Going deeper with convolutions

CVPR 2015

GIST EECS
윤준영

Paper Info

Paper : Going deeper with convolutions

Authors : Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet,
Scott Reed, Dragomir Anguelov, Dumitru Erhan,
Vincent Vanhoucke, Andrew Rabinovich

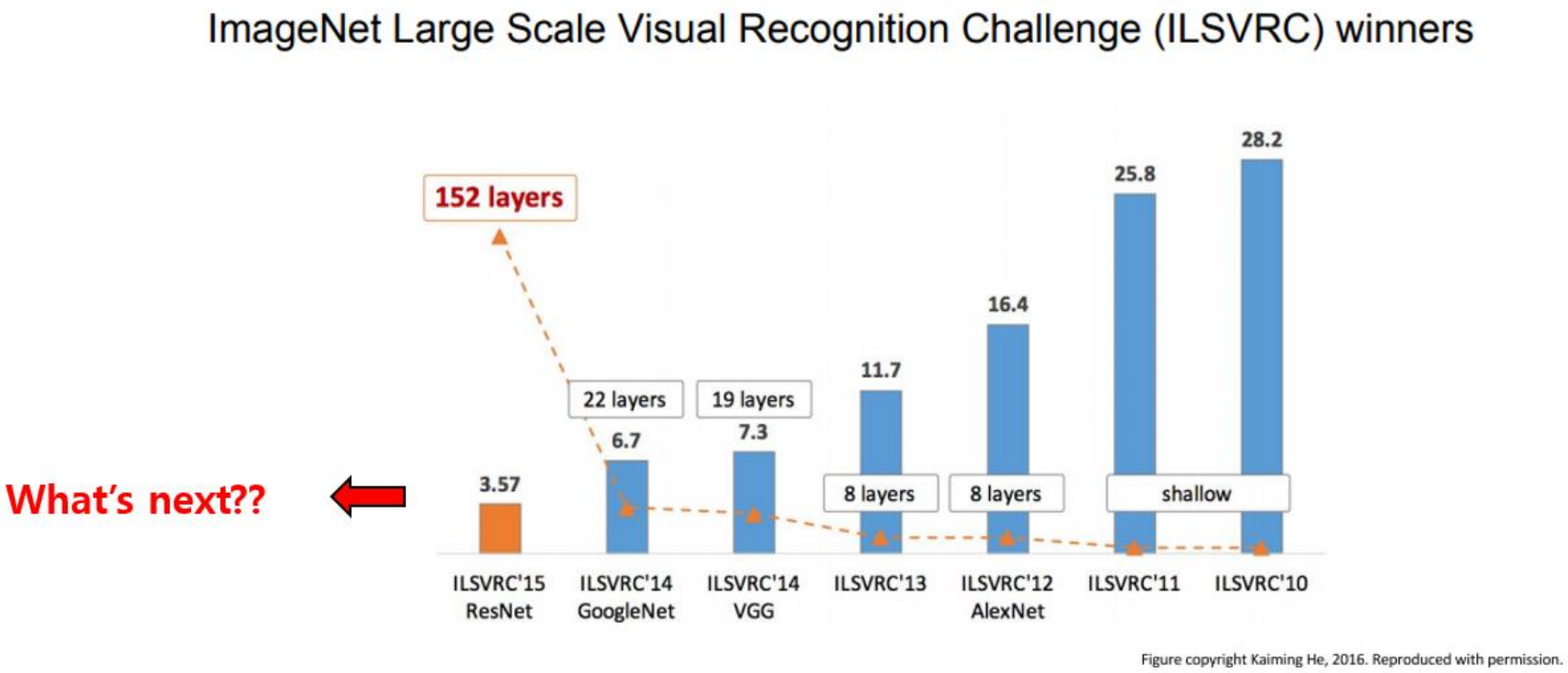
Journal : Computer Vision and Pattern Recognition (CVPR)

Citations : 27825

Main Problem

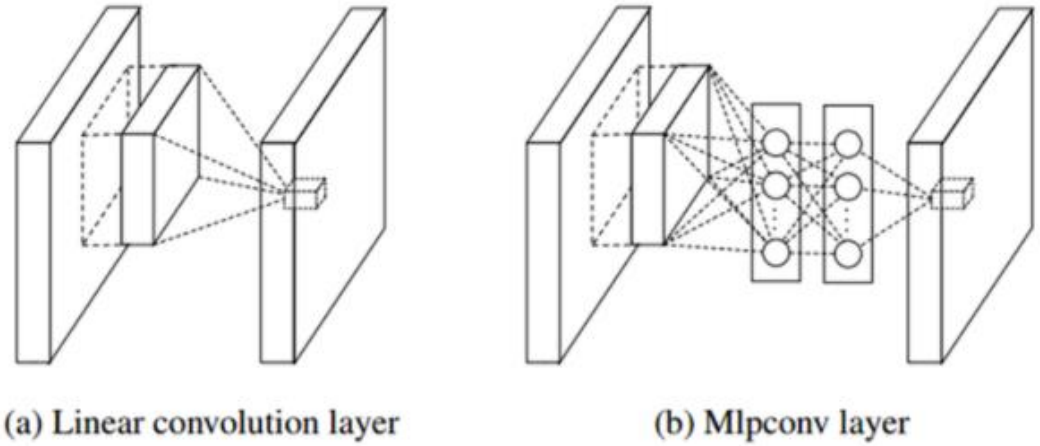
Network should go deeper

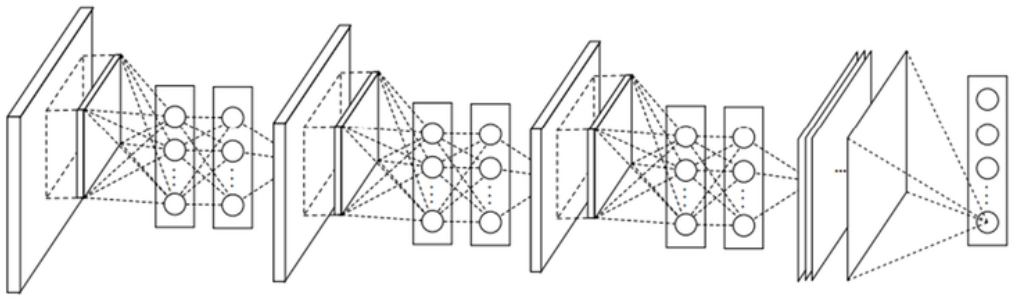
- Deeper -> too much computational budget



Background

NIN(Network In Network)

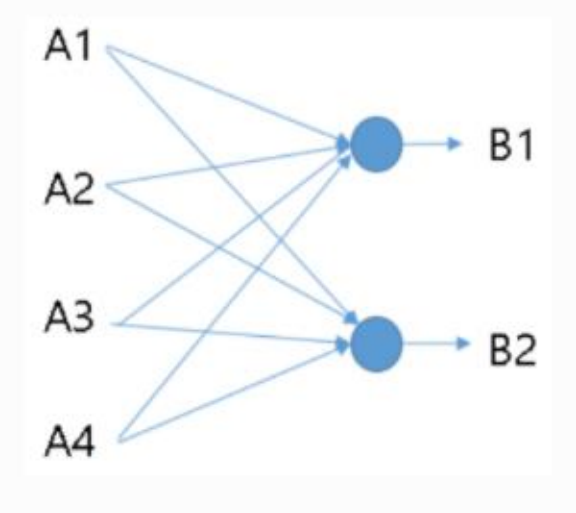


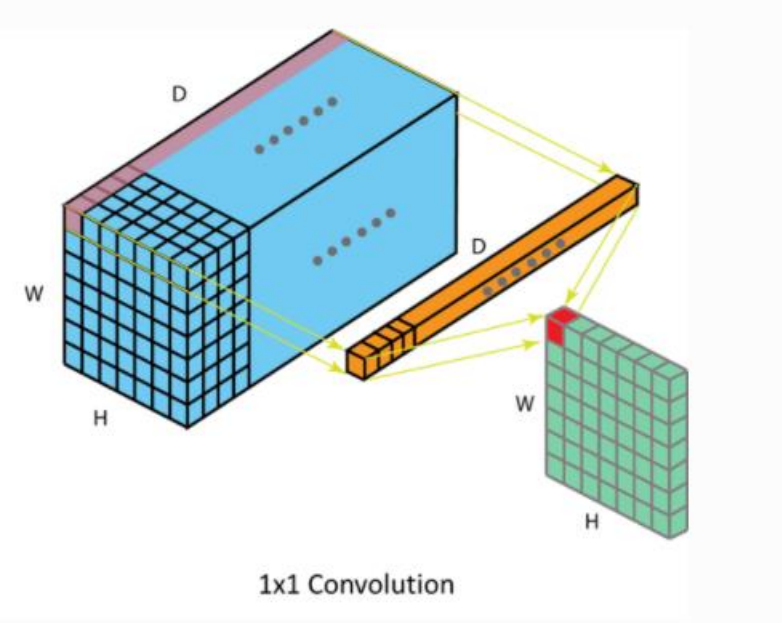


- To get non-linear feature
- Used MLP(Multi-Layer Perceptron) instead of filter
- 1x1 convolution -> dimension reduction
- Use Global average pooling(GAP)

Background

1x1 convolution

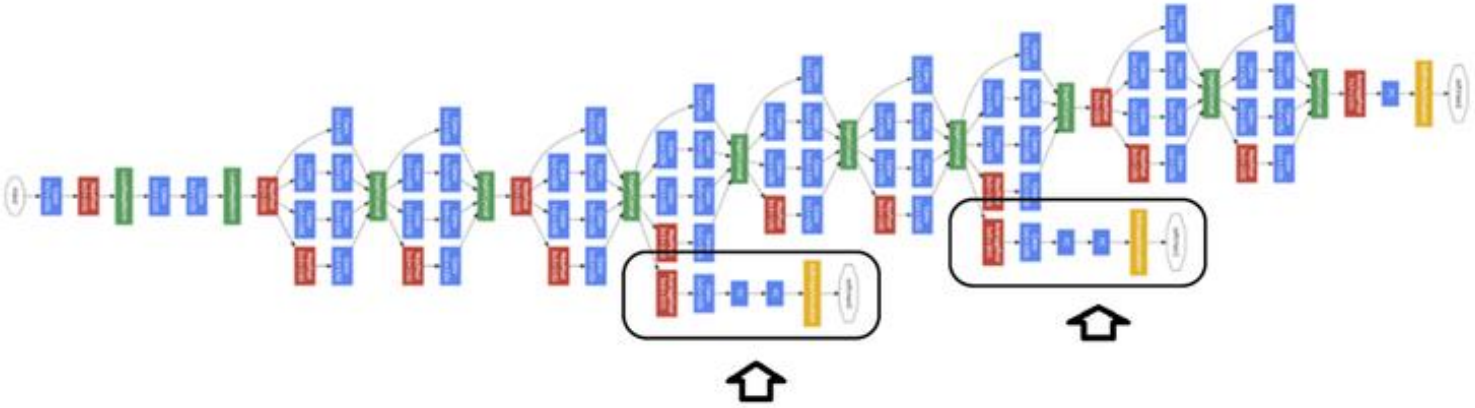




- dimension reduction
- feature–map decrease
- Free parameter decrease
- Use RELU to increase non-linearity

Background

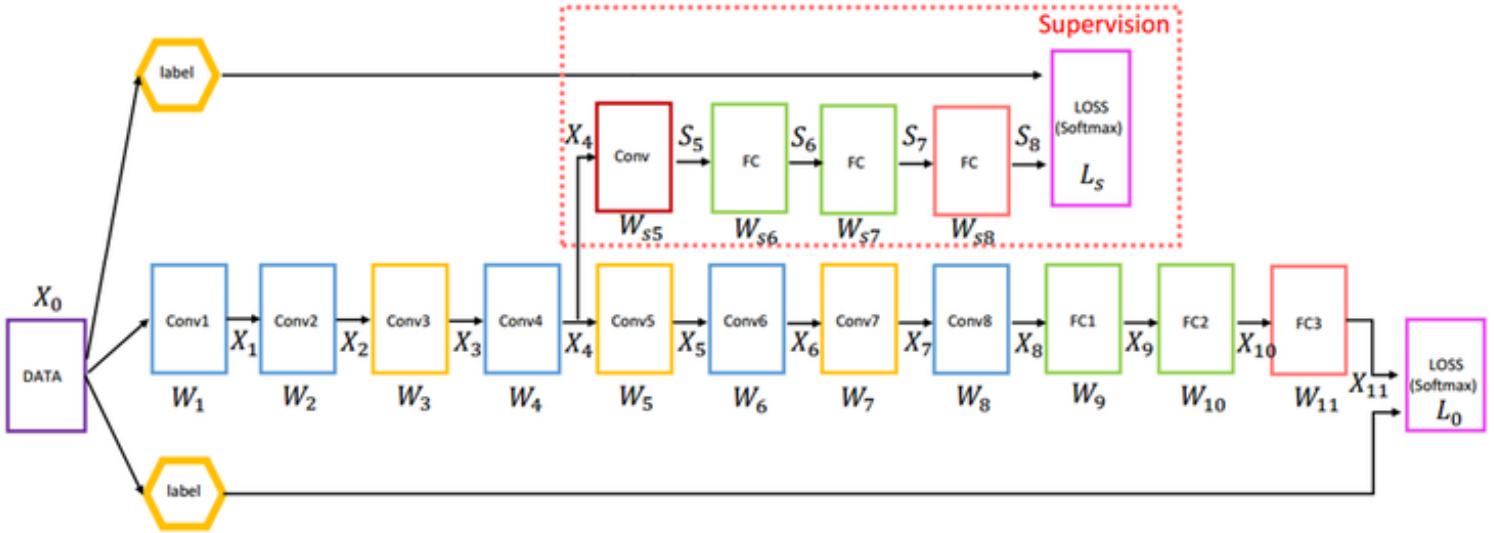
Auxiliary classifier

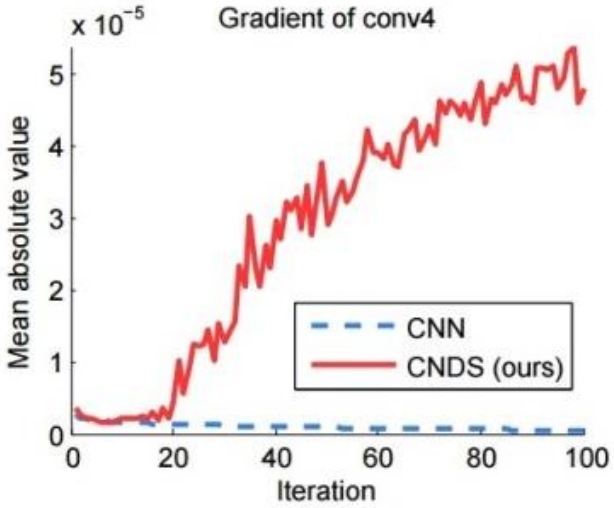


- Vanishing gradient problem
- Back-propagation result is returned in classifier
- Increases the gradient signal

Background

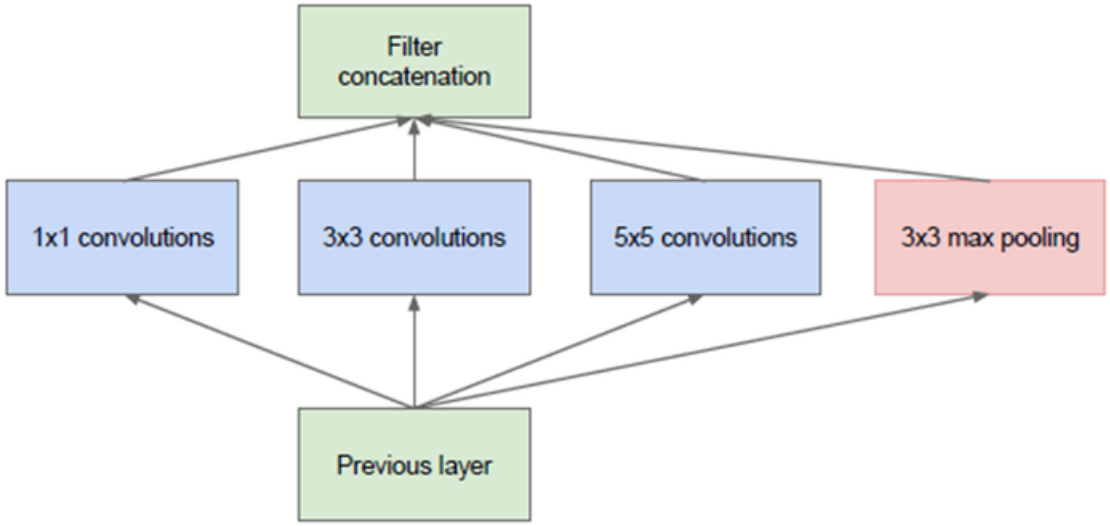
Auxiliary classifier





- Training Deeper Convolutional Networks with Deep SuperVision – Liwei Wang
- Gradient signal changes dramatically
- It is only used on training, it is removed in Test

Architectural detail



(a) Inception module, naïve version

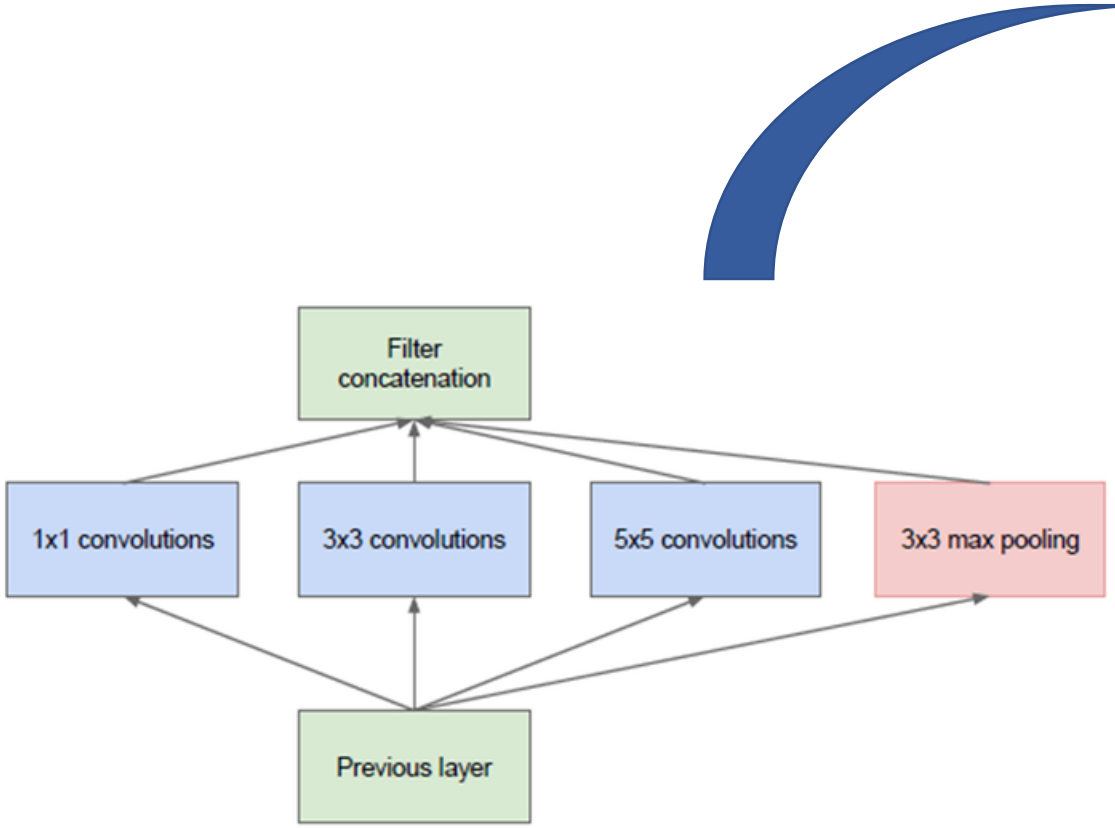
- native version
  - Arora et al suggestion

    -> "analyze the correlation statics of the last layer and cluster them into group of units with high correlation"
  - Used several convolution parallel
  - To avoid patch-alignment issue

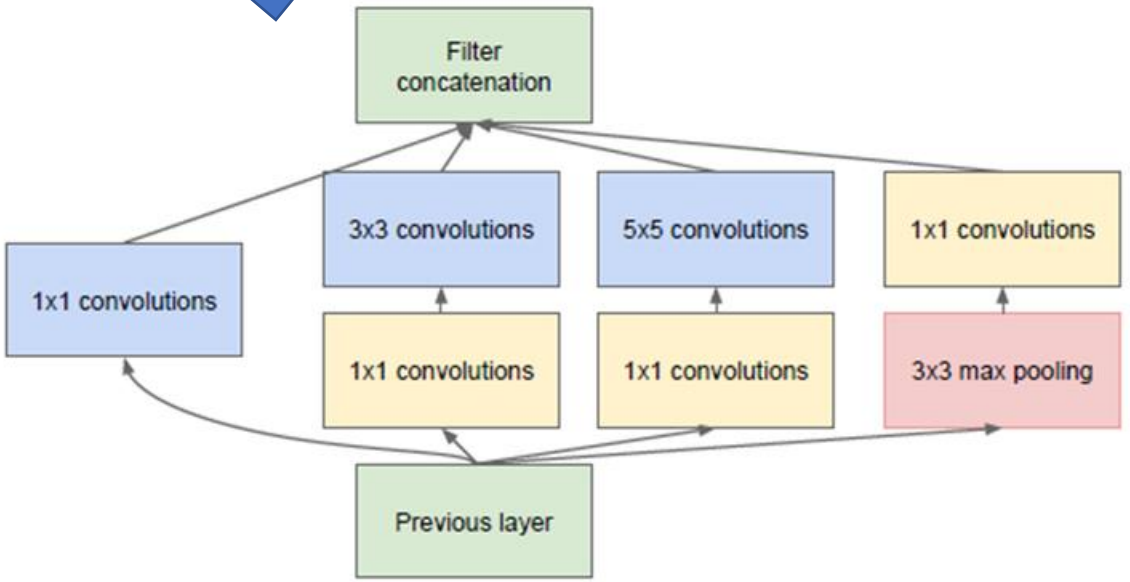
    -> 1x1, 3x3, 5x5 filter is used

Architectural detail

3x3, 5x5 filter is an expensive unit



(a) Inception module, naïve version

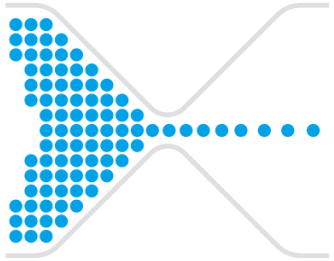


(b) Inception module with dimensionality reduction

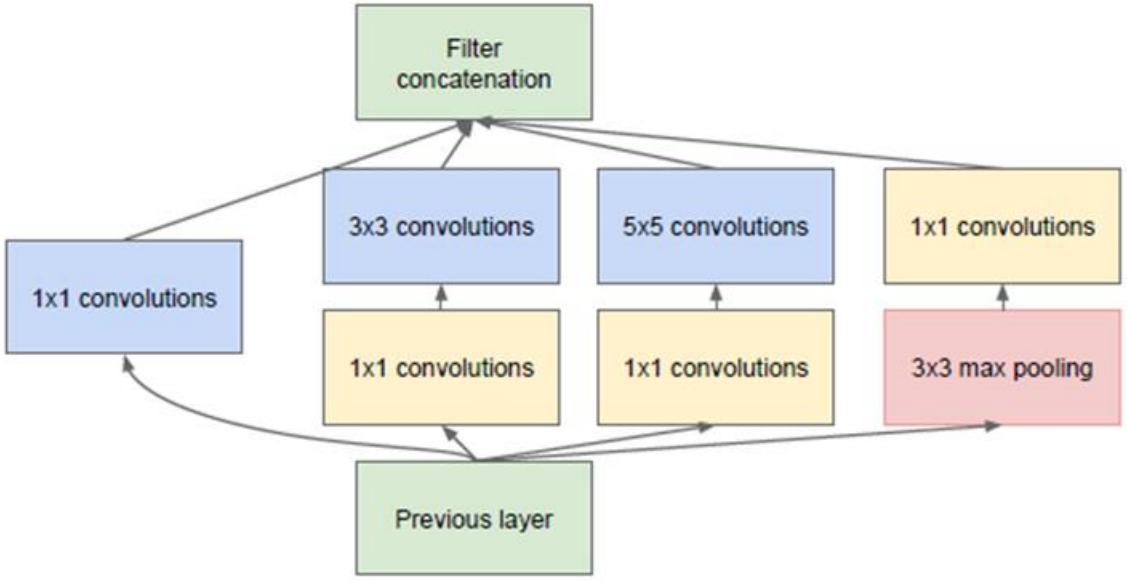
Architectural detail

- Bottleneck structure

  Deeper, free parameter reduce



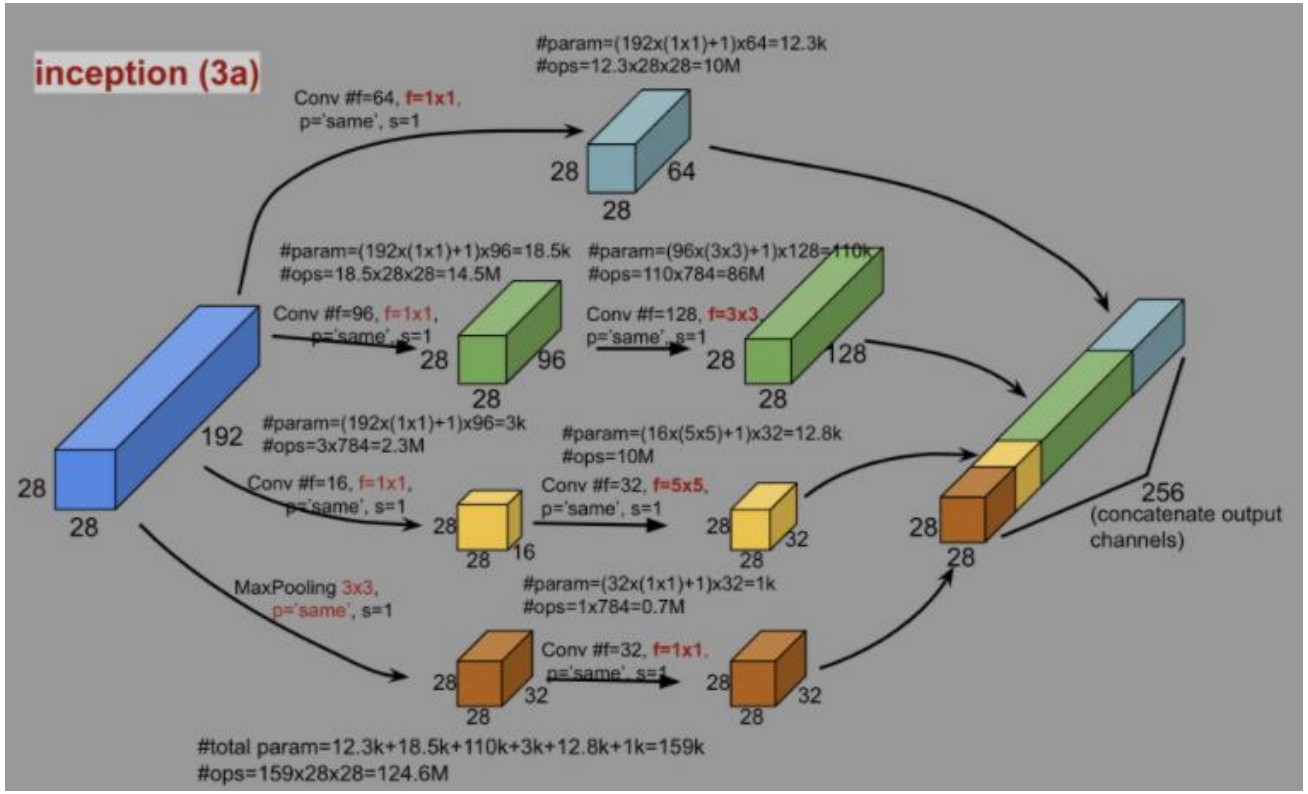
- Rectified linear activation



(b) Inception module with dimensionality reduction

Architectural detail

- 1x1 convolution
  - 28x28 feature map total 64
- 3x3 convolution
  - 28x28 feature map total 128
- 5x5 convolution
  - 28x28 feature map total 32
- Max pooling
  - 28x28 feature map total 32

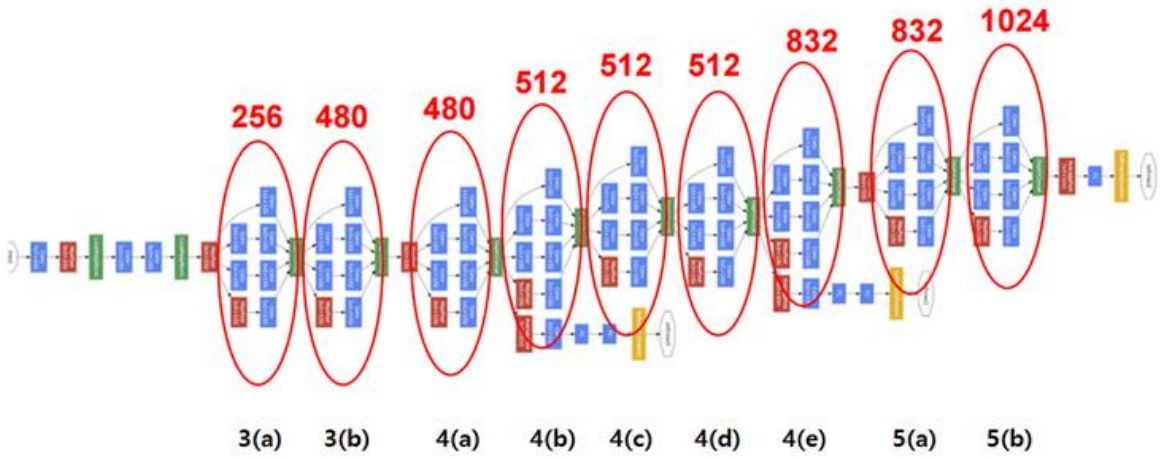


GoogLeNet

Total 9 Inception
22 layers

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								

Table 1: GoogLeNet incarnation of the Inception architecture

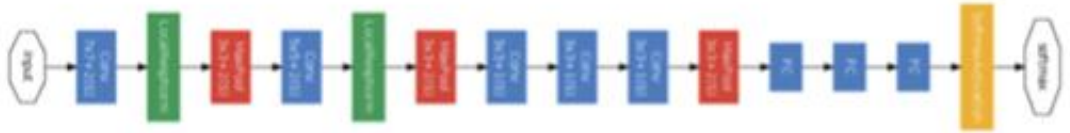


- Red circle: Inception
- Blue module: Convolution layer
- Red module: Max pooling
- Yellow module: Softmax layer
- Green module: etc function

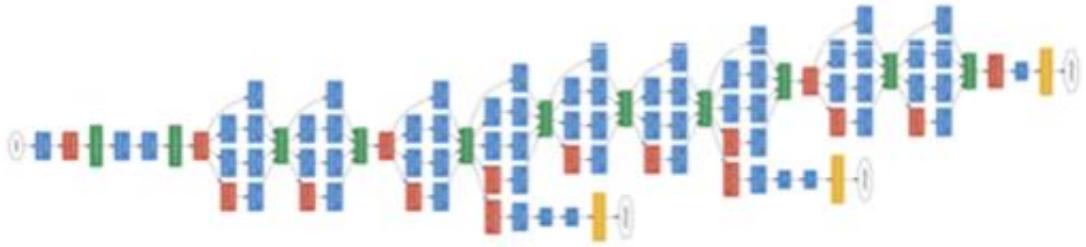
GoogLeNet

Deeper & 1/12 free parameter

	params	*FLOPs*
Krizhevsky, Suskever and Hinton (2012)	60M	2B



Szegedy et al (2014)	5M	1.5B



Result

- ILSVRC 2014

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Classification performance

Team	Year	Place	mAP	external data	ensemble	approach
UvA-Euvision	2013	1st	22.6%	none	?	Fisher vectors
Deep Insight	2014	3rd	40.5%	ImageNet 1k	3	CNN
CUHK DeepID-Net	2014	2nd	40.7%	ImageNet 1k	?	CNN
GoogLeNet	2014	1st	43.9%	ImageNet 1k	6	CNN

Detection performance

Conclusion

- Sparse structure by readily available dense building
-> improve NN for computer vision

- Verify the strength of Inception architecture

- Suggest the way to DNN