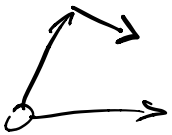
Lecture 9.
Various CNN Architectures

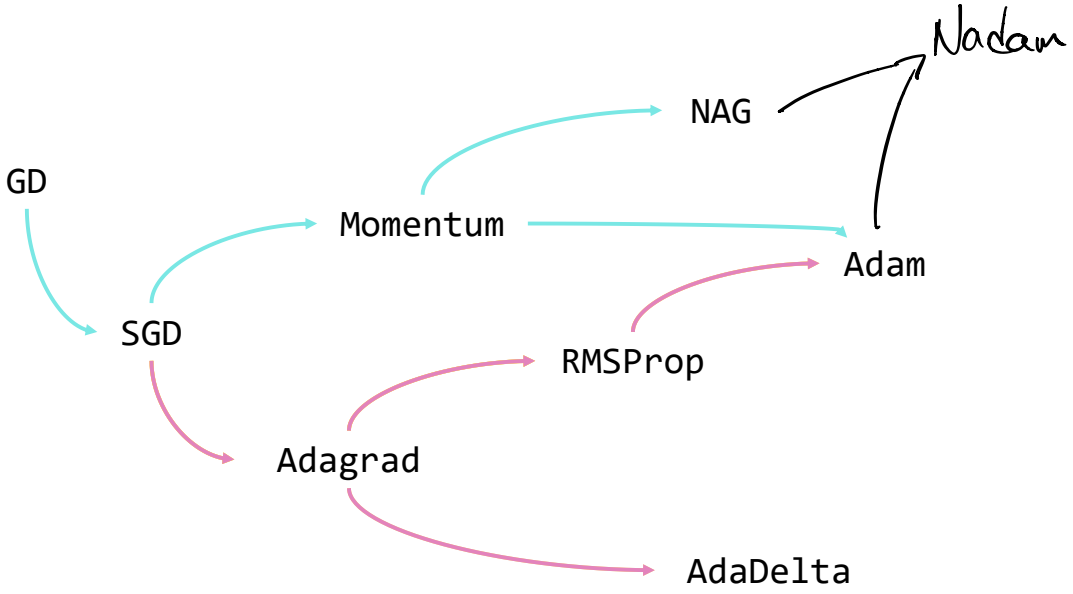
Review

Optimizer



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

Review

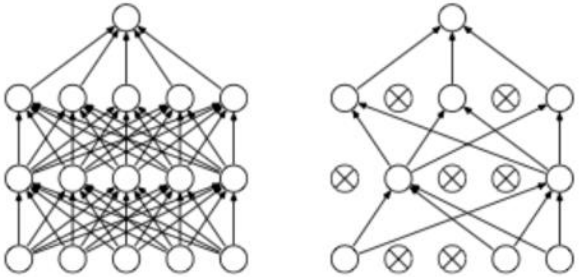


Review

Regularization → overfitting X

- Dropout
- L1, L2 weight decay. w^2 $|w|$
- BN
- Data augmentation

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 14 2nd
3. GoogLeNet 14 1st
4. ResNet 15 1st

2010

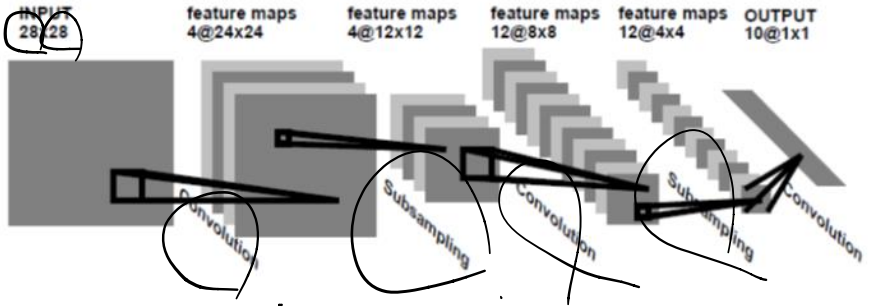
ILSVRC

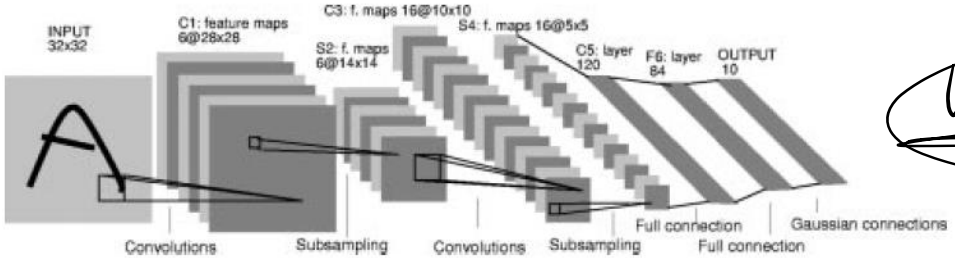
ImageNet Large Scale Visual Recognition Competition

LeNet(1990 – 1998)

Yann Lecun.

MLP

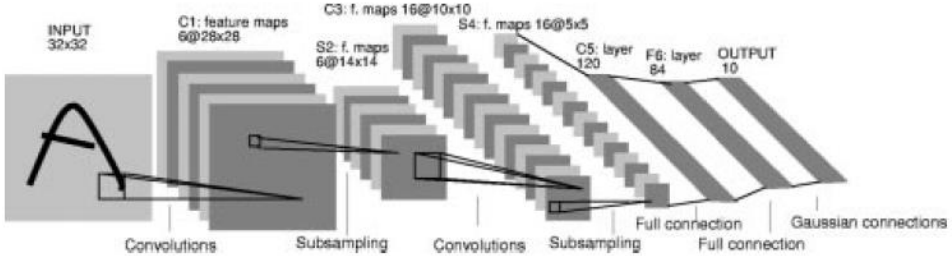




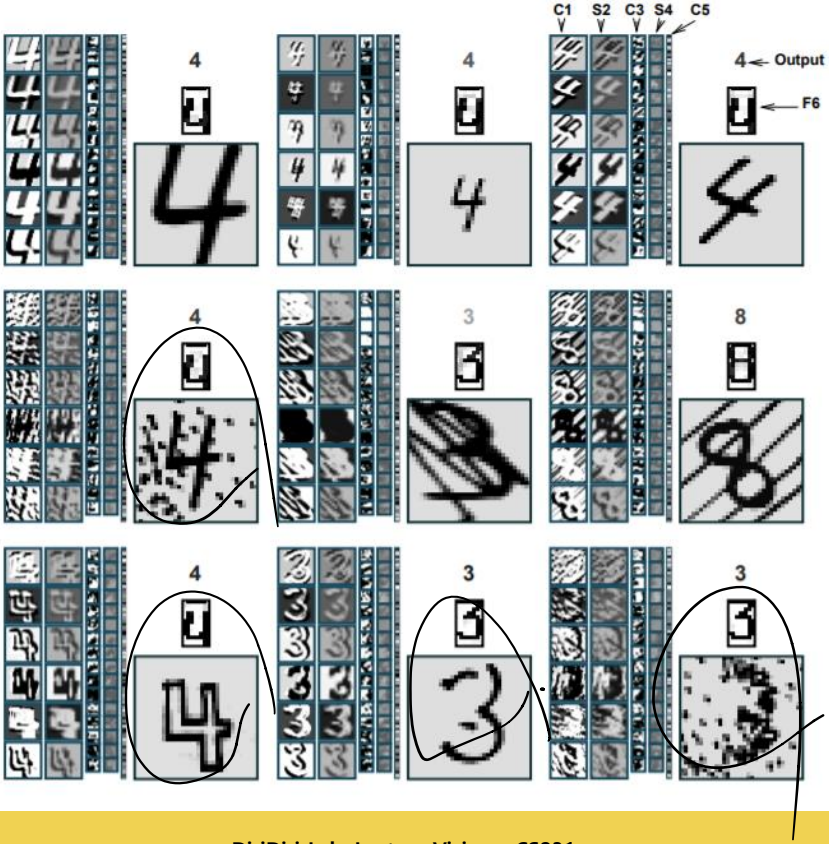
LeNet-1

LeNet-5

LeNet 5(1998)

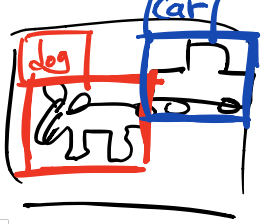


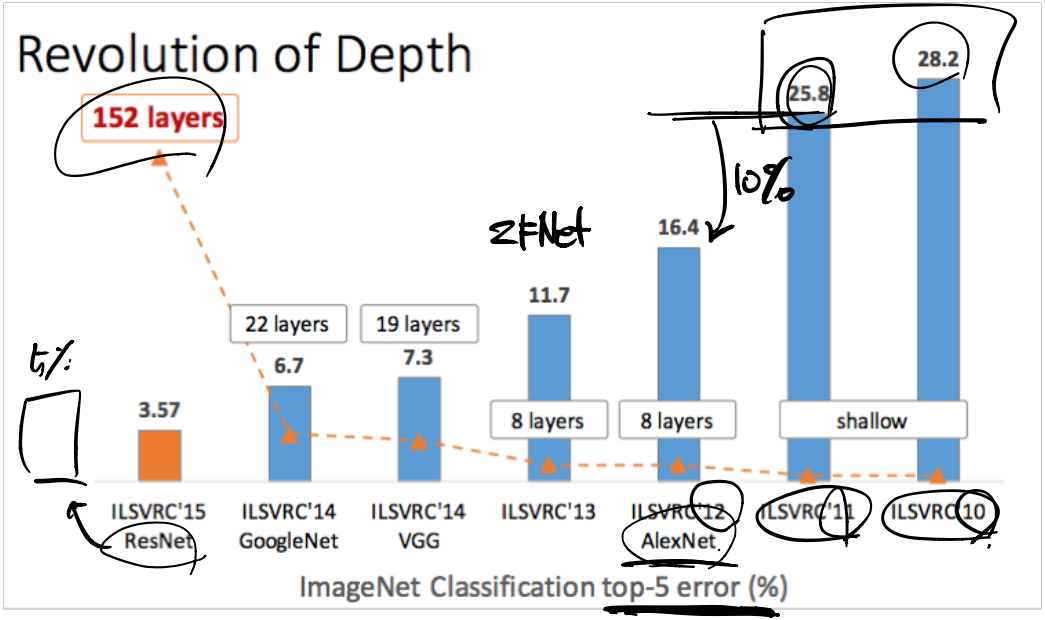
Channel	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X


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

CNN Architectures

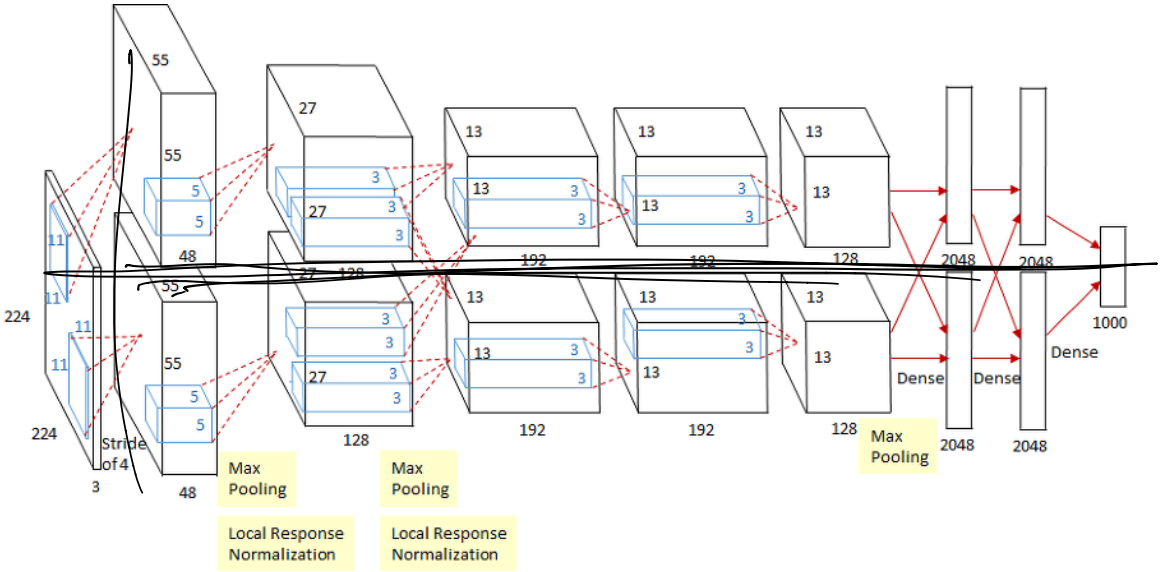
Classification
Localization





AlexNet(2012)

GPU -> GTX 580 x2.



AlexNet(2012)

ReLU.
Conv – Pool – Conv – Pool
↑
Norm

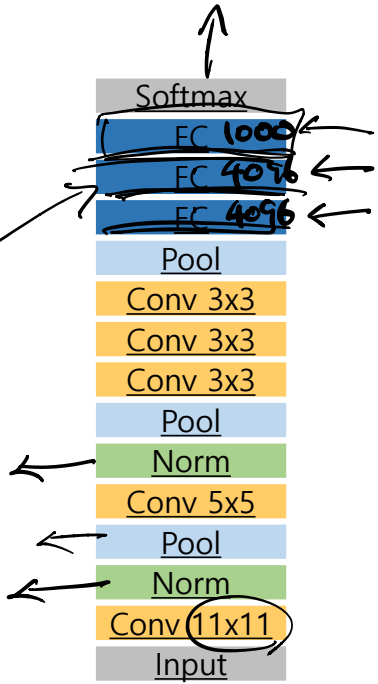
Dropout 0.5

4096 x 4096. = 2

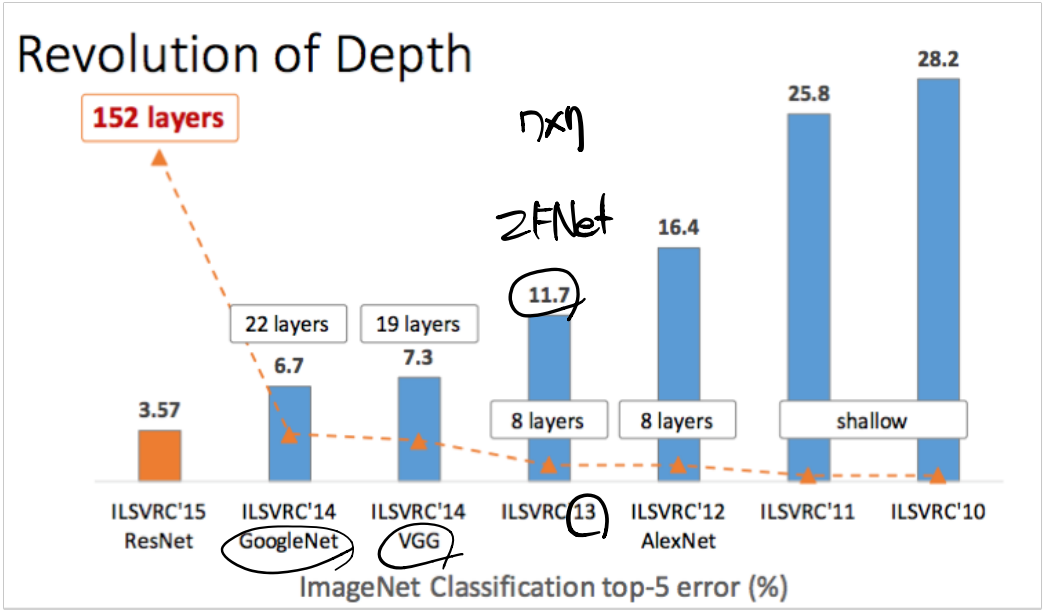
1. GPU 2개를 이용한 특이한 구조
2. ReLU 처음으로 사용한 모델
3. 현재는 사용하지 않는 Normalization layer를 사용
4. Data Augmentation / Dropout 으로 overfitting 방지
5. SGD + Momentum (0.9)

lr = 0.01

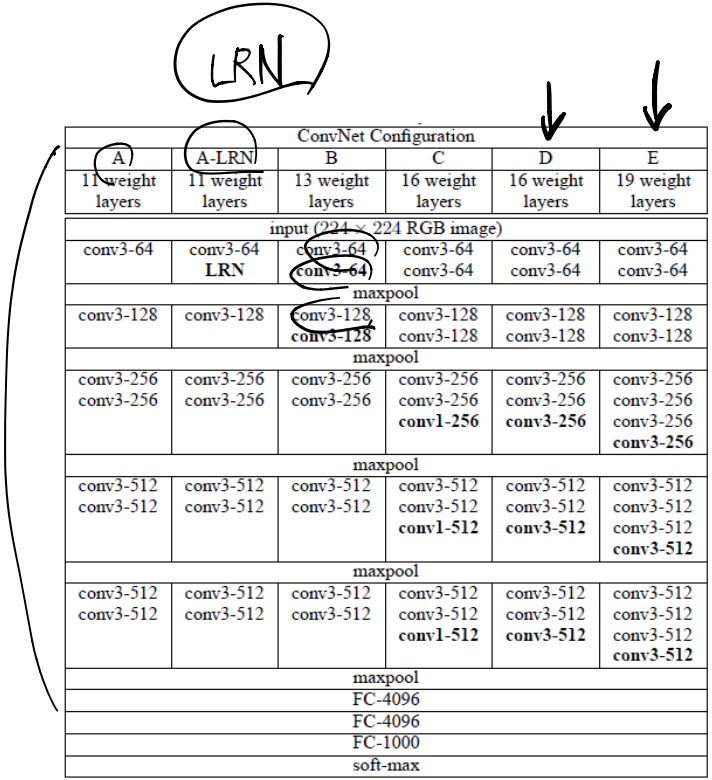
L2 weight decay.
0.0005



CNN Architectures

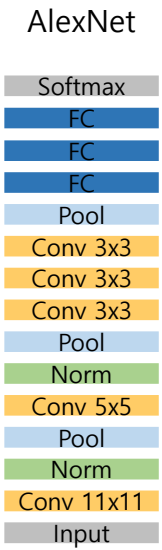


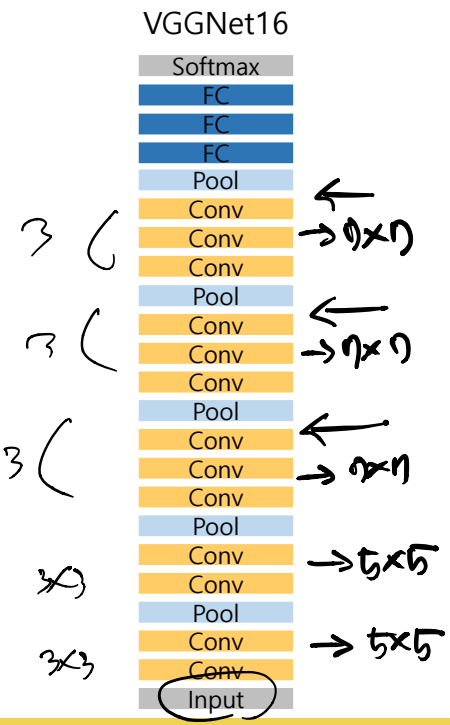
VGGNet



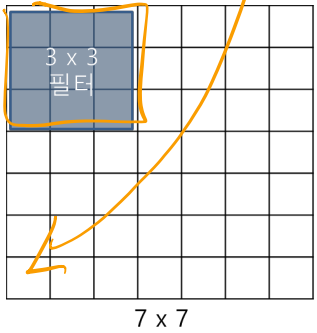
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	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

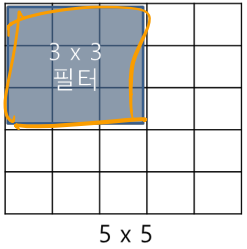

AlexNet
Softmax
FC
FC
FC
Pool
Conv 3x3
Conv 3x3
Conv 3x3
Pool
Norm
Conv 5x5
Pool
Norm
Conv 11x11
Input


VGGNet 16
VGGNet16
Softmax
FC
FC
FC
Pool
Conv
Conv
Conv
Pool
Conv
Conv
Conv
Pool
Conv
Conv
Conv
Pool
Conv
Conv
Pool
Conv
Conv
Input

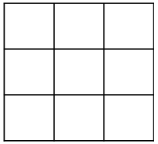
VGGNet

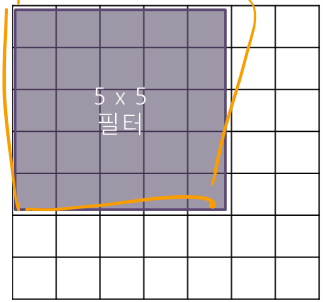


Stride 1로
컨볼루션

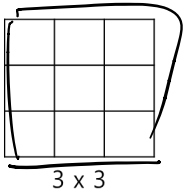


Stride 1로
컨볼루션

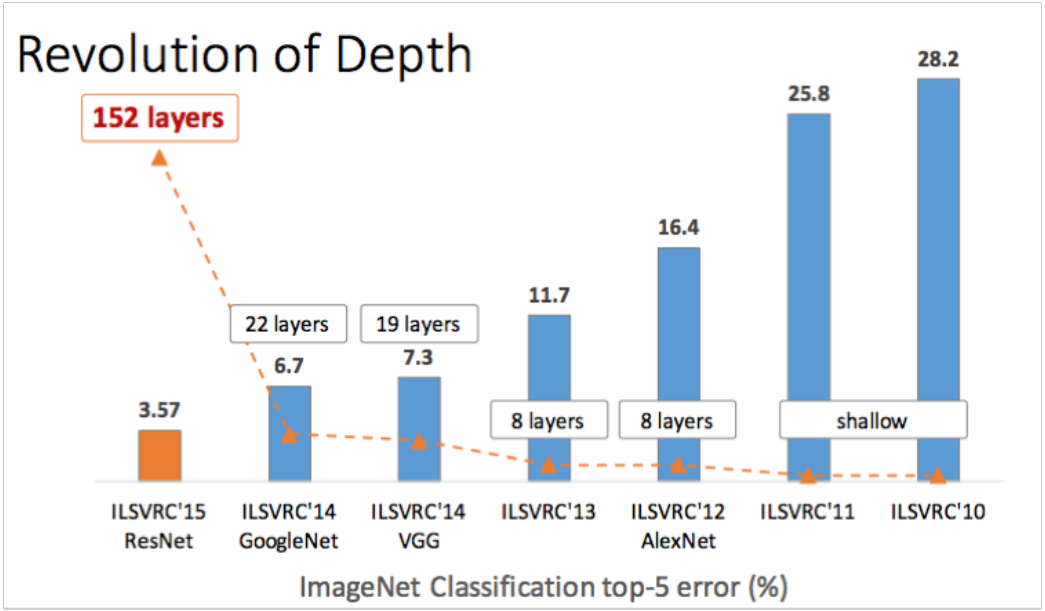




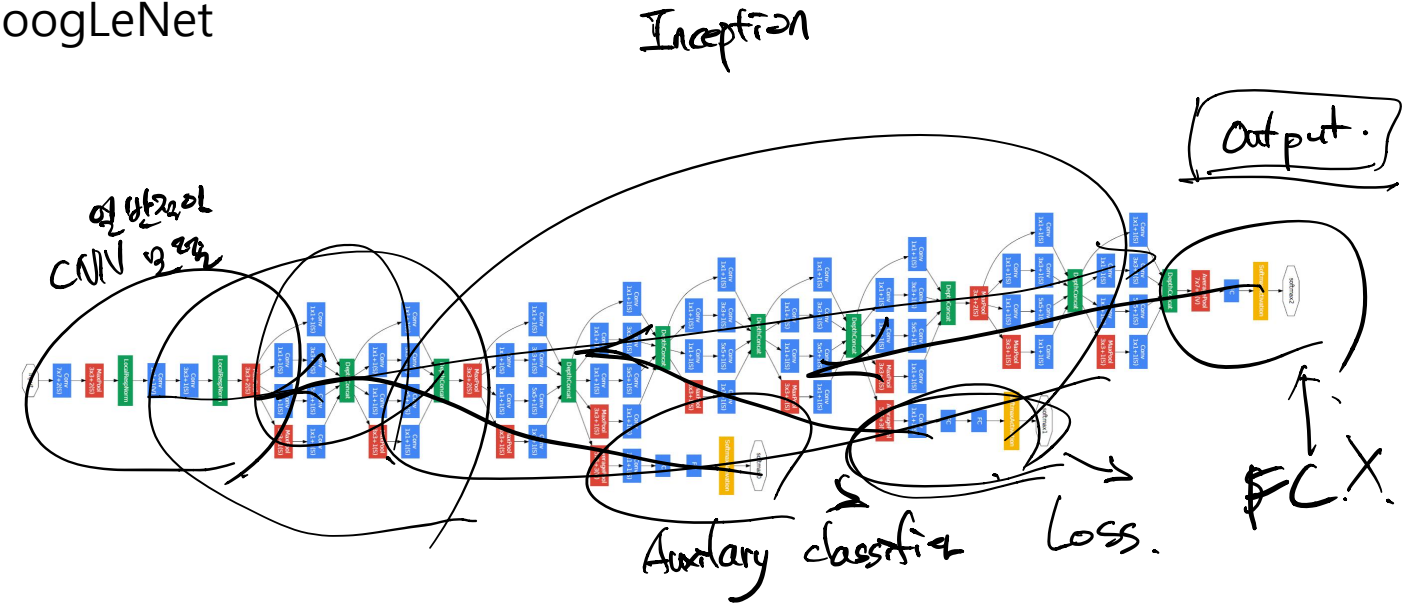
Stride 1로
컨볼루션



CNN Architectures

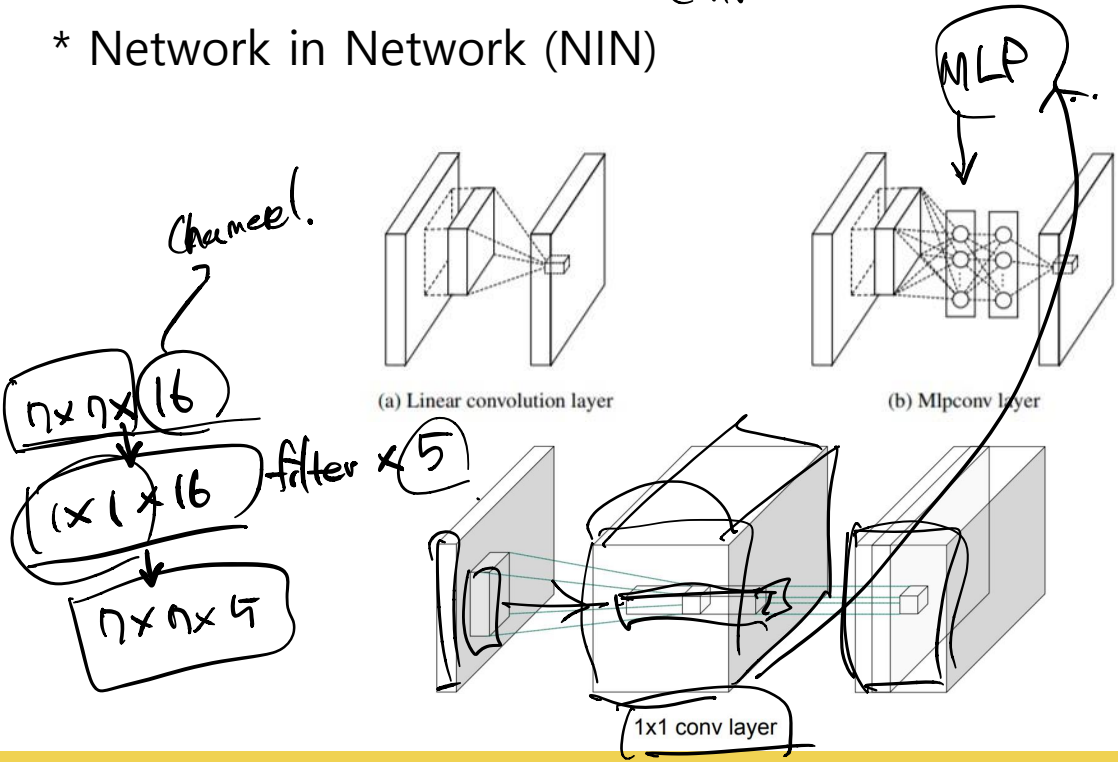


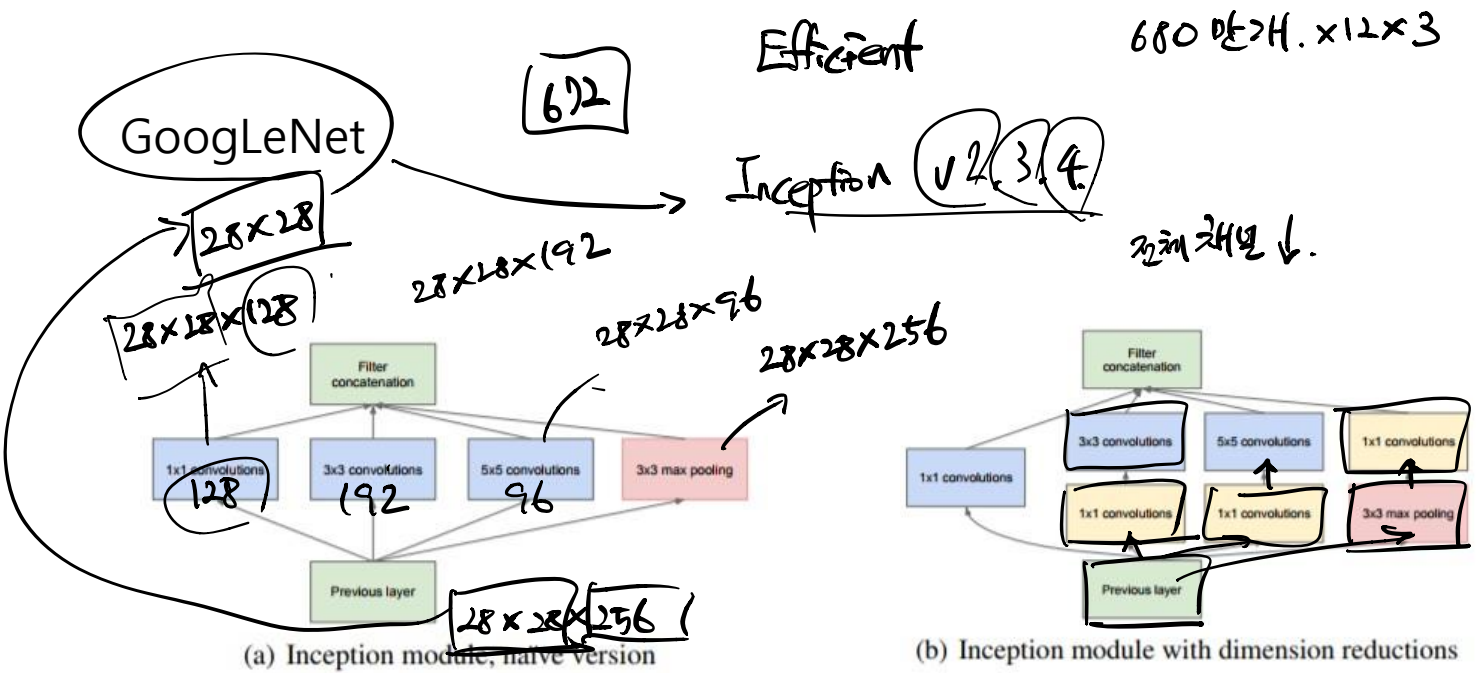
GoogLeNet



Conv

* Network in Network (NIN)

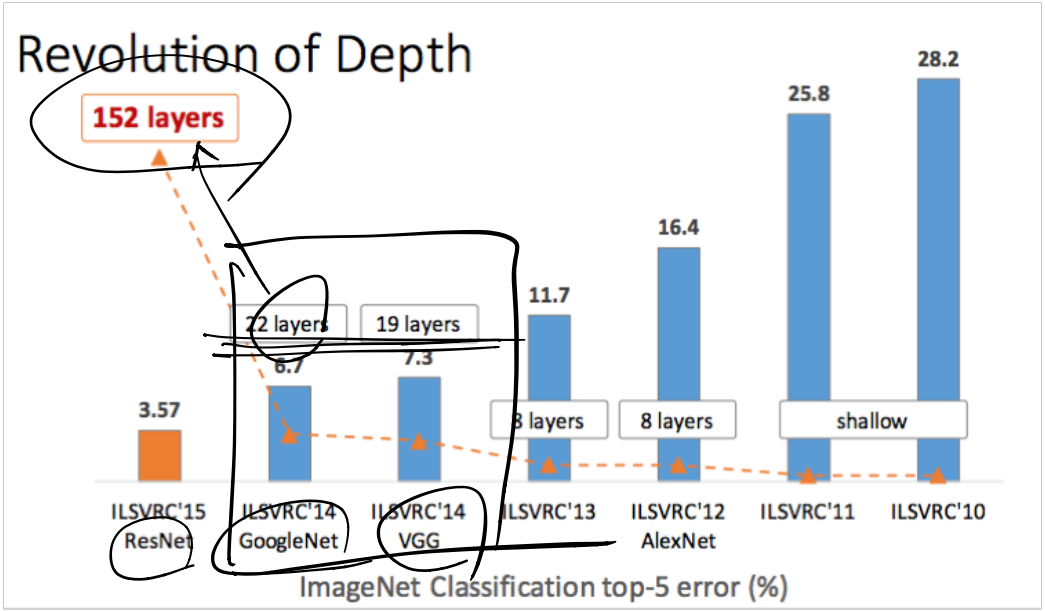




(a) Inception module, naive version

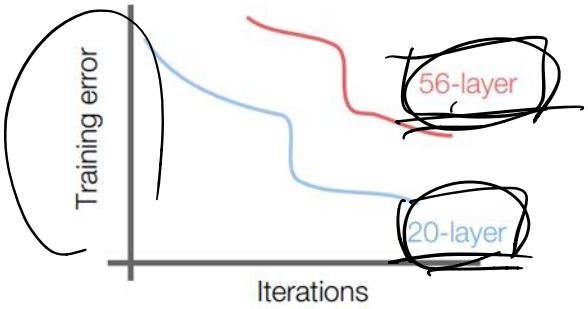
(b) Inception module with dimension reductions

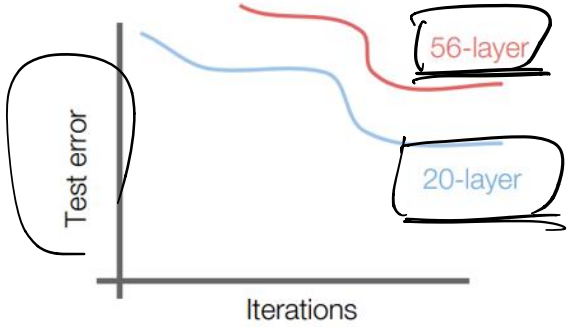
CNN Architectures



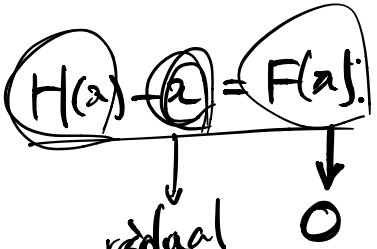
ResNet

overfitting X.





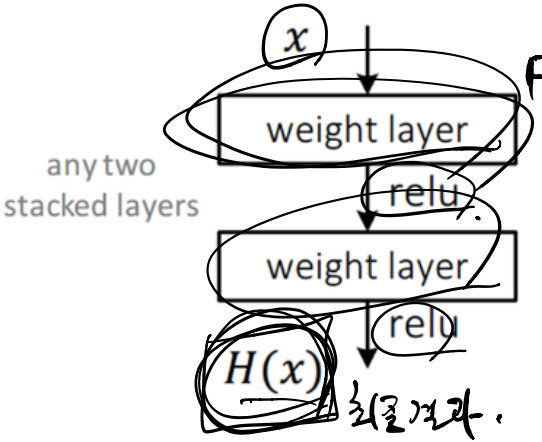
ResNet



H(x) = F(x) + x.

residual learning.

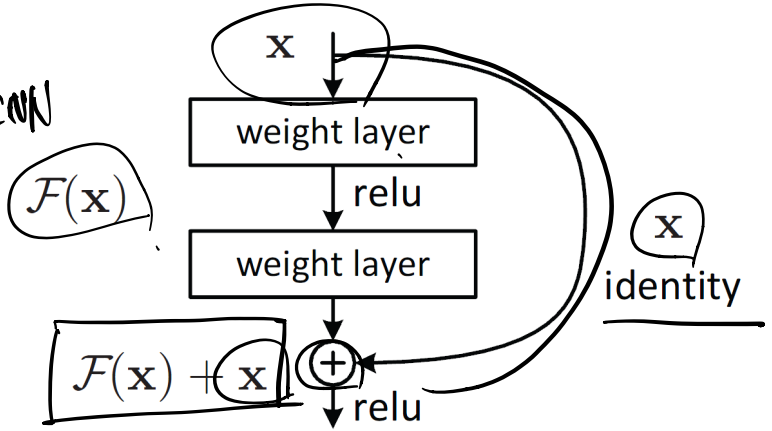
이전.



FC/CNN

최종결과.

H(x) - x.

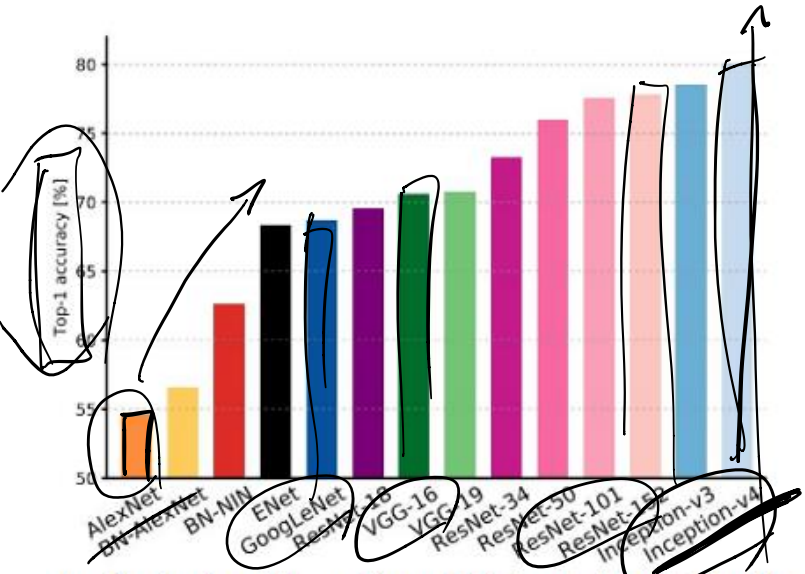


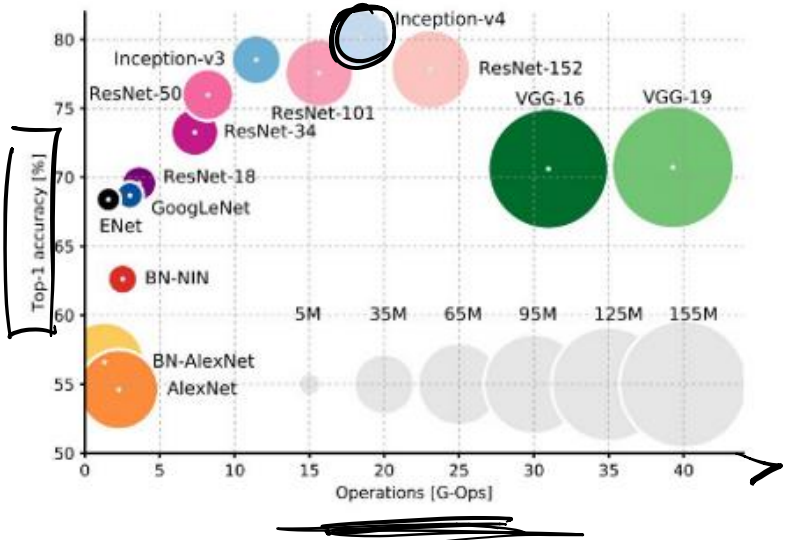
ResNet

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**

GoogleNet + ResNet

Transfer Learning

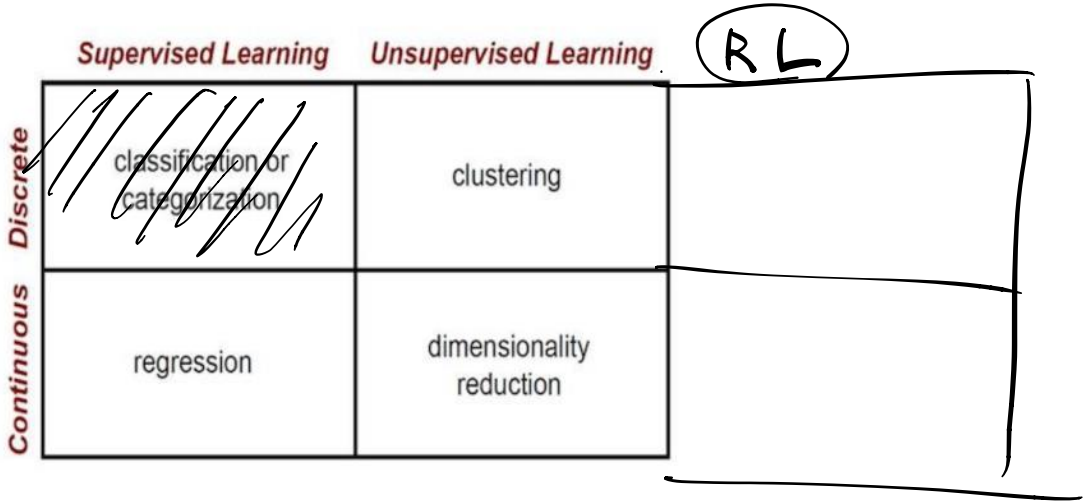




An Analysis of Deep Neural Network Models for Practical Applications, 2017.

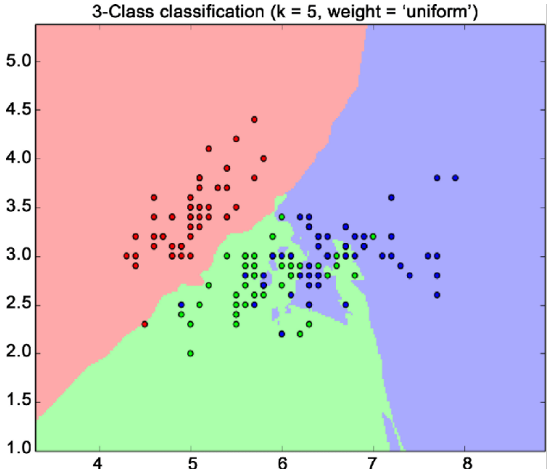
9회차 동안...

ML

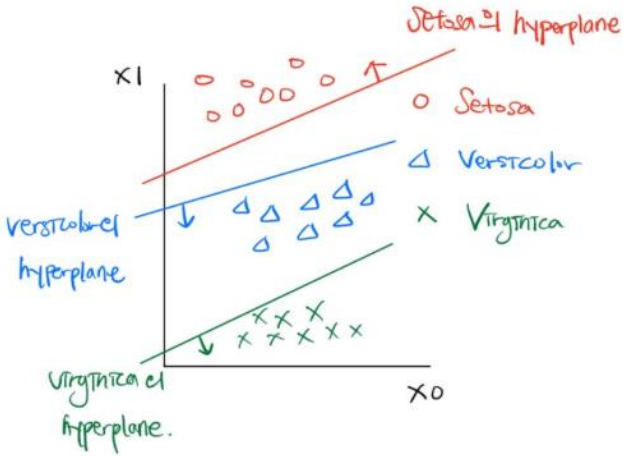

Supervised Learning
Unsupervised Learning
RL
Discrete
Continuous
classification or categorization
clustering
regression
dimensionality reduction

9회차 동안...

KNN

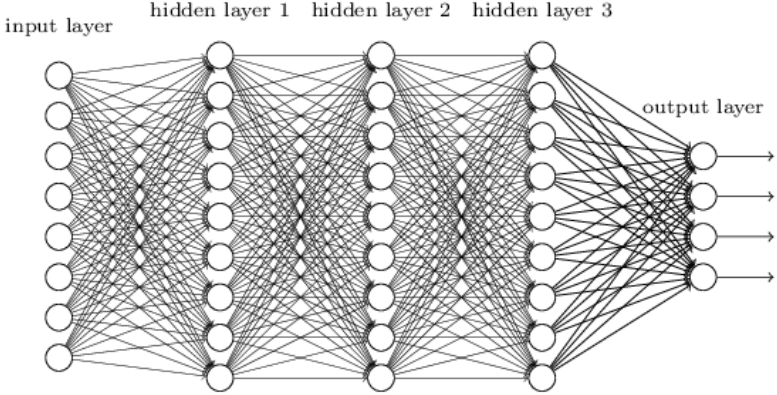


Linear Classifier

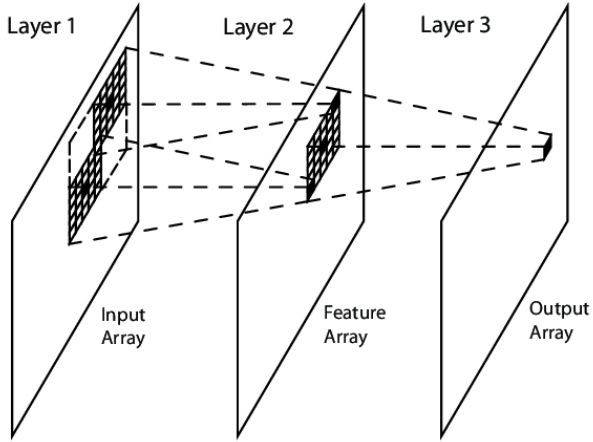


9회차 동안...

MLP



CNN

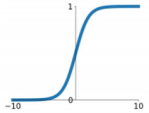


9회차 동안...

Activation Function

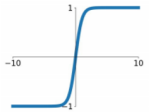
Sigmoid

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



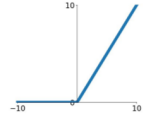
tanh

$\tanh(x)$



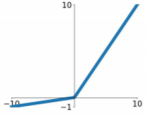
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

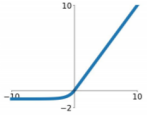


Maxout

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

ELU

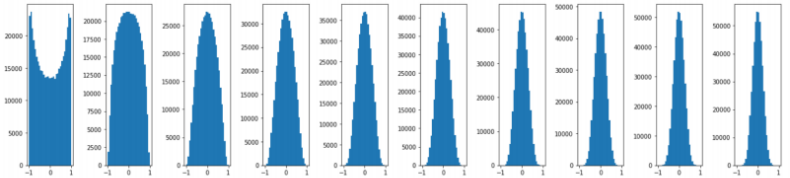
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Xavier initialize

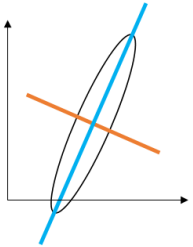
$np.random.randn(size = (D, H))/np.sqrt(D)$

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

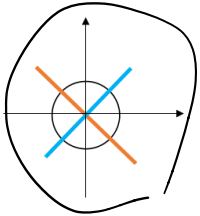


9회차 동안...

Before normalization

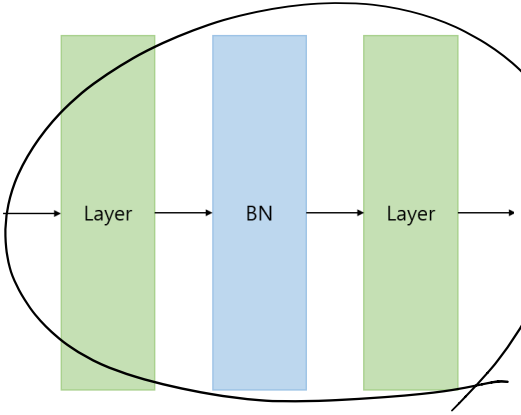


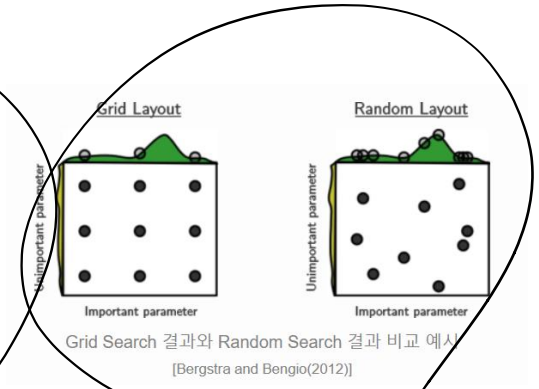
After normalization



$$\frac{x - x_{min}}{x_{max} - x_{min}}$$

$$\frac{x - \mu}{\sigma}$$





9회차 동안...

Crop

Flip

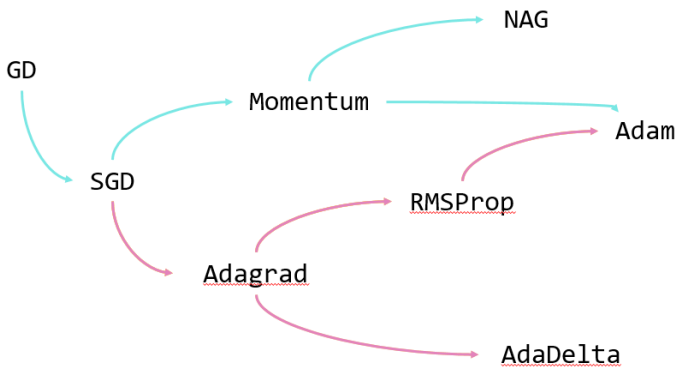
Noise

Scale

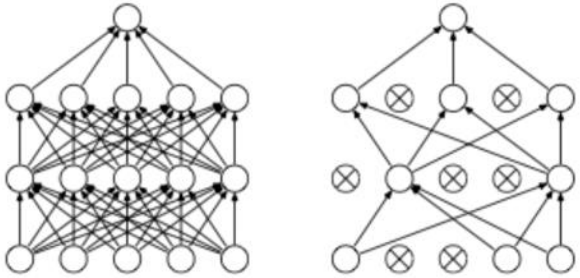
Rotate

Translation

9회차 동안...



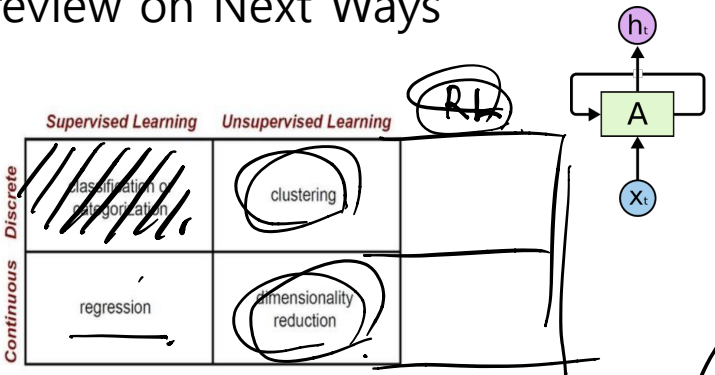
Dropout: A Simple Way to Prevent Neural Networks from Overfitting [Srivastava et al. 2014]

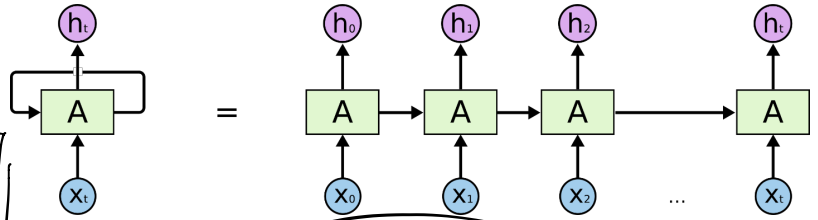


(a) Standard Neural Net (b) After applying dropout.

Preview on Next Ways

RNN → LSTM





denseNet.

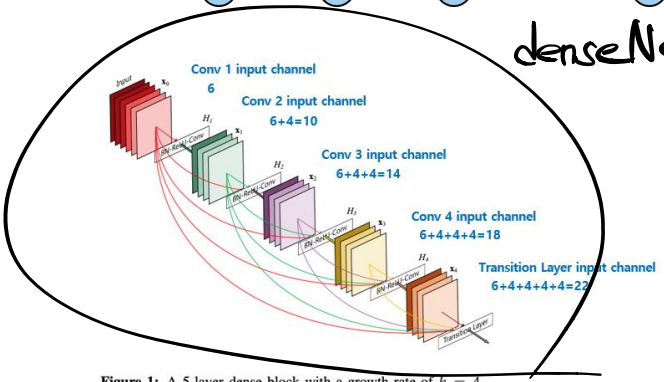


Figure 1: A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.

kaggle