Visualizing CNNs and Deeper Deep Architectures

Visualizing Deep Networks

See slides here:
http://places.csail.mit.edu/slide_iclr2015.pdf

Beyond AlexNet

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan & Andrew Zisserman 2015

**These are the "VGG" networks.
Including what you use in Project 6**

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					

Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**

Going Deeper with Convolutions

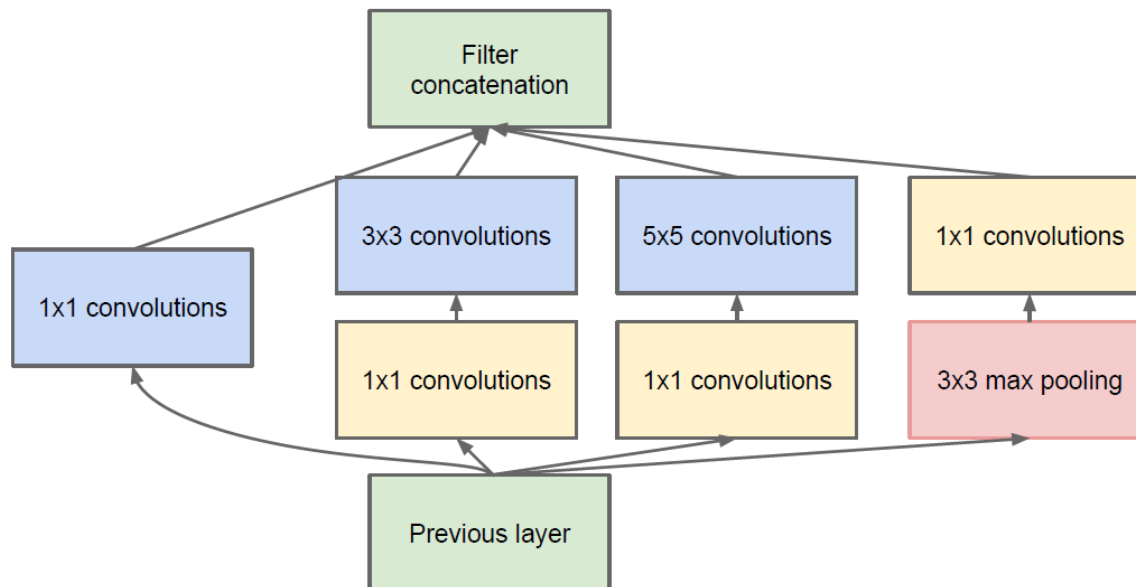**Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich**
2015

This is the "Inception" architecture or "GoogLeNet"

*The architecture blocks are called "Inception" modules and the collection of them into a particular net is "GoogLeNet"
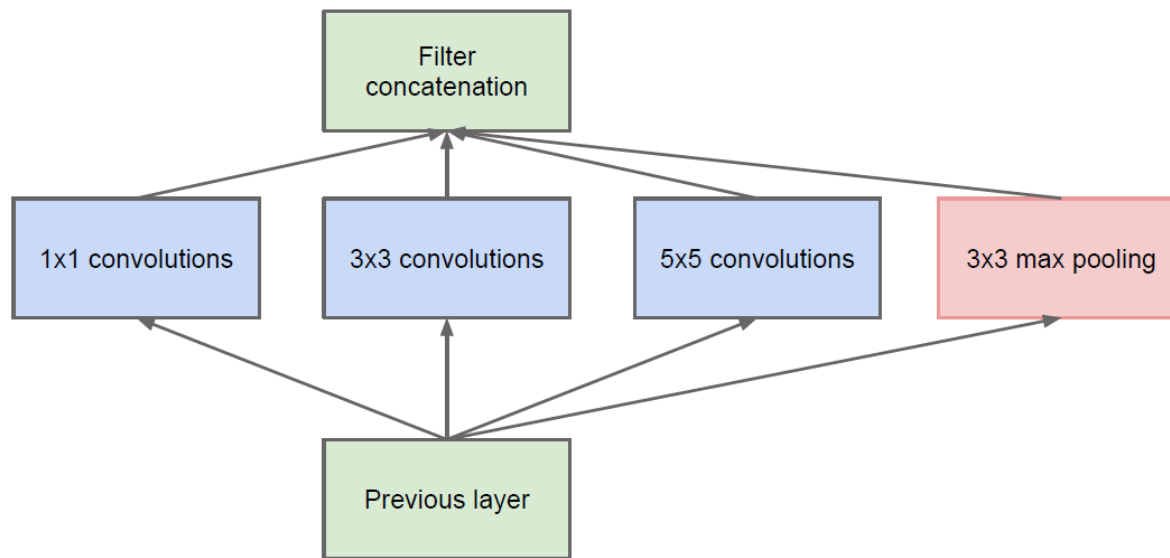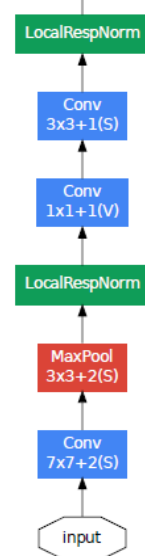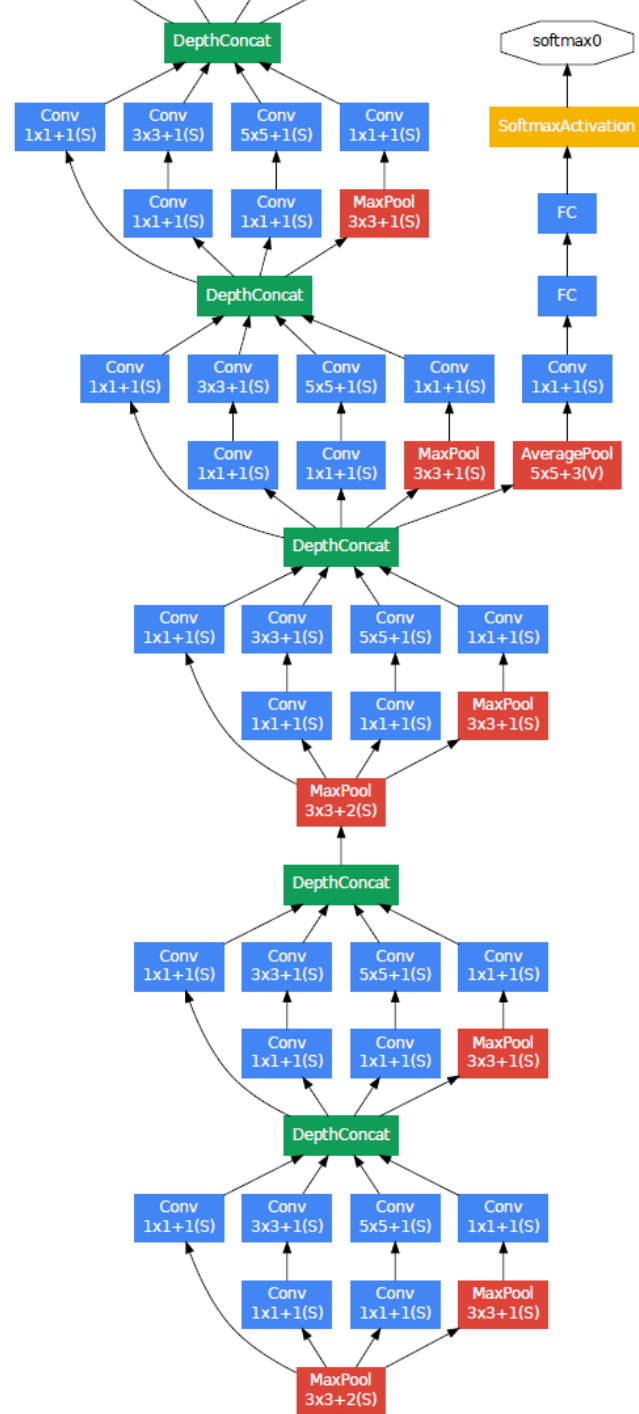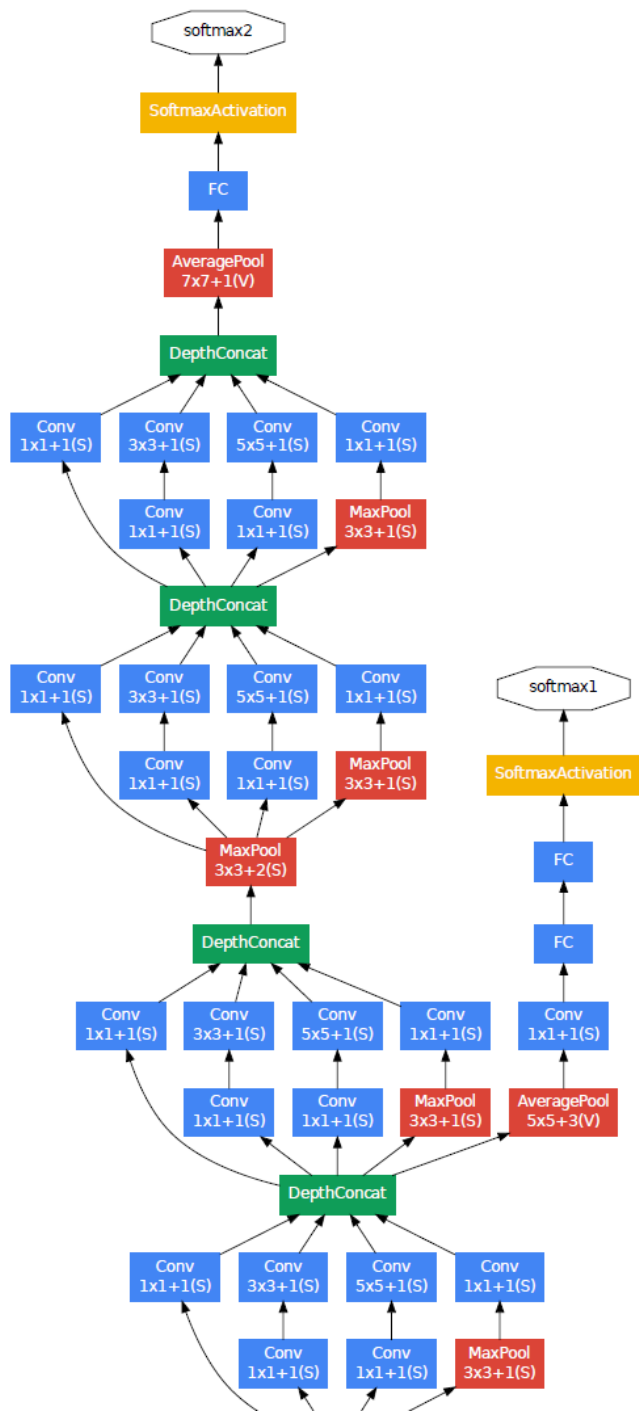
(a) Inception module, naïve version

(b) Inception module with dimensionality reduction

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								

Only 6.8 million parameters. AlexNet ~60 million, VGG up to 138 million

softmax2

SoftmaxActivation

FC

AveragePool
7x7+1(V)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

MaxPool
3x3+2(S)

softmax1

SoftmaxActivation

FC

FC

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)   AveragePool
5x5+3(V)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

MaxPool
3x3+2(S)

softmax0

SoftmaxActivation

FC

FC

Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)   AveragePool
5x5+3(V)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

DepthConcat

Conv
1x1+1(S)   Conv
3x3+1(S)   Conv
5x5+1(S)   Conv
1x1+1(S)

Conv
1x1+1(S)   Conv
1x1+1(S)   MaxPool
3x3+1(S)

MaxPool
3x3+2(S)

LocalRespNorm

Conv
3x3+1(S)

Conv
1x1+1(V)

LocalRespNorm

MaxPool
3x3+2(S)

Conv
7x7+2(S)

input

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

Table 2: Classification performance.

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%

Surely it would be ridiculous to go any deeper…

- ResNet

Surely it would be ridiculous to go any deeper…

- ResNet
- See the following slides: http://kaiminghe.com/cvpr16resnet/cvpr2016_deep_residual_learning_kaiminghe.pdf