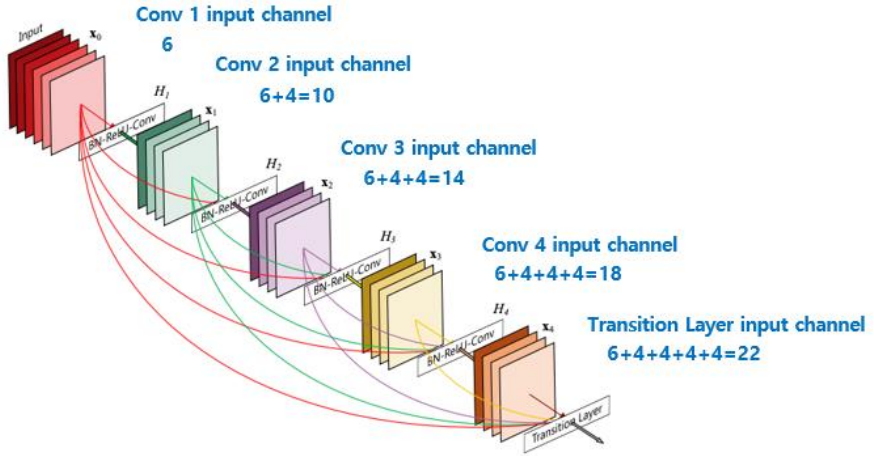


Figure 1: A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.