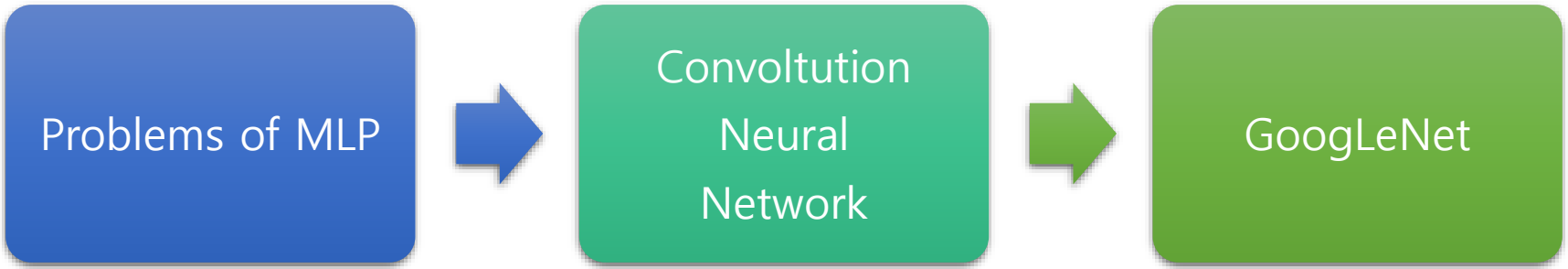
GoogLeNet

Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

Jeon Hyun Ho

Contents



Introduction

- Artificial Neural Network (ANN)

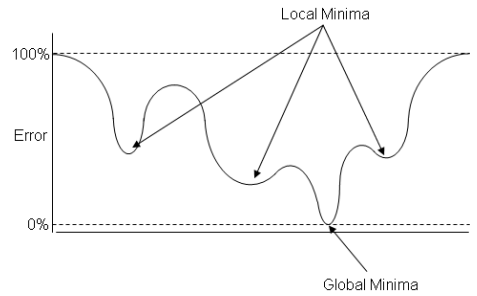
- 40년 대에 개발된 알고리즘

부흥

- 70 ~ 80년대 Back Propagation 알고리즘을 통해 Learning이 가능해짐
- 80년대 후반에 뉴런의 개수가 무한하다면 어떠한 비선형 함수도 표현할 수 있음이 수학적으로 증명됨 (Universal approximation theorem)

몰락

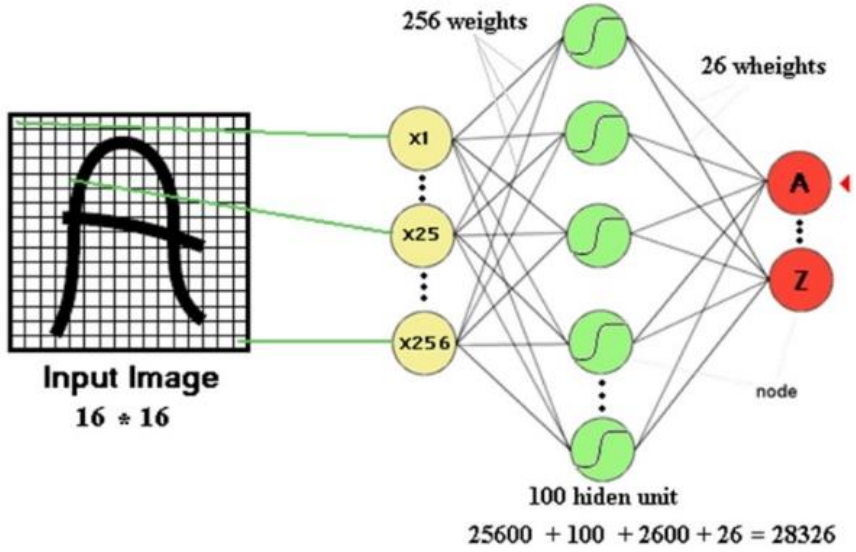
- 복잡한 데이터를 분류(Classification)하기 위해서는 여러 층의 신경망이 필요함
- 층이 깊어지면서 Learning 과정에서 local minima에 수렴하는 문제가 발생

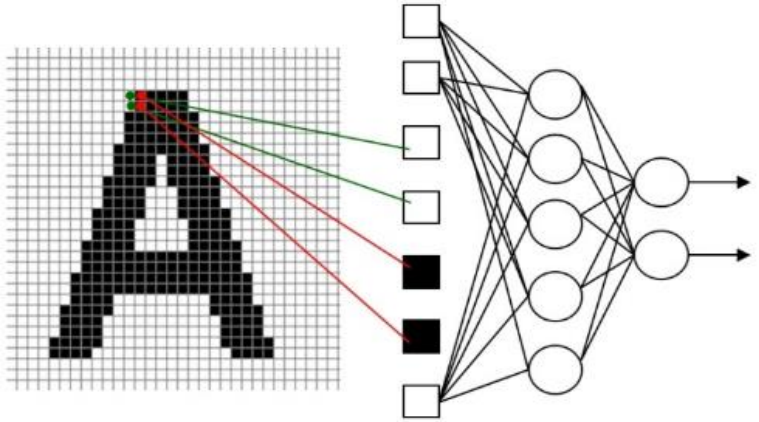


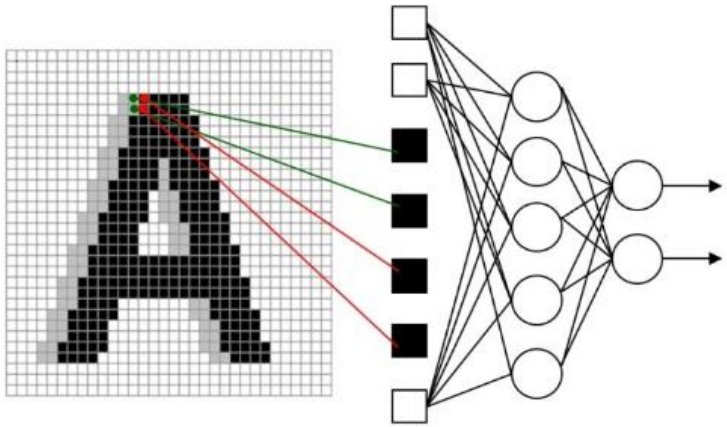
Problems

- Problems of MLP

- Training time
- Network size
- Parameters

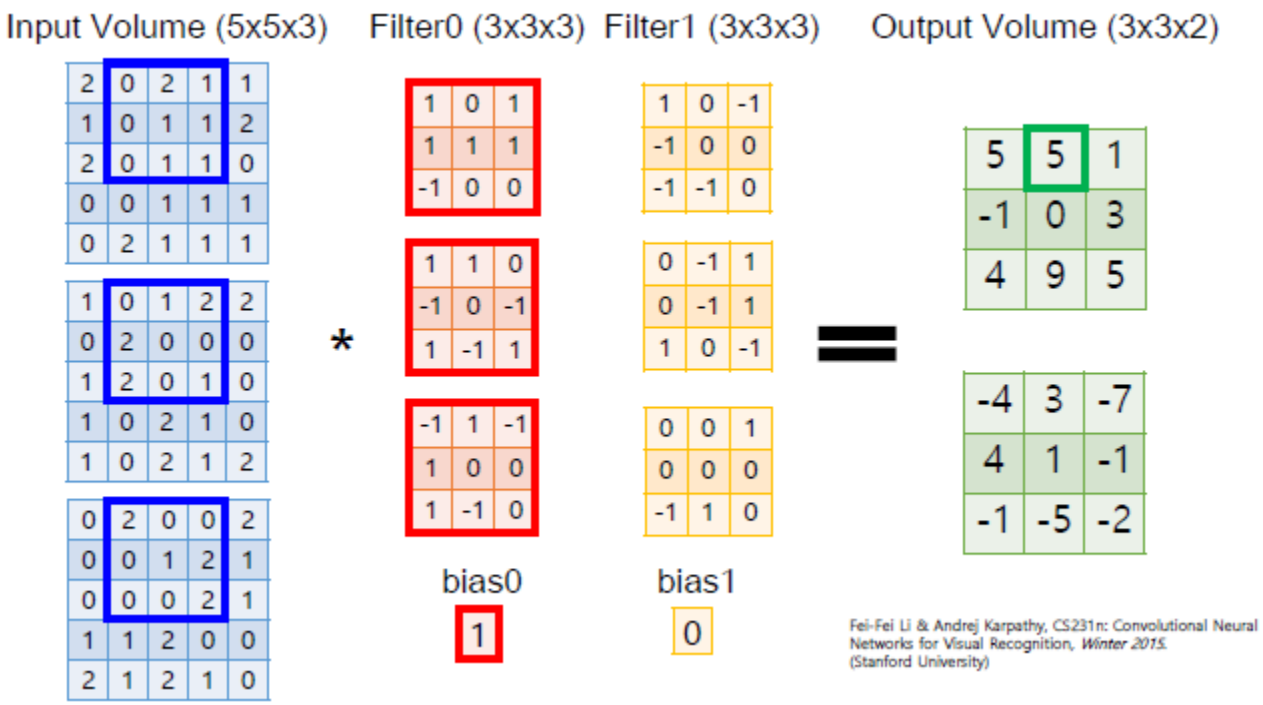


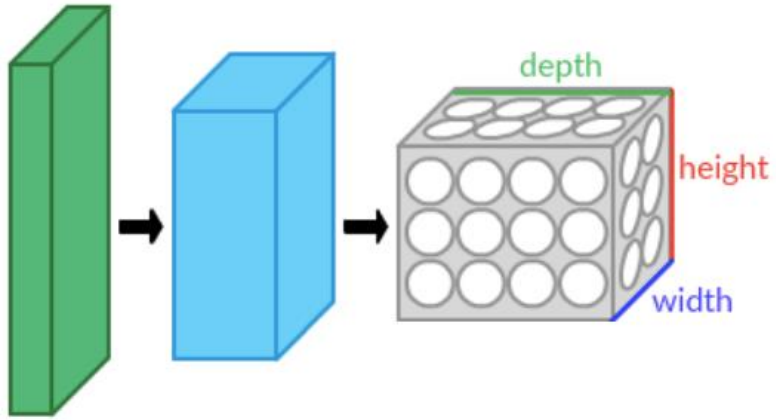




CNN

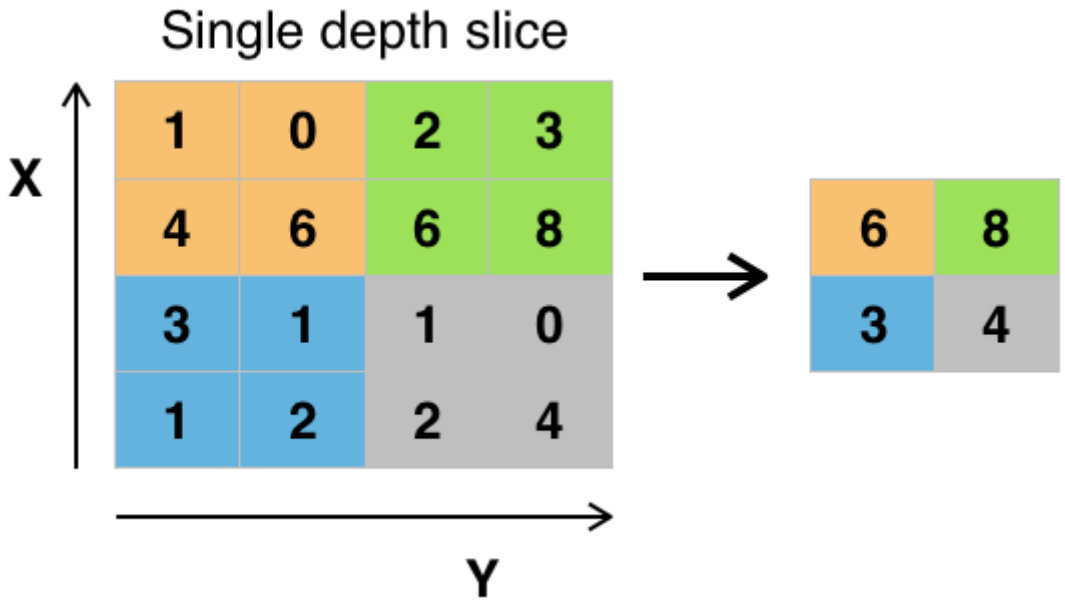
- Convolution





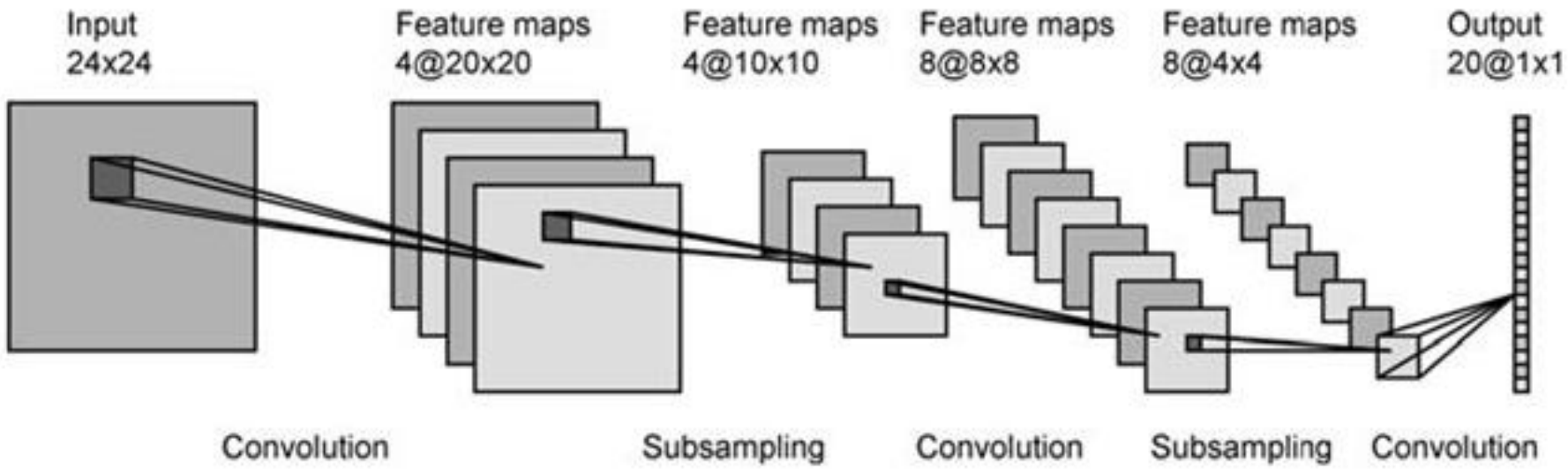
CNN

- Sub sampling (Pooling)
  - 2x2 filter and stride = 2



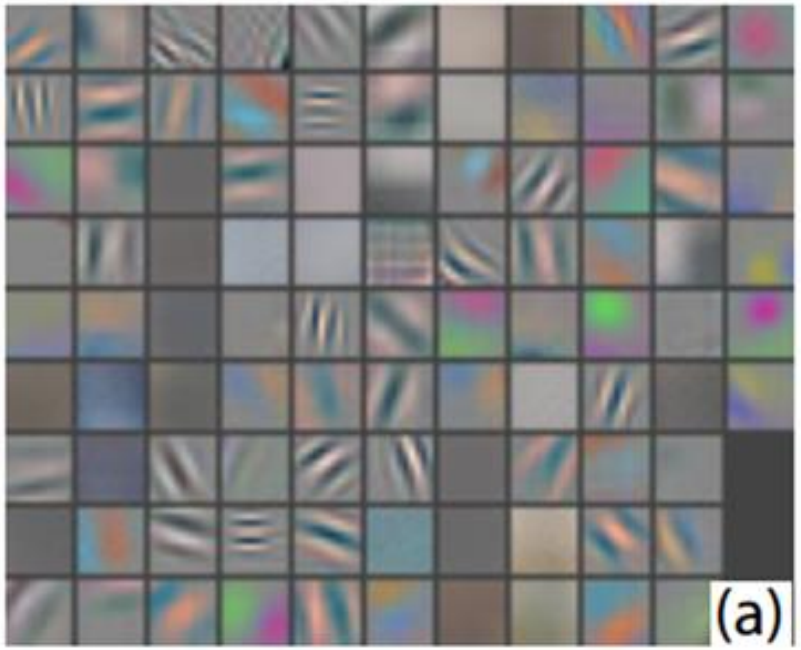
CNN

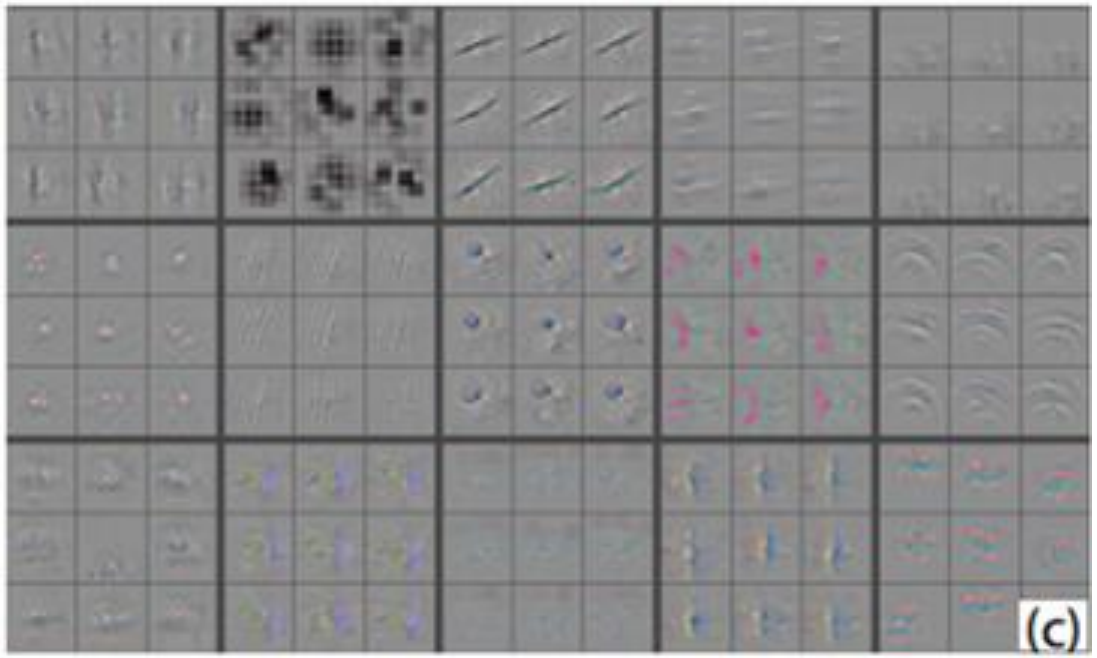
- 기본적인 CNN의 구조



CNN

- Convolution 반복의 결과 - 1



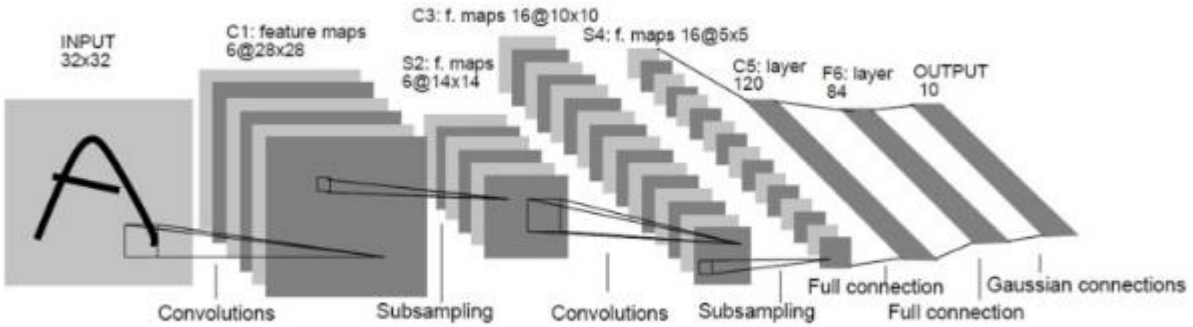


CNN

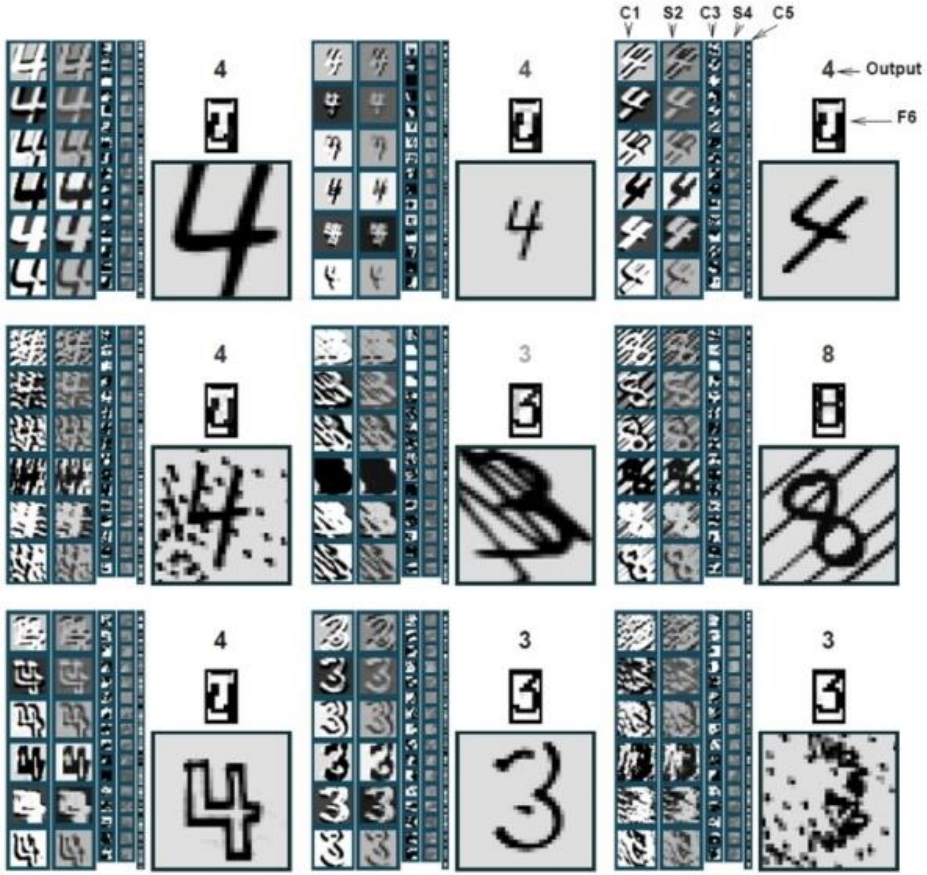
- Convolution 반복의 결과 - 2

CNN

- LeNet - Yann LeCun



LeNet의 구조



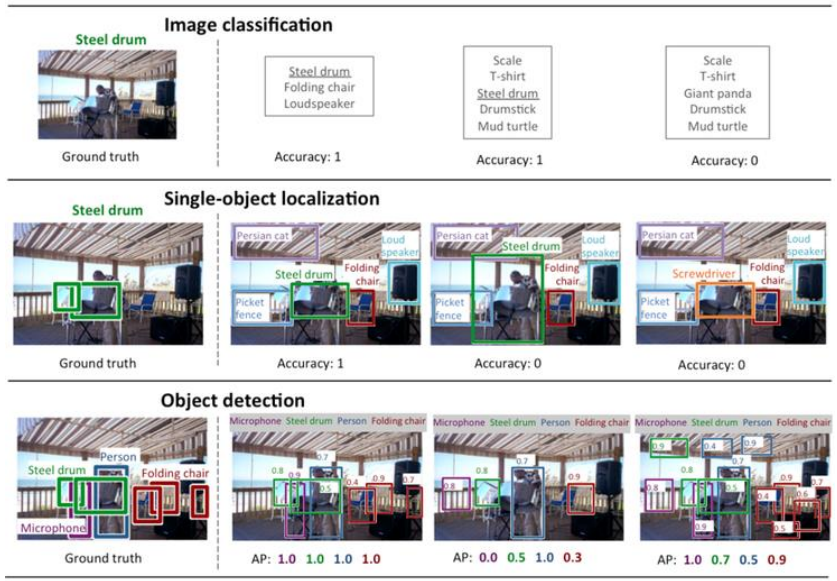
결과

CNN

- ImageNet

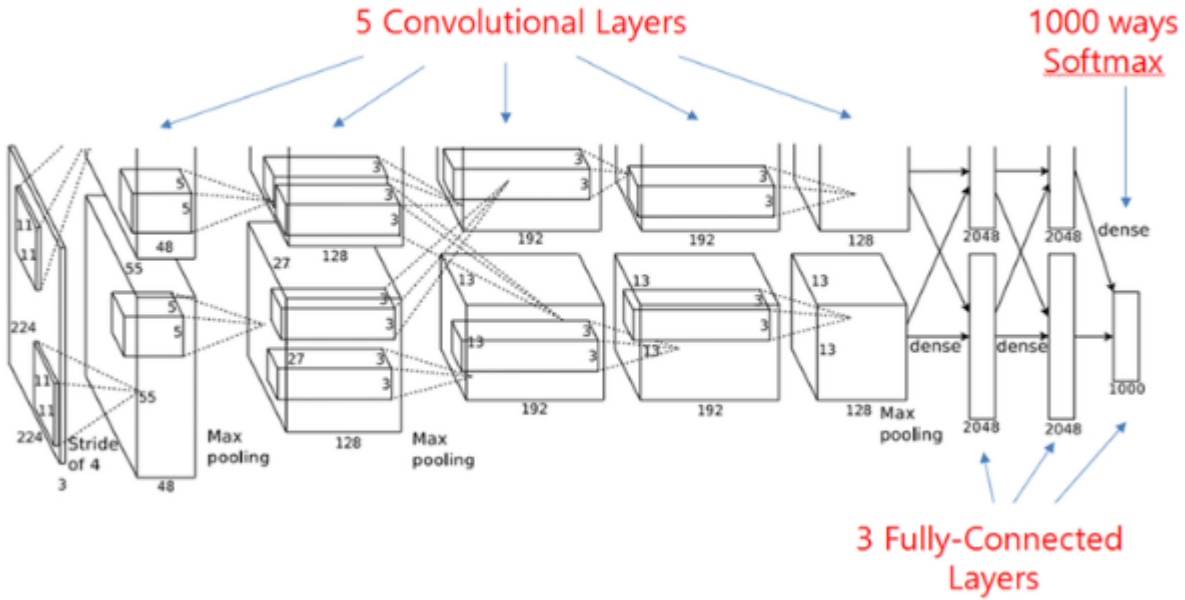
- 세계 최대의 영상 데이터 베이스
- 약 22000종류, 1500만장의 이미지
- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)를 2010년부터 매년 개최

IMAGENET

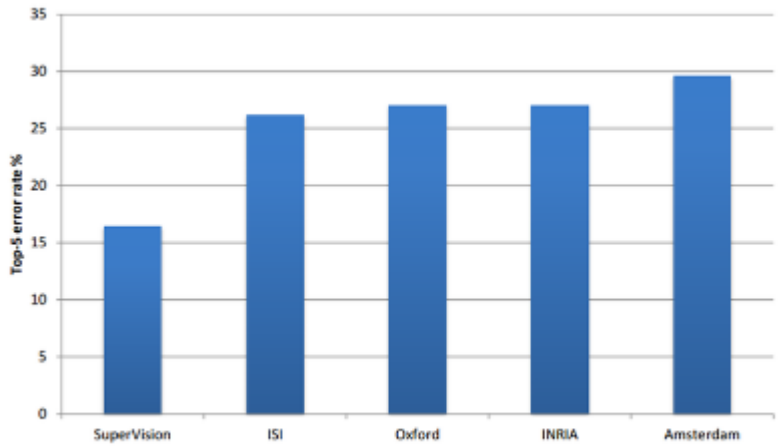


CNN

- AlexNet - Krizhevsky, Hinton



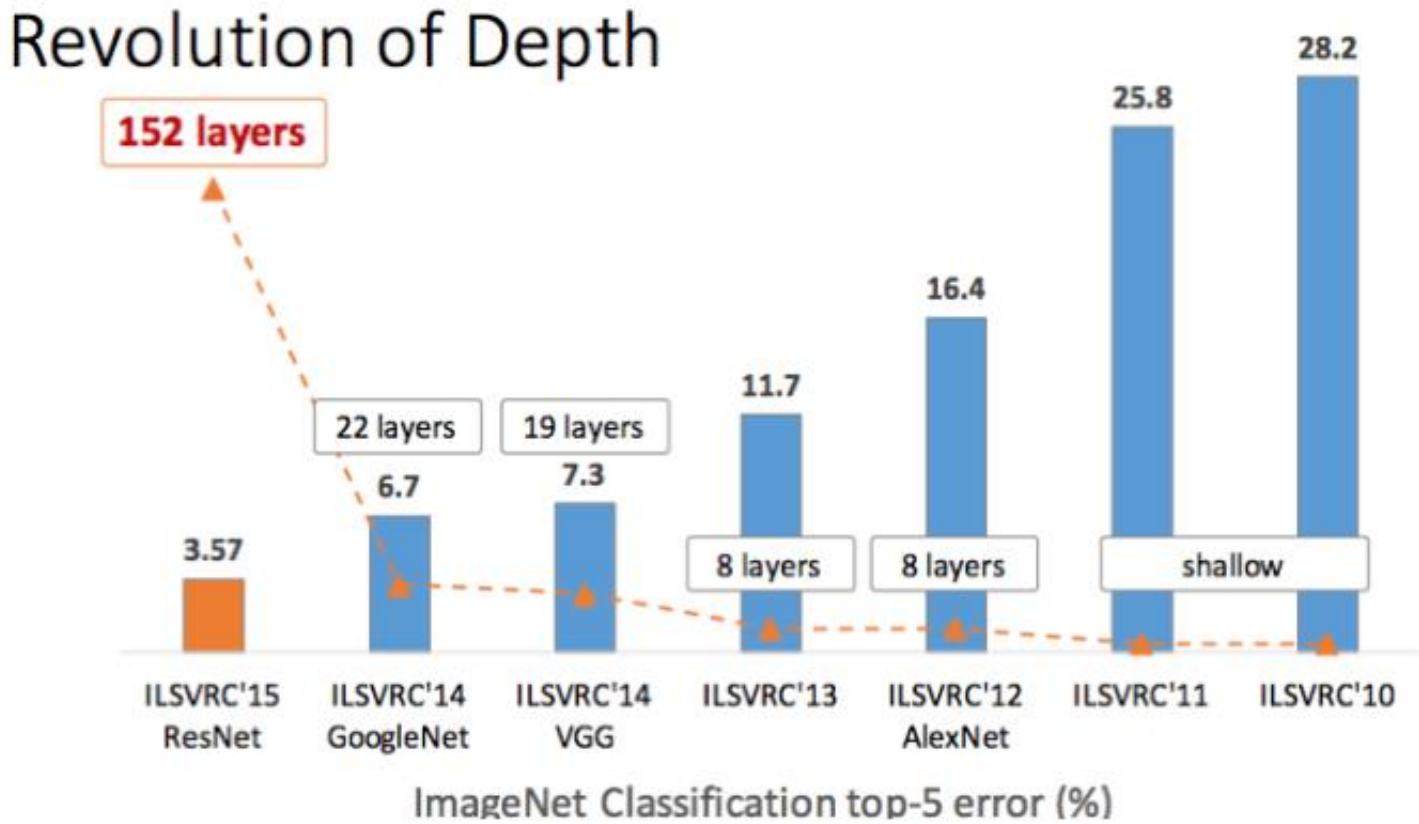
AlexNet의 구조



에러율

GoogLeNet

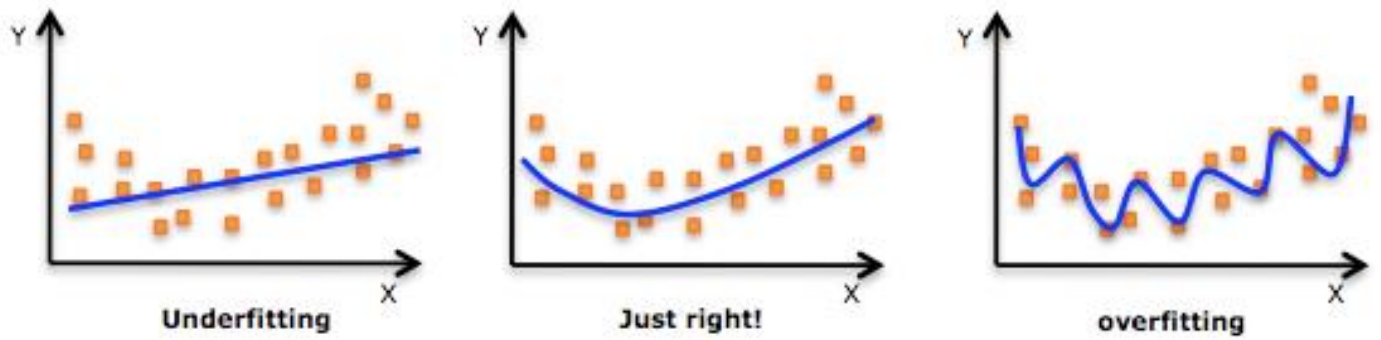
- ILSVRC 2014 ~ - deeper and deeper



GoogLeNet

- Deep NN - Side effect

- Overfitting
- Label
- Computation load
- Vanishing gradient

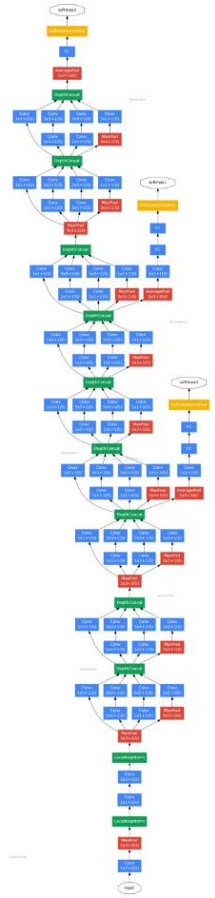


overfitting

GoogLeNet

- GoogLeNet

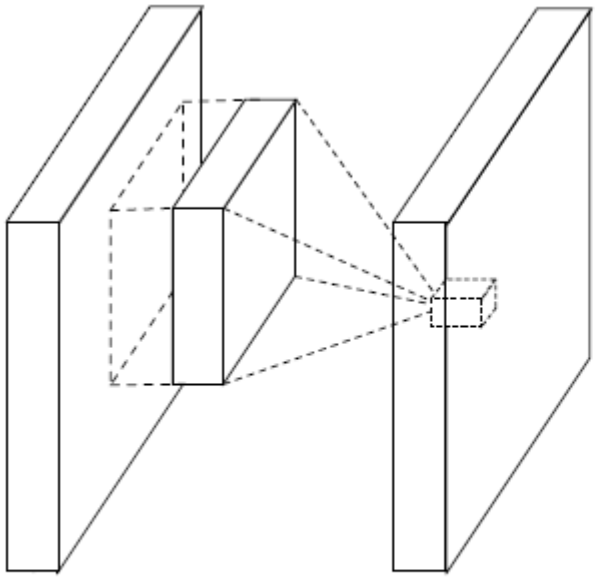
영화 '인셉션'



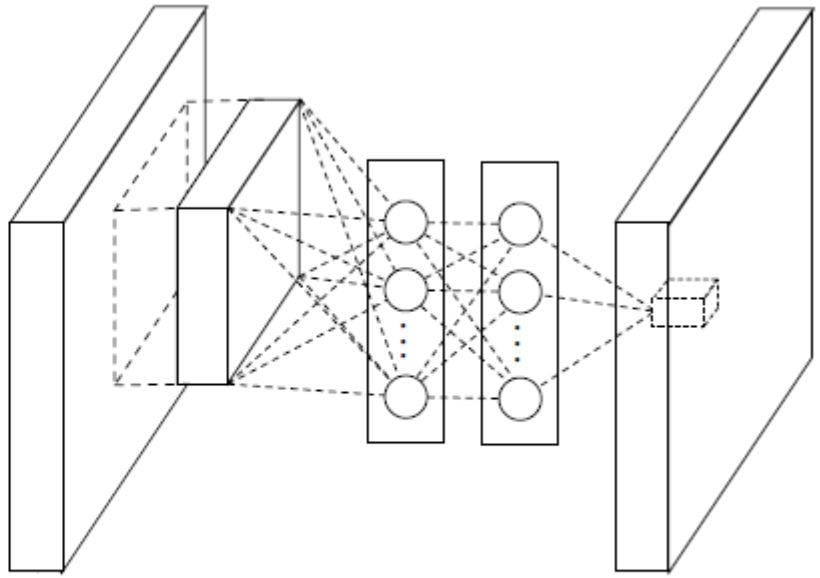
GoogLeNet의 구조

GoogLeNet

- GoogLeNet - NIN



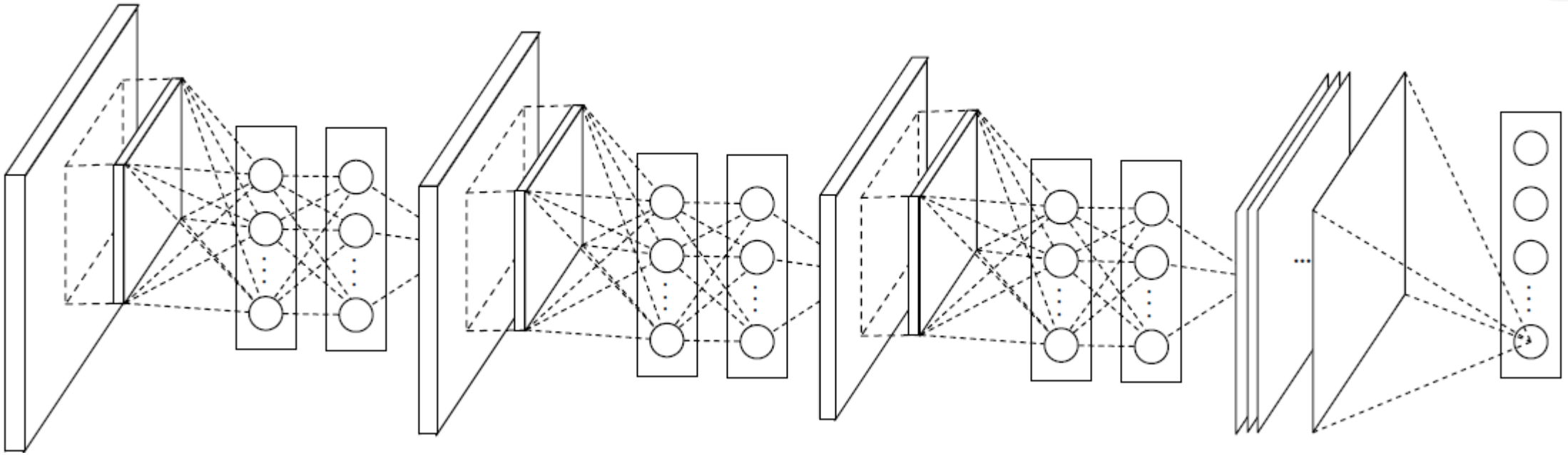
(a) Linear convolution layer



(b) Mlpconv layer

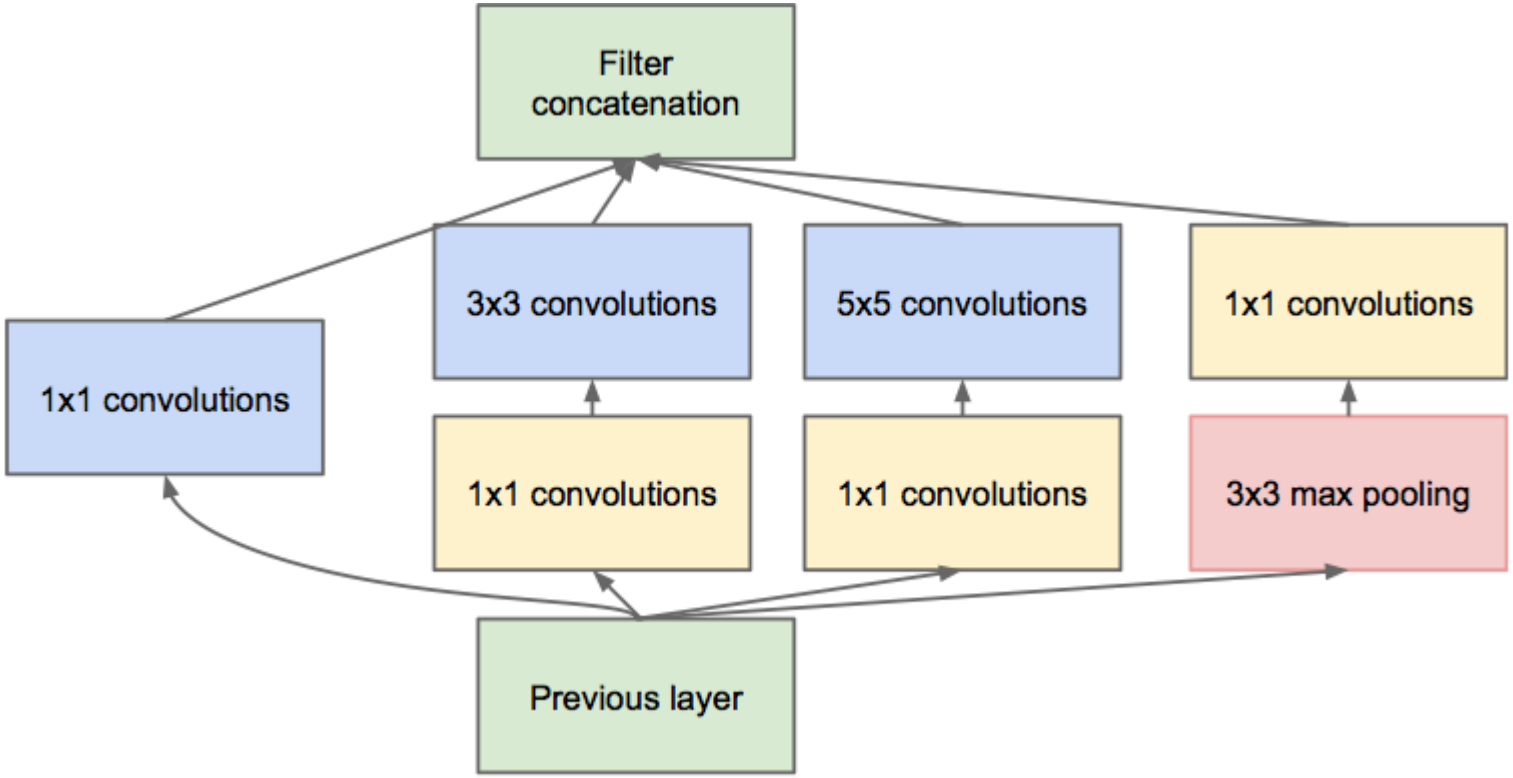
GoogLeNet

- GoogLeNet - NIN



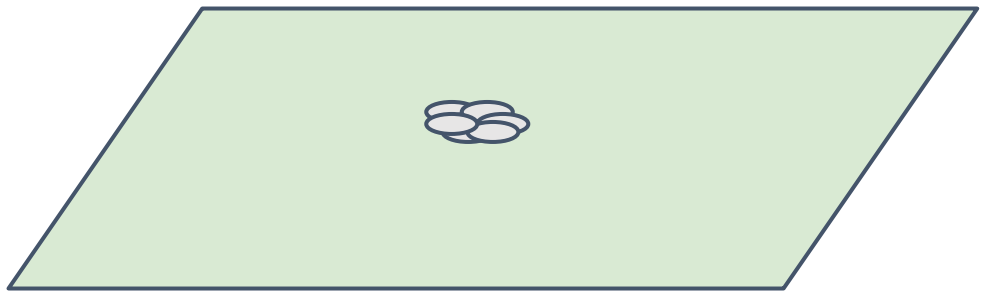
GoogLeNet

- Inception module



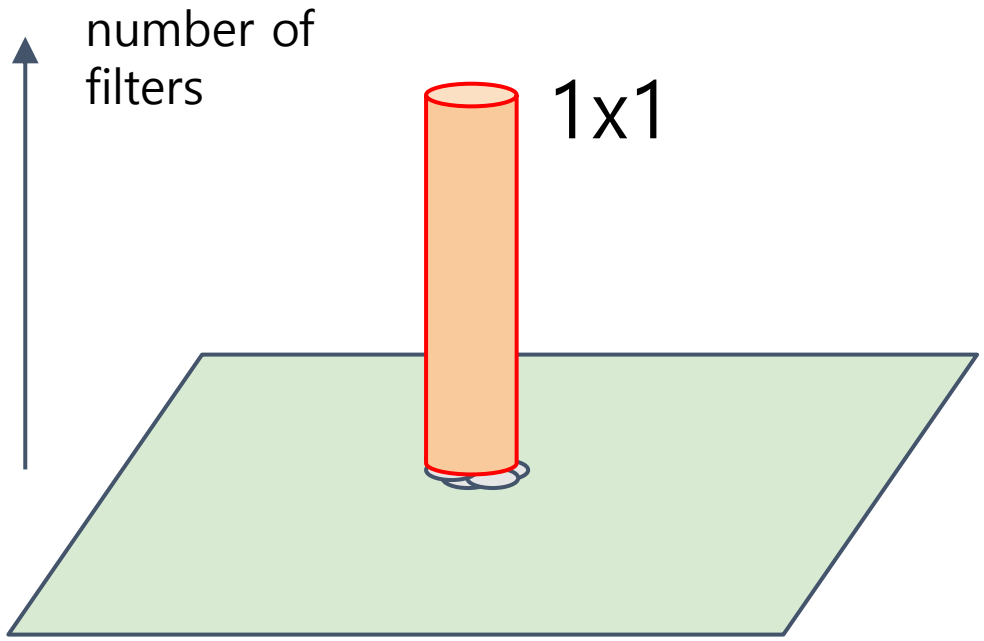
GoogLeNet

- Inception concept



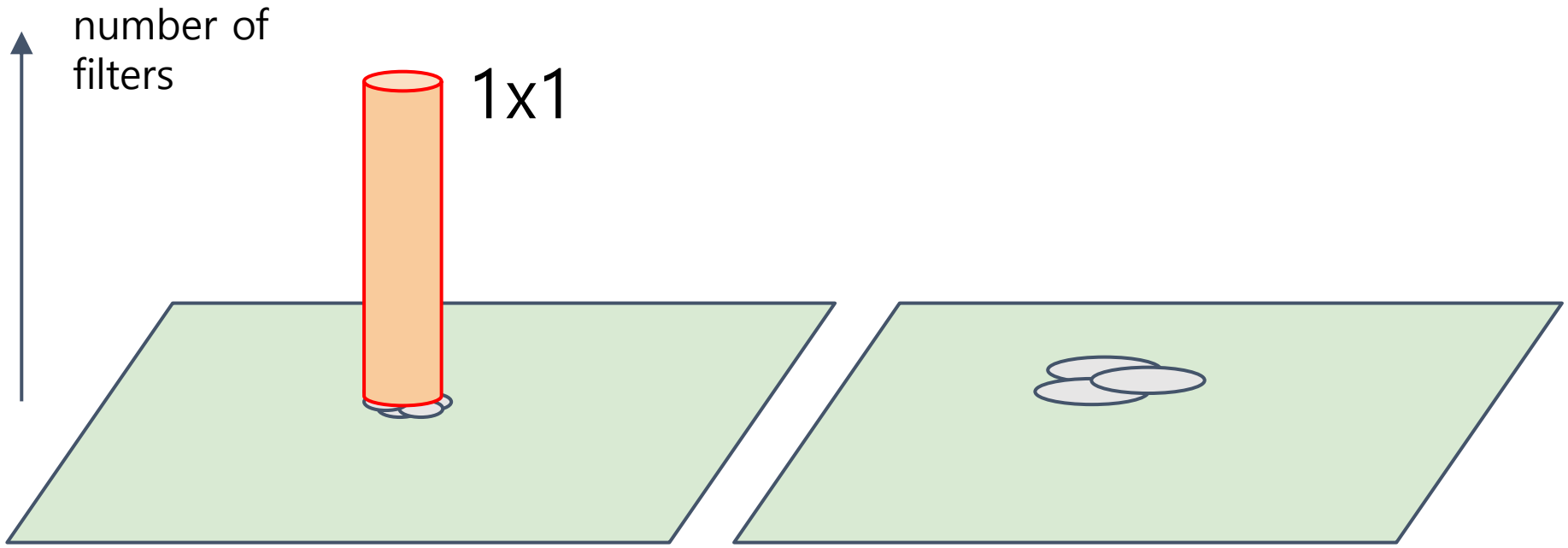
GoogLeNet

- Inception concept



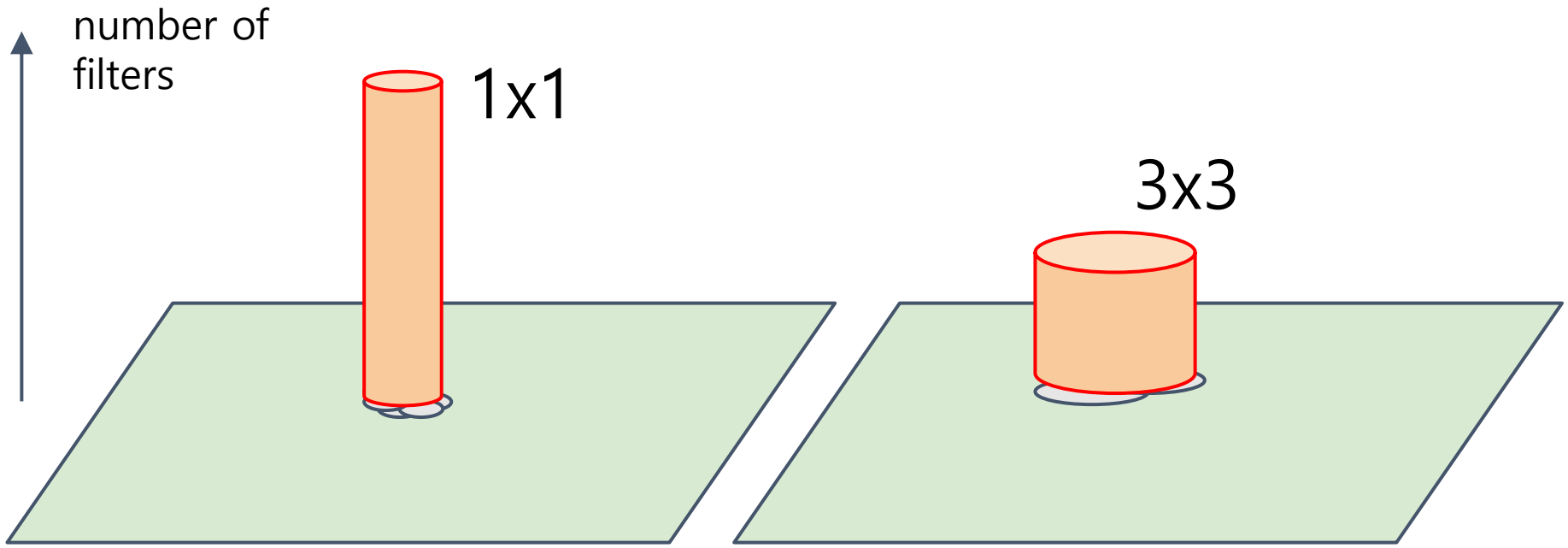
GoogLeNet

- Inception concept



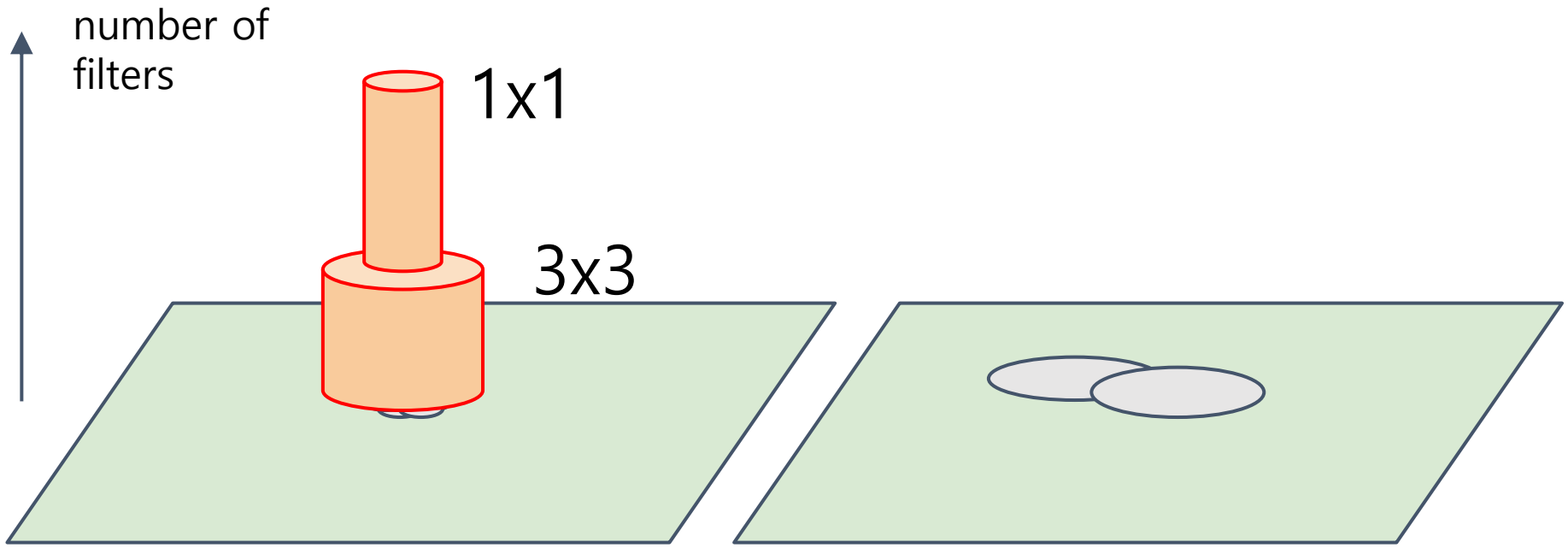
GoogLeNet

- Inception concept



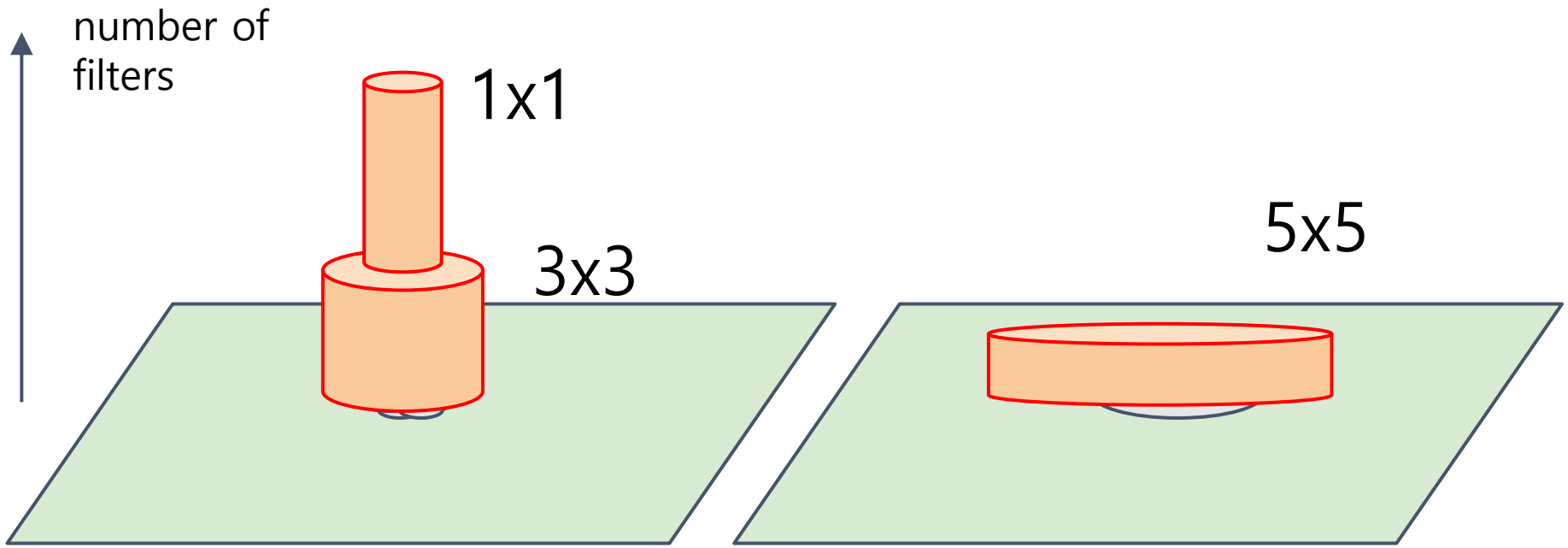
GoogLeNet

- Inception concept



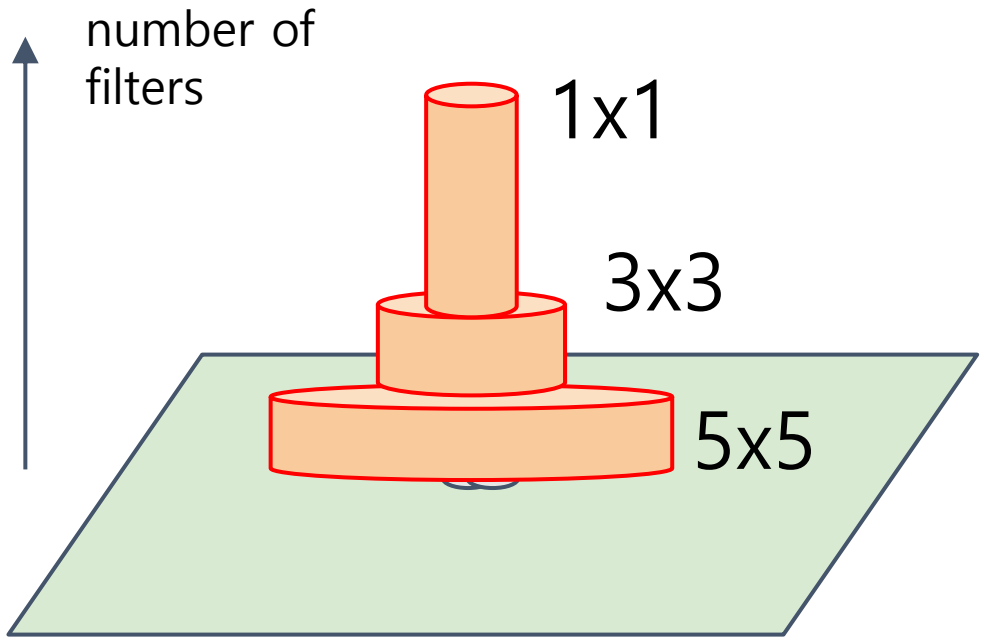
GoogLeNet

- Inception concept



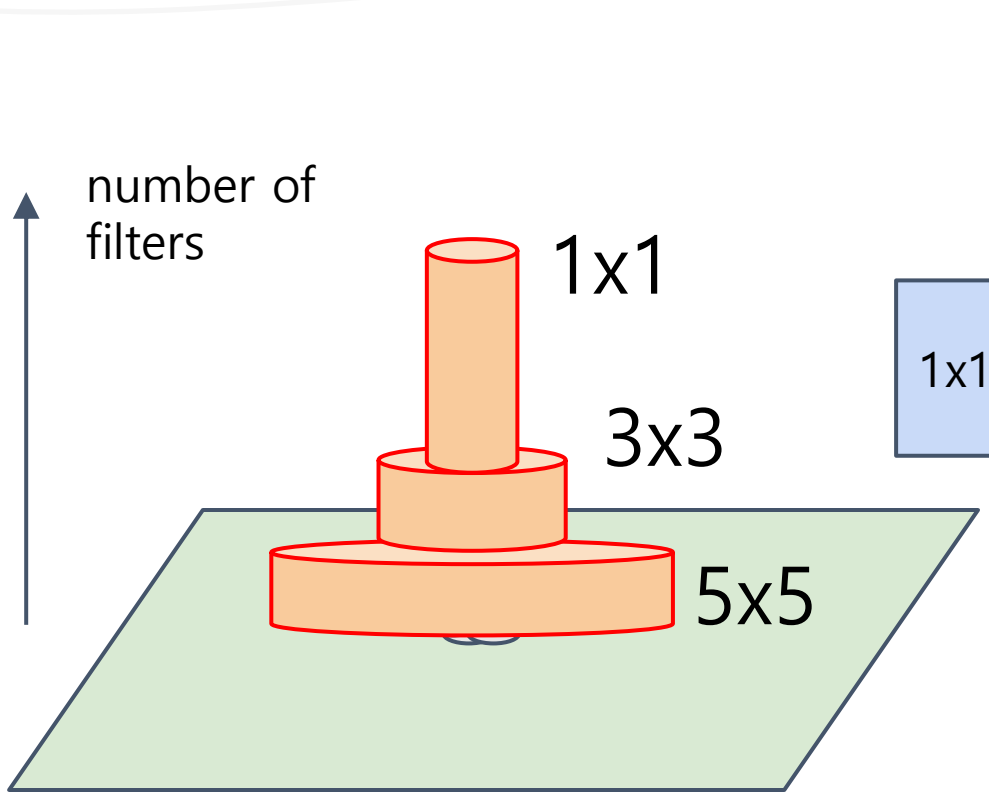
GoogLeNet

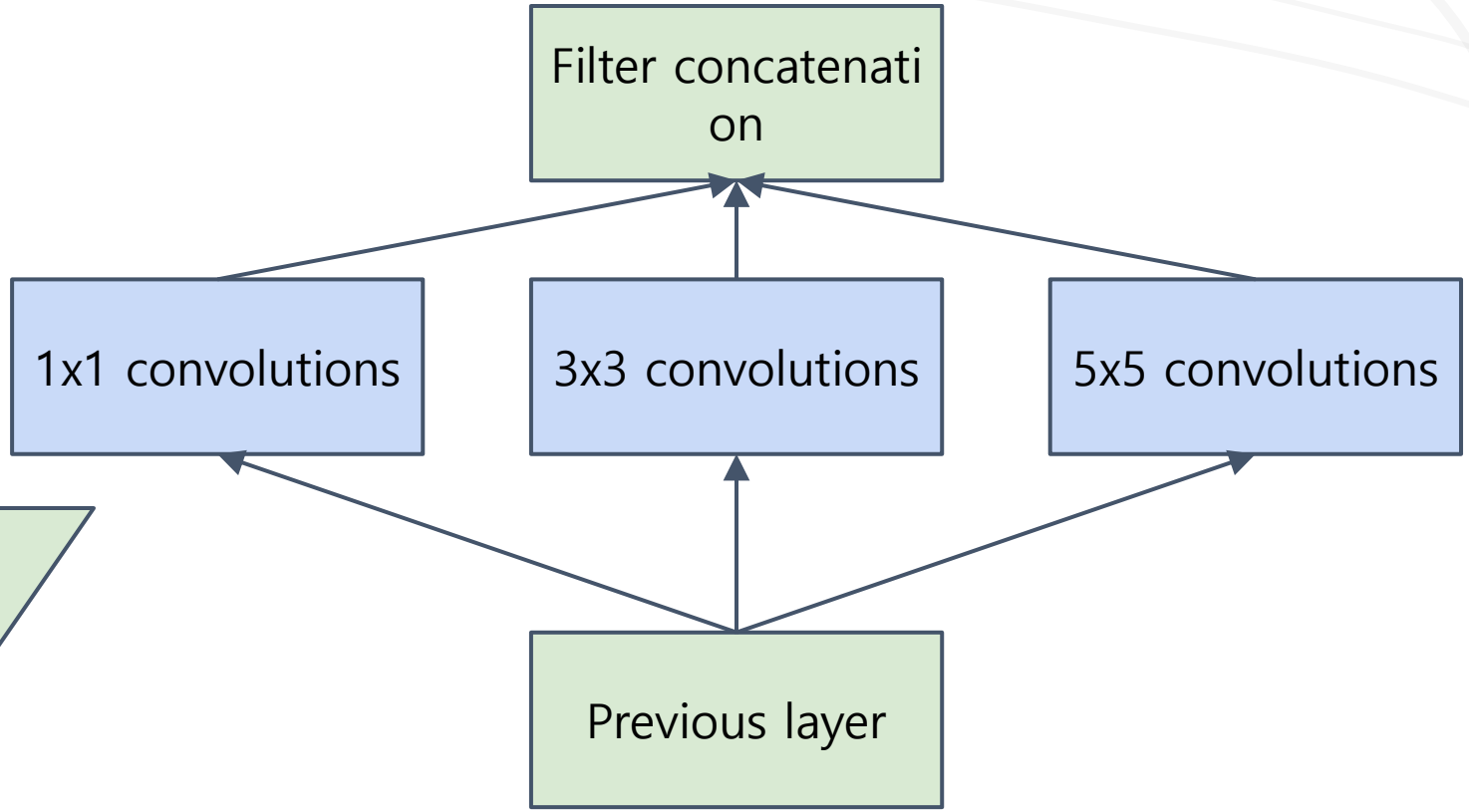
- Inception concept



GoogLeNet

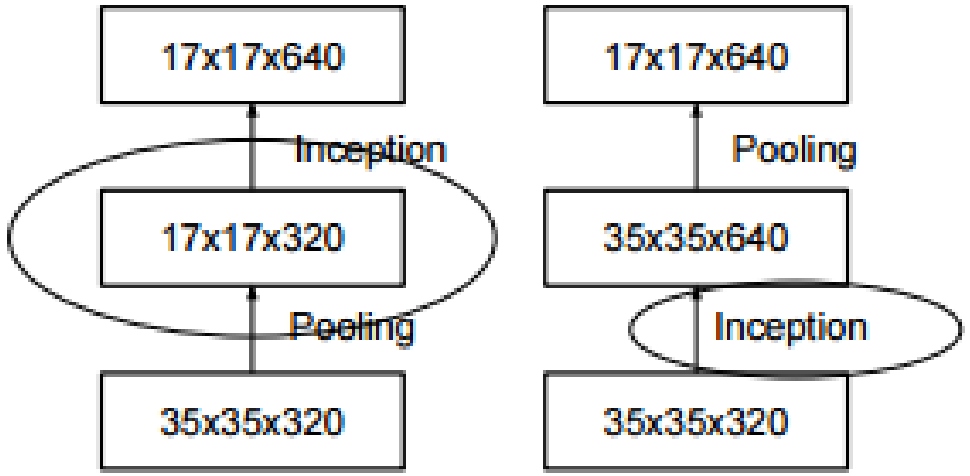
- Inception concept





GoogLeNet

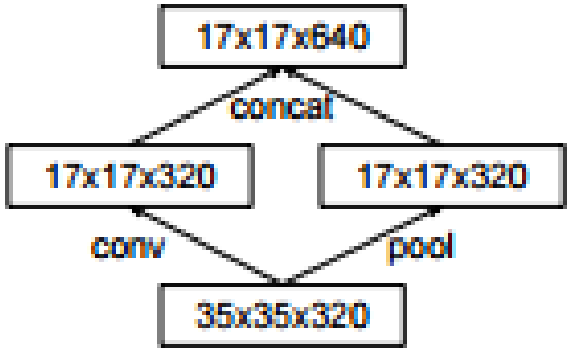
- Inception concept



Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *arXiv preprint arXiv:1512.00567* (2015).

GoogLeNet

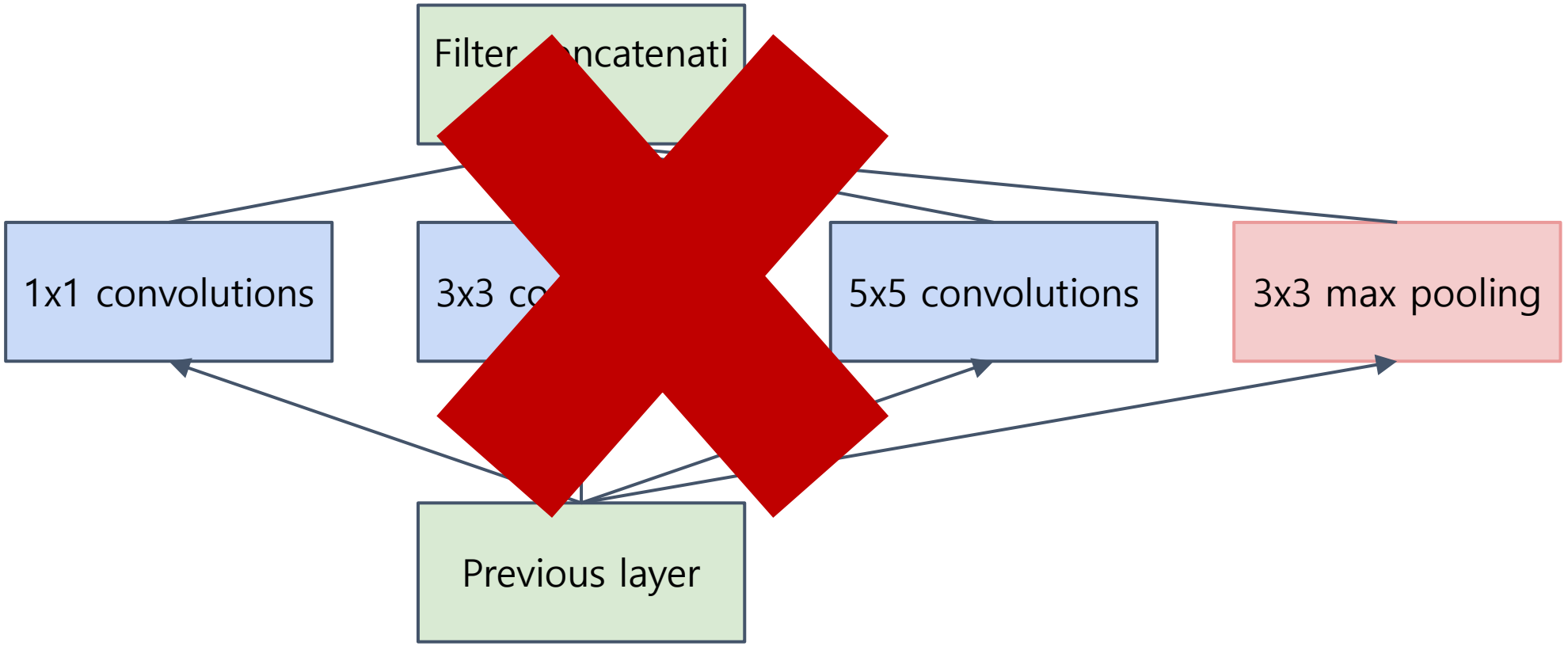
- Inception concept



Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *arXiv preprint arXiv:1512.00567* (2015).

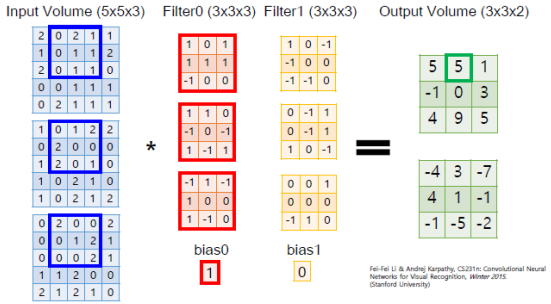
GoogLeNet

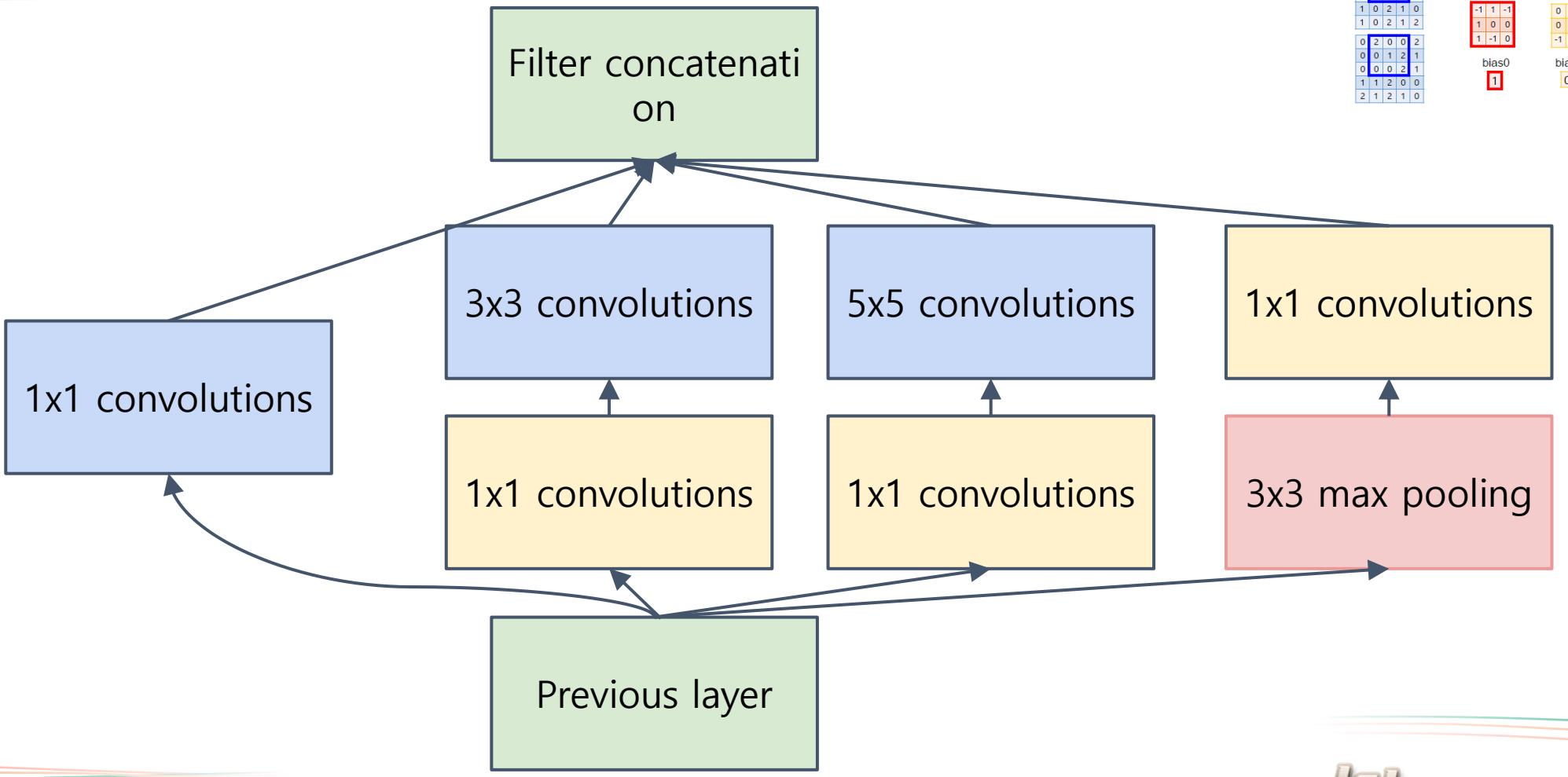
- Inception concept



GoogLeNet

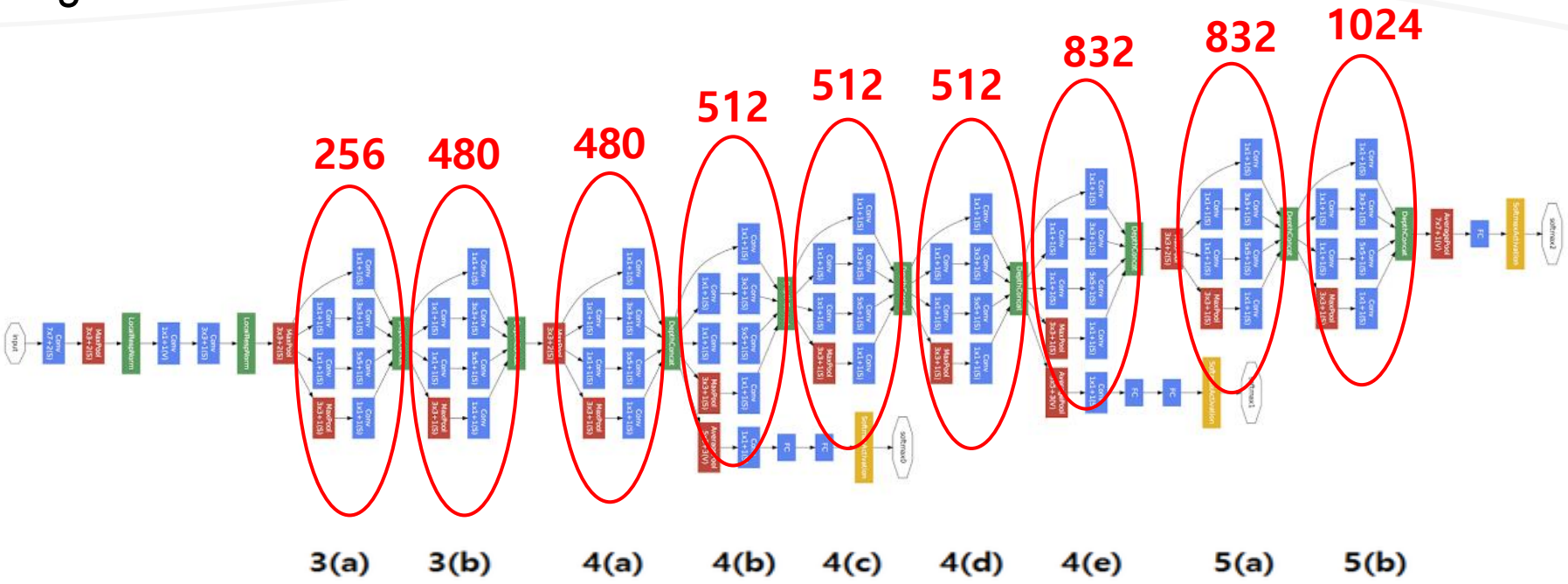
- Inception 1x1 convolution





GoogLeNet

- GoogLeNet



9 **Inception** modules

Convolution
Pooling
Softmax
Other

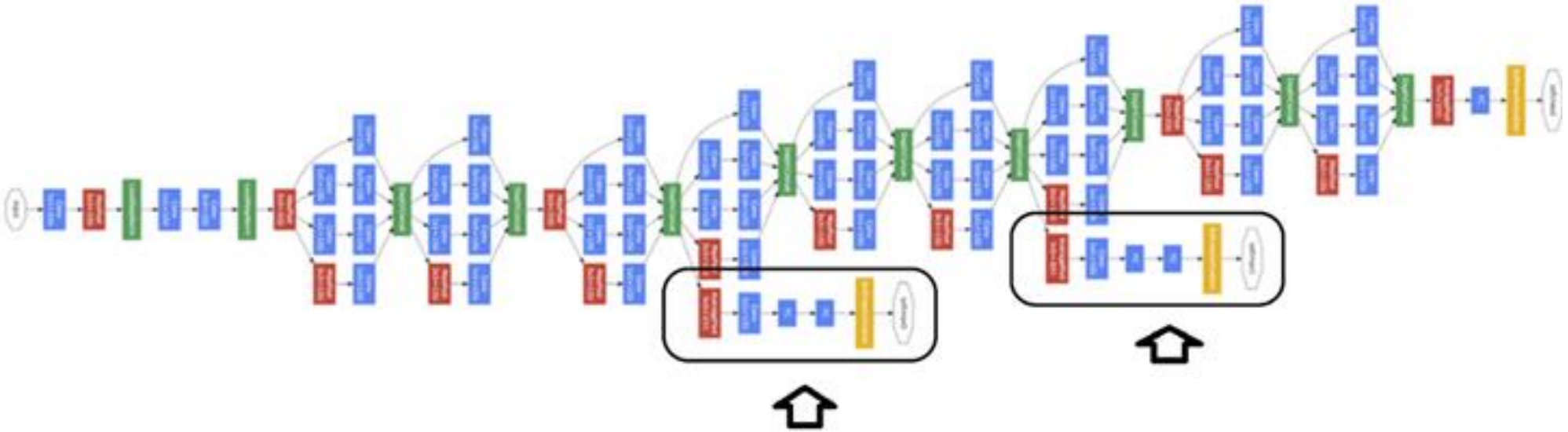
GoogLeNet

- GoogLeNet

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

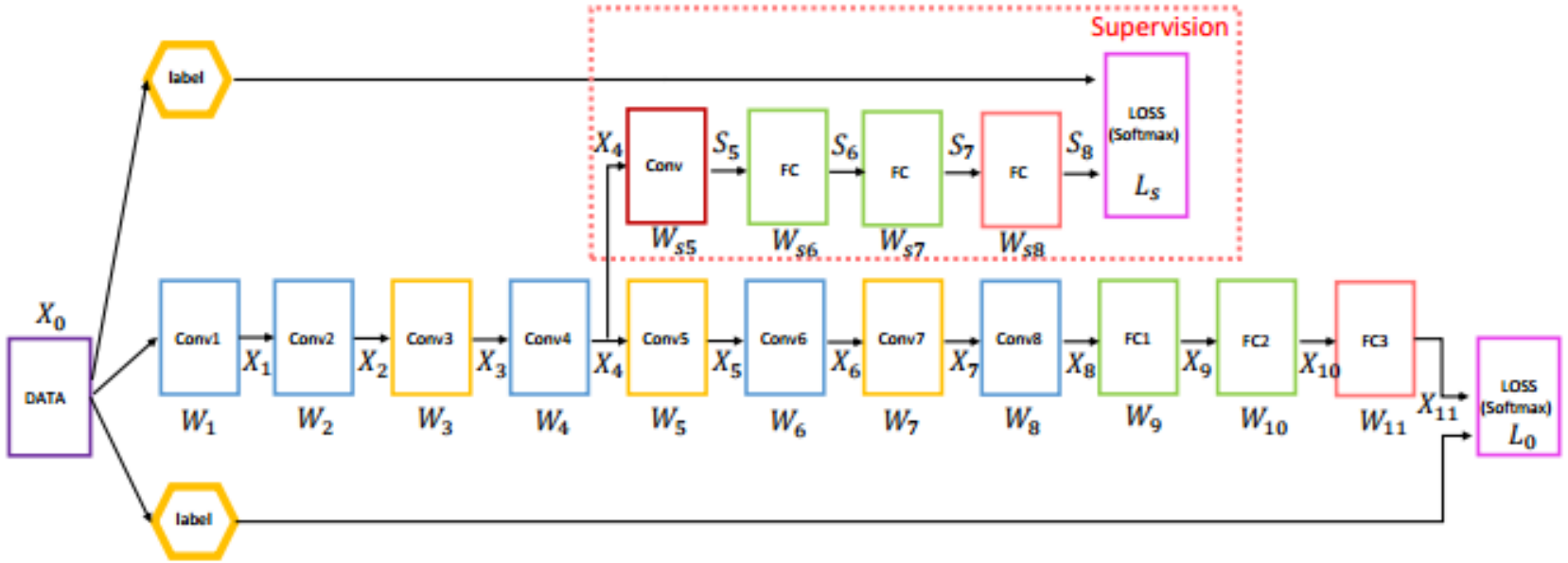
GoogLeNet

- Auxiliary classifier



GoogLeNet

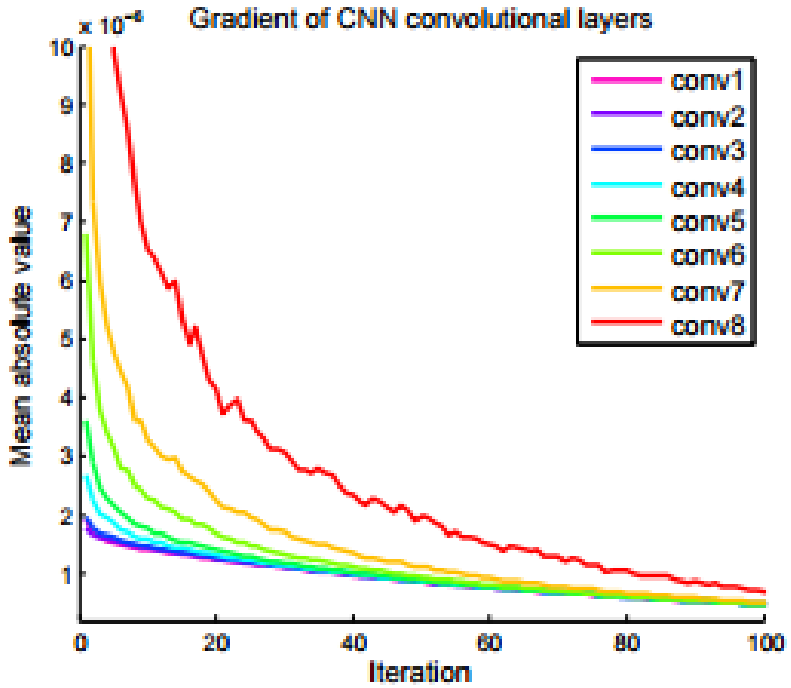
- Auxiliary classifier

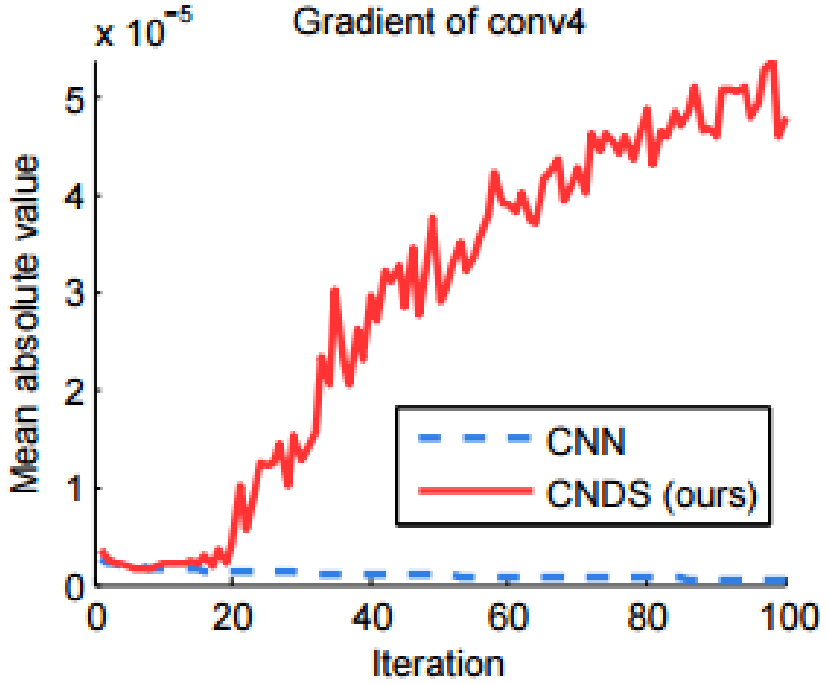


Wang, Liwei, et al. "Training deeper convolutional networks with deep supervision." *arXiv preprint arXiv:1505.02496* (2015).

GoogLeNet

- Auxiliary classifier

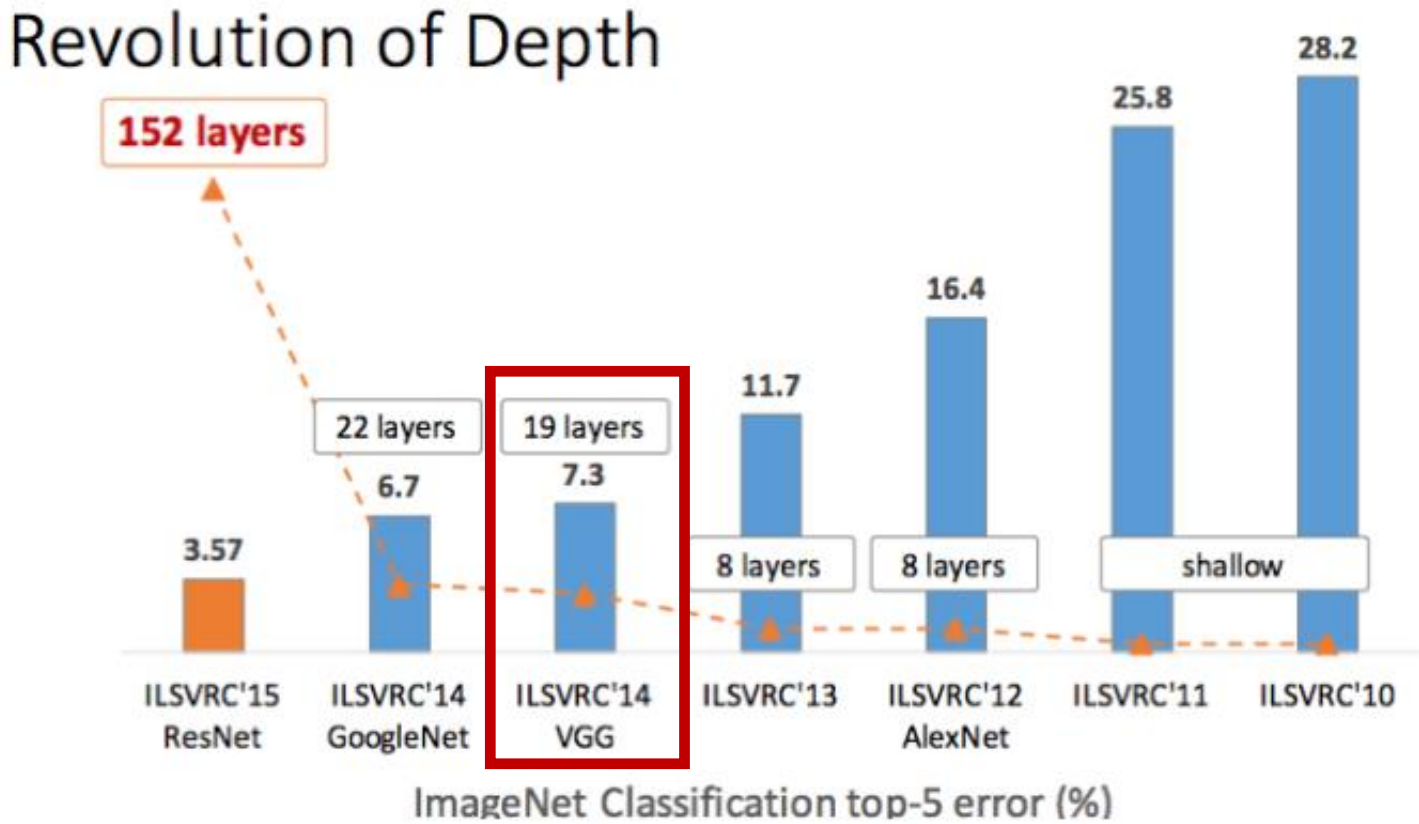




Wang, Liwei, et al. "Training deeper convolutional networks with deep supervision." *arXiv preprint arXiv:1505.02496* (2015).

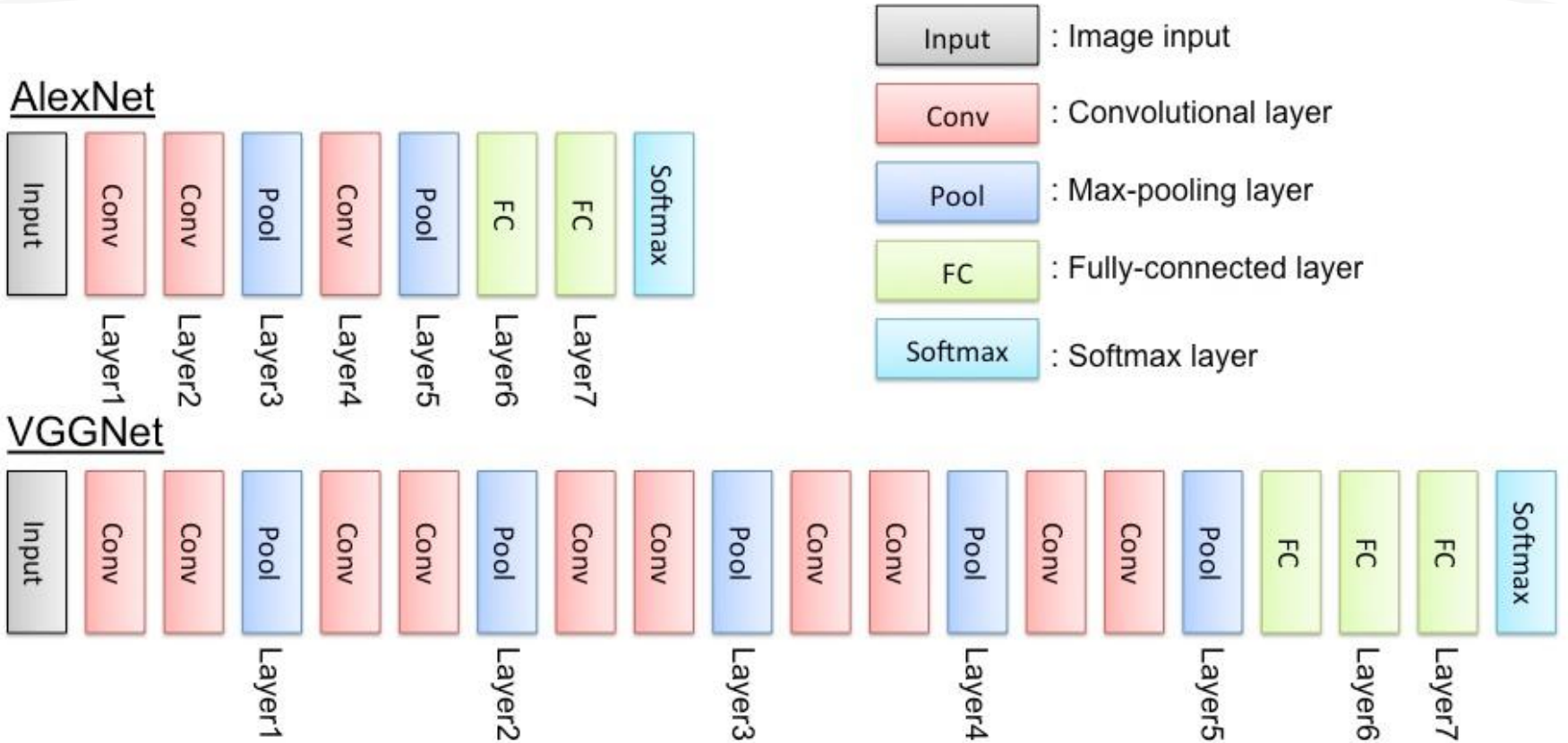
GoogLeNet

- VGGNet



GoogLeNet

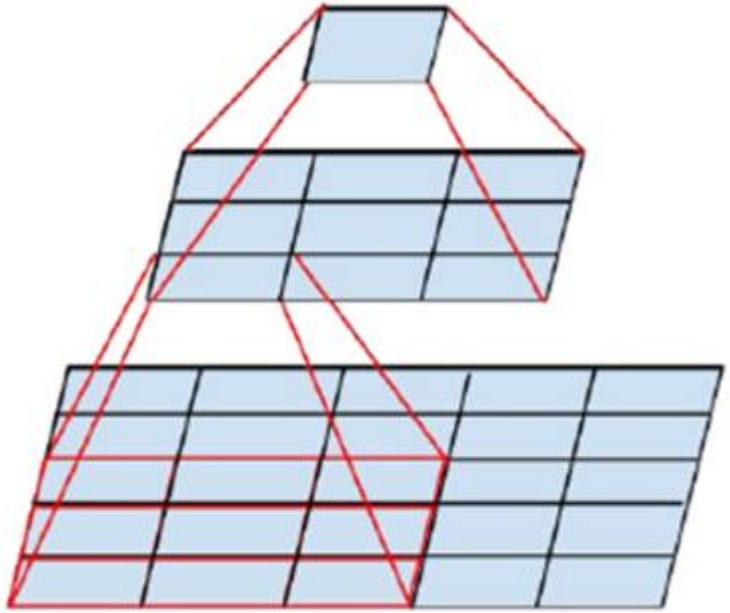
- VGGNet

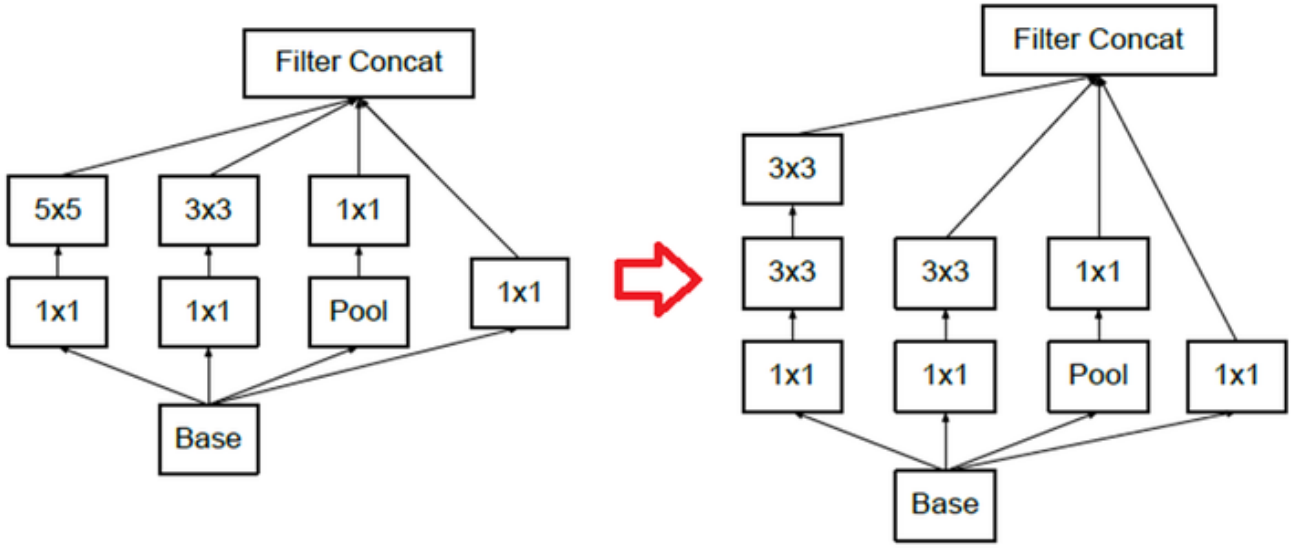


Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

GoogLeNet

- Inception – V2~ : Factoriziong Convolutions (VGG Net)





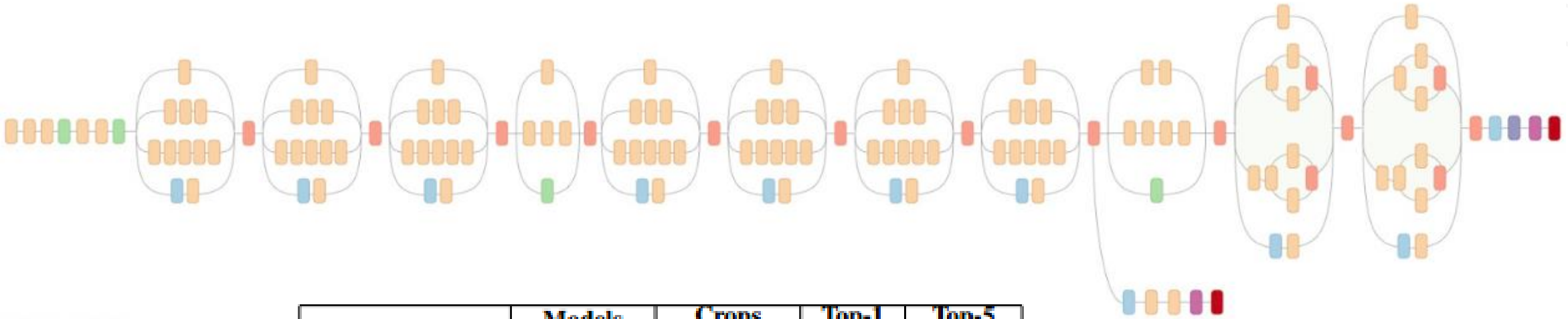
GoogLeNet

- Inception–v2

Network	Top-1 Error	Top-5 Error	Cost Bn Ops
GoogLeNet [20]	29%	9.2%	1.5
BN-GoogLeNet	26.8%	-	**1.5**
BN-Inception [7]	25.2%	7.8	2.0
Inception-v2	23.4%	-	3.8
Inception-v2 RMSProp	23.1%	6.3	3.8
Inception-v2 Label Smoothing	22.8%	6.1	3.8
Inception-v2 Factorized 7 × 7	21.6%	5.8	4.8
Inception-v2 BN-auxiliary	**21.2%**	**5.6%**	4.8

GoogLeNet

- Inception-v3

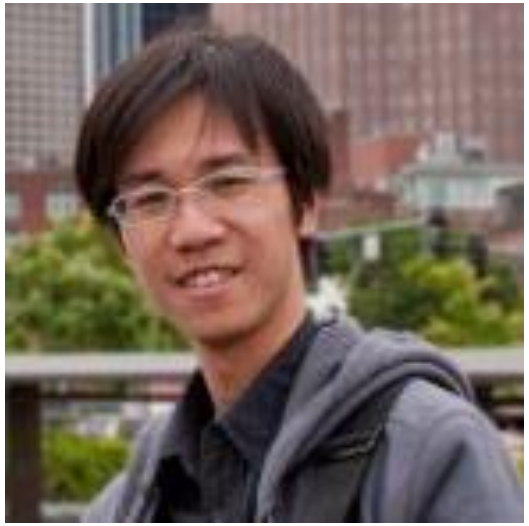


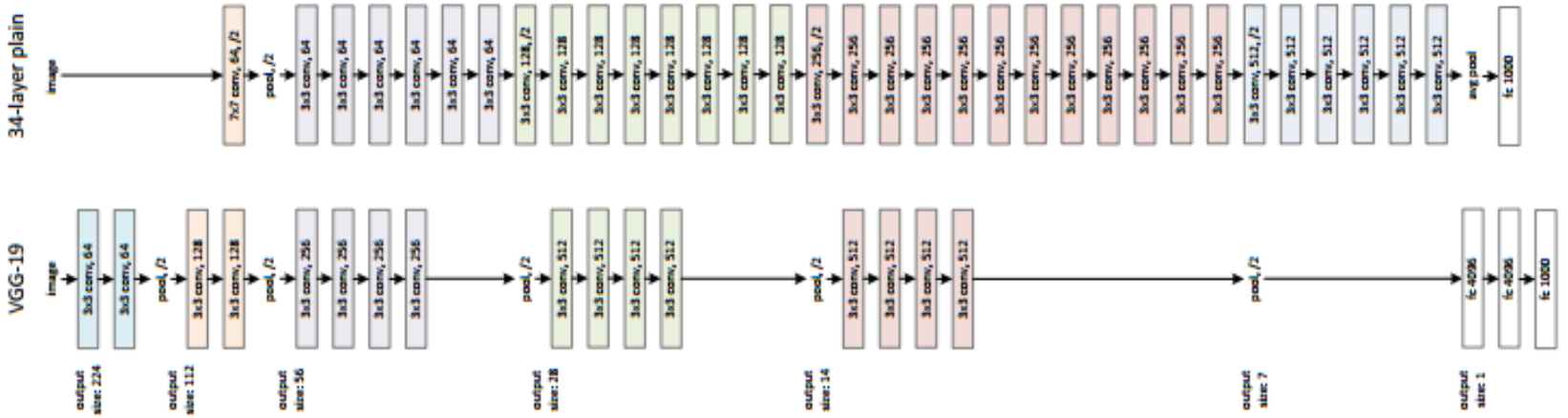
Convolution
AvgPool
MaxPool
Concat
Dropout
Fully connected
Softmax

Network	Models Evaluated	Crops Evaluated	Top-1 Error	Top-5 Error
VGGNet [18]	2	-	23.7%	6.8%
GoogLeNet [20]	7	144	-	6.67%
PReLU [6]	-	-	-	4.94%
BN-Inception [7]	6	144	20.1%	4.9%
Inception-v3	4	144	**17.2%**	**3.58%** *

GoogLeNet

- ResNet

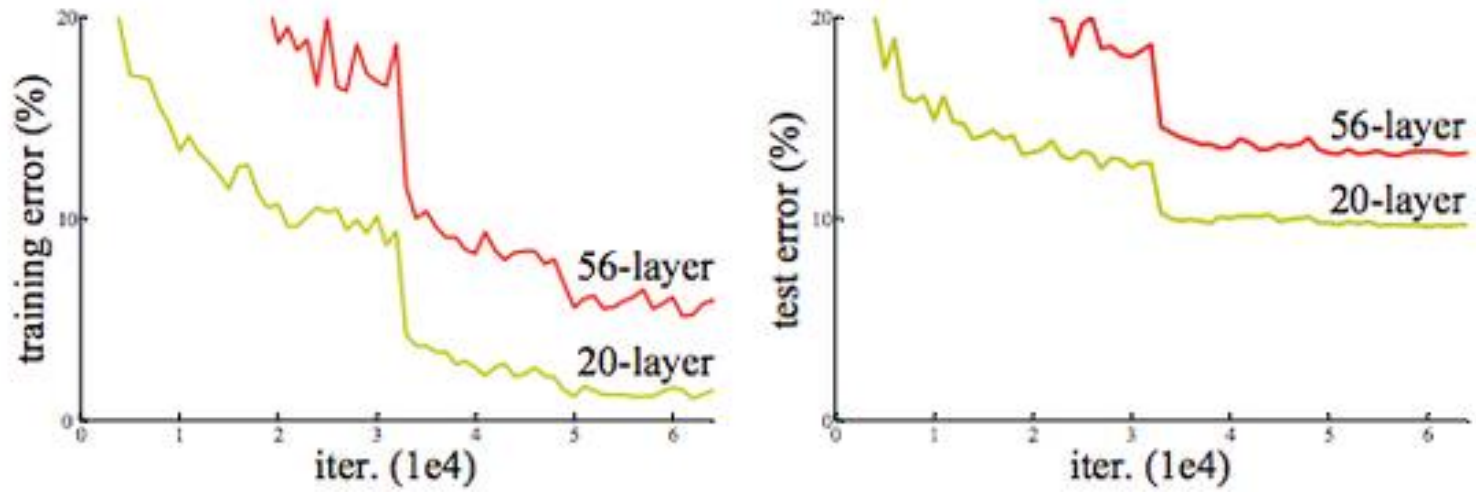




He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385* (2015).

GoogLeNet

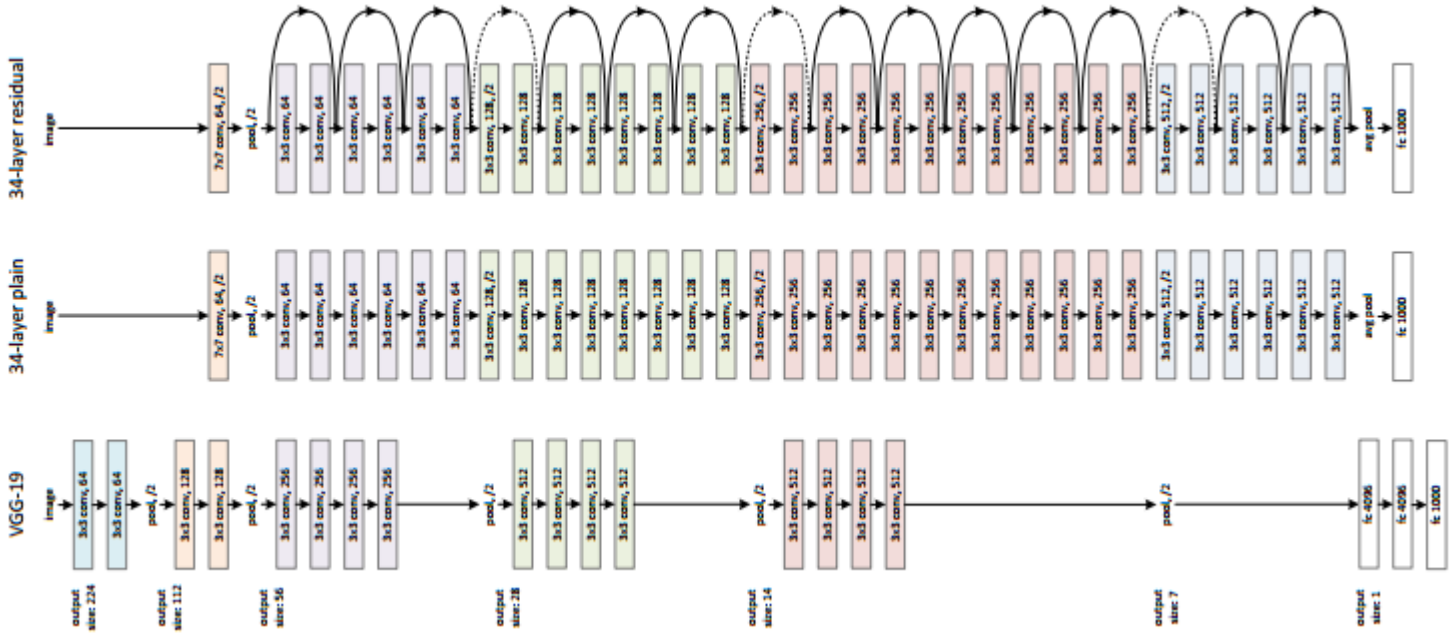
- ResNet



He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385* (2015).

GoogLeNet

- ResNet

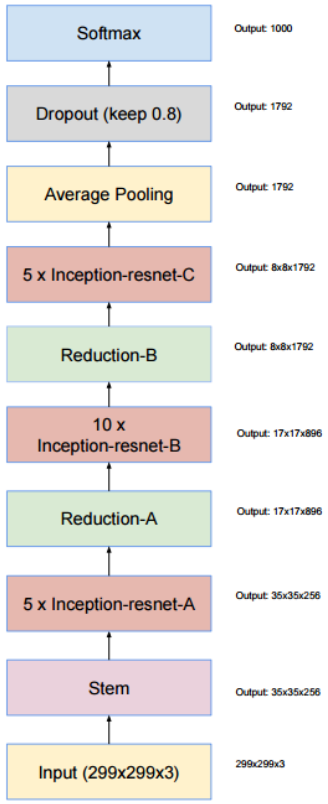


method	top-5 err. (**test**)
VGG [41] (ILSVRC'14)	7.32
GoogLeNet [44] (ILSVRC'14)	6.66
VGG [41] (v5)	6.8
PReLU-net [13]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385* (2015).

GoogLeNet

- Inception-v4 + ResNet



Network	Models	Top-1 Error	Top-5 Error
ResNet-151 [5]	6	–	3.6%
Inception-v3 [15]	4	17.3%	3.6%
Inception-v4 + 3× Inception-ResNet-v2	4	16.5%	3.1%

Szegedy, Christian, Sergey Ioffe, and Vincent Vanhoucke. "Inception-v4, inception-resnet and the impact of residual connections on learning." *arXiv preprint arXiv:1602.07261* (2016).

GoogLeNet

- Classification failure cases

Groundtruth: ????

GoogLeNet

- Classification failure cases

Groundtruth: **coffee mug**

GoogLeNet

- Classification failure cases

Groundtruth: **coffee mug**

GoogLeNet:

- **table lamp**
- **lamp shade**
- **printer**
- **projector**
- **desktop computer**

GoogLeNet

- Classification failure cases

Groundtruth: ???

GoogLeNet

- Classification failure cases

Groundtruth: **hay**

GoogLeNet

- Classification failure cases

Groundtruth: **hay**

GoogLeNet:

- **sorrel (horse)**
- **hartebeest**
- **Arabian camel**
- **warthog**
- **gazelle**

Q&A