

Open Elective Course [OE]

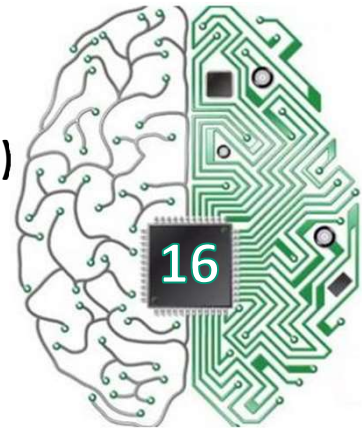
Course Code: CSO507

Winter 2023-24

Lecture#

Deep Learning

Unit-4: Convolutional Neural Networks (Part-IV)

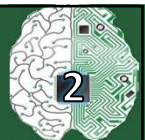
**Course Instructor:**

Dr. Monidipa Das

Assistant Professor

Department of Computer Science and Engineering

Indian Institute of Technology (Indian School of Mines) Dhanbad, Jharkhand 826004, India



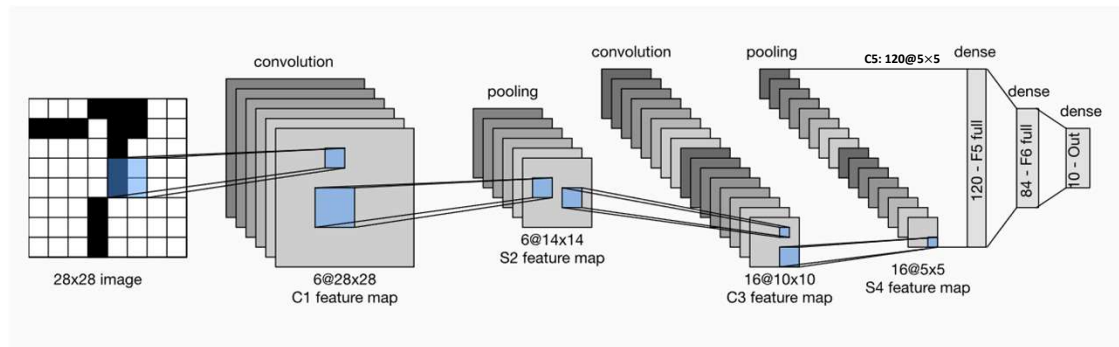
ConvNets: Popular Architectures

LeNet-5



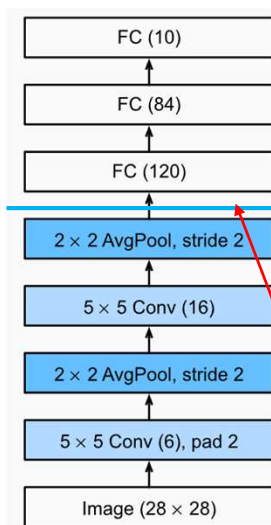
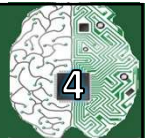
LeCun et al. in 1998

- At a high level, LeNet (LeNet-5) consists of two parts:
 - (i) a convolutional encoder consisting of two convolutional layers
 - (ii) a dense block consisting of three fully connected layers.



Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

LeNet-5



As we go through the network:

Spatial size **decreases**
(using pooling or strided conv)

Number of channels **increases**
(total "volume" is preserved!)

Layer	Output Size	Weight Size
Input	1 x 28 x 28	
Conv ($C_{out} = 6, K=5, P=2, S=1$)	6 x 28 x 28	6 x 1 x 5 x 5
Sigmoid	6 x 28 x 28	
AvgPool ($K=2, S=2$)	6 x 14 x 14	
Conv ($C_{out}=16, K=5, P=0, S=1$)	16 x 10 x 10	16 x 6 x 5 x 5
Sigmoid	16 x 10 x 10	
AvgPool ($K=2, S=2$)	16 x 5 x 5	
Flatten	120	
Linear (120->84)	84	120 x 84
Sigmoid	84	
Linear (84->10)	10	84 x 10

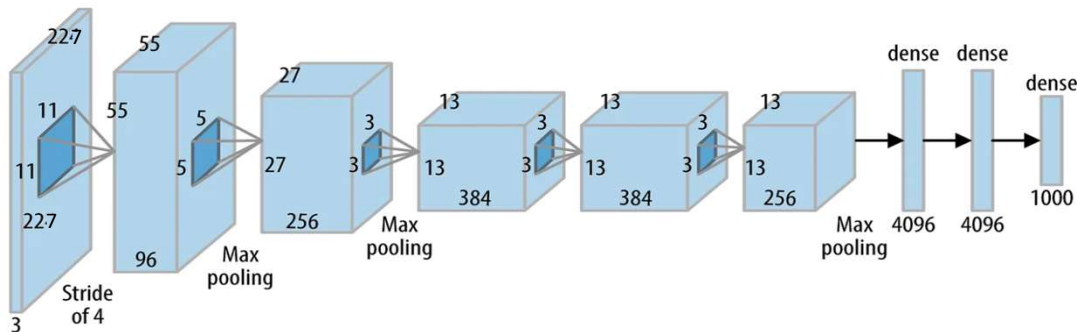
Layer	Output Size	Weight Size
Conv ($C_{out} = 120, K = 5$)	120 x 1 x 1	120 x 16 x 5 x 5

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

AlexNet

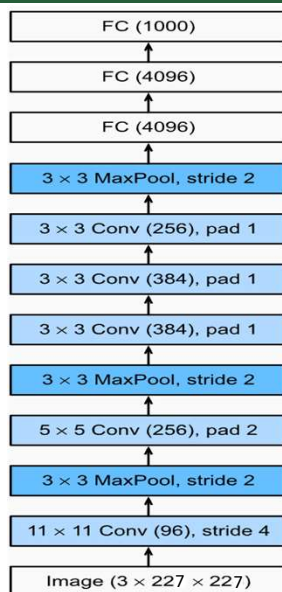
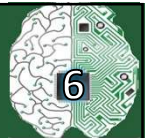


- [ImageNet LSVRC-2012 \[Krizhevsky et al.\]](#)
- The heart of AlexNet is its architecture, comprising of five Convolutional (CONV) layers and three Fully Connected (FC) layers.
- AlexNet opts for the Rectified Linear Unit (ReLU) as its activation function



Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

AlexNet



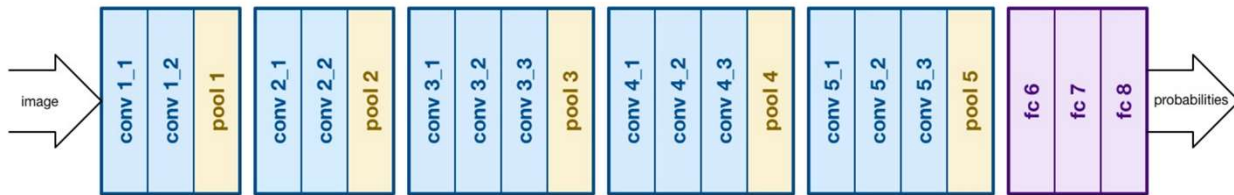
AlexNet Network - Structural Details													
Input			Output			Layer	Stride	Pad	Kernel size		in	out	# of Param
227	227	3	55	55	96	conv1	4	0	11	11	3	96	34944
55	55	96	27	27	96	maxpool1	2	0	3	3	96	96	0
27	27	96	27	27	256	conv2	1	2	5	5	96	256	614656
27	27	256	13	13	256	maxpool2	2	0	3	3	256	256	0
13	13	256	13	13	384	conv3	1	1	3	3	256	384	885120
13	13	384	13	13	384	conv4	1	1	3	3	384	384	1327488
13	13	384	13	13	256	conv5	1	1	3	3	384	256	884992
13	13	256	6	6	256	maxpool5	2	0	3	3	256	256	0
						fc6			1	1	9216	4096	37752832
						fc7			1	1	4096	4096	16781312
						fc8			1	1	4096	1000	4097000
Total						62.378.344							

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

VGGNet



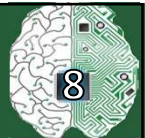
- 16 layers (not including pooling)
- 3x3 convolutions only
- 138 million parameters
- Pre-trained model is available for download and is a standard in the field



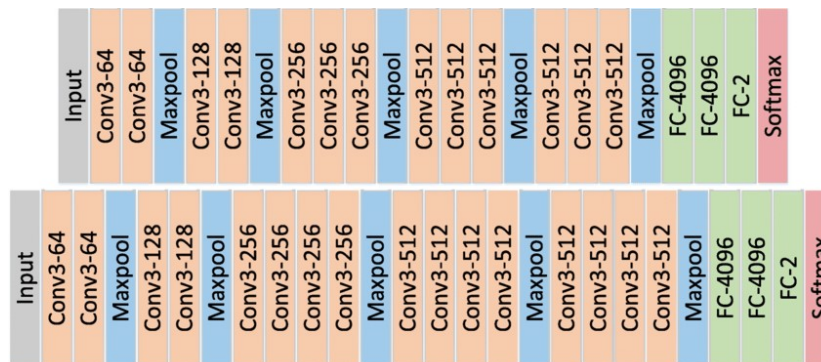
Simonyan & Zisserman (2014)

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

VGGNet

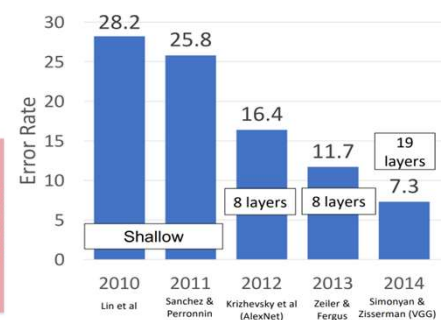


- There exist multiple versions of VGGNet such as VGG16, VGG19, etc., differentiated by the total number of layers in the network.



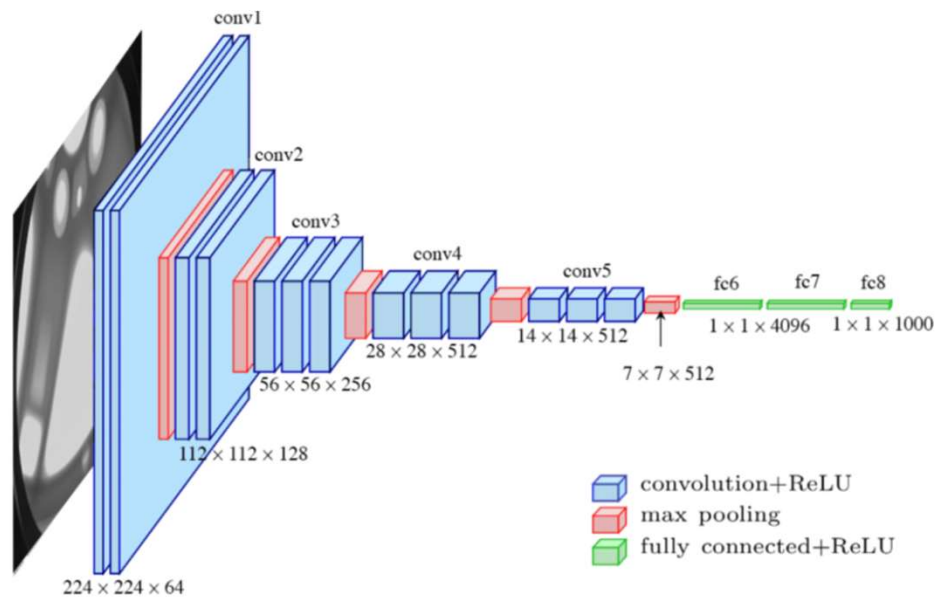
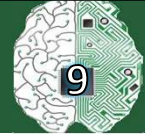
Network structures of VGG16 (top) and VGG19 (bottom)

Simonyan and Zisserman, 2014



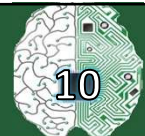
Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

VGG16

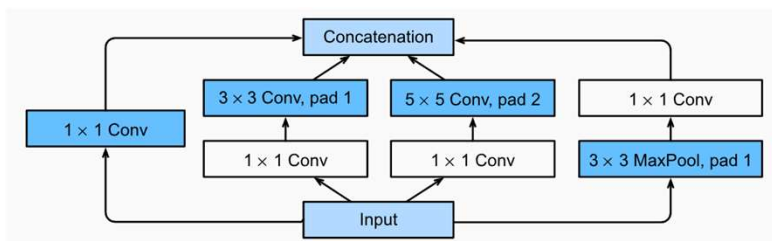


Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

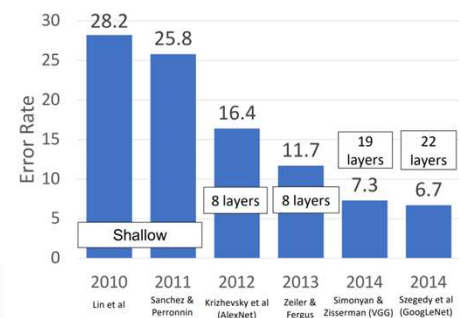
GoogLeNet



- **Focus on Efficiency**
- Uses a structure that combined the strengths of Network in Network (NiN) (Lin et al., 2013), repeated blocks (Simonyan and Zisserman, 2014), and a number of convolution kernels applied parallelly.
- The basic convolutional block in GoogLeNet is called an **Inception block**

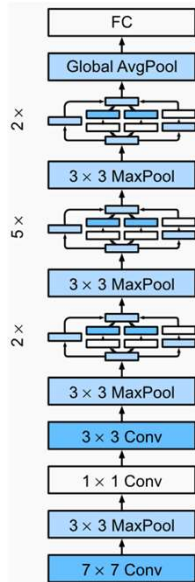
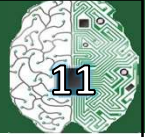


ImageNet Classification Challenge



Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

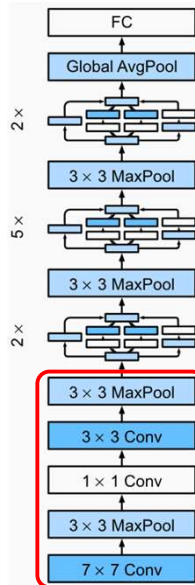
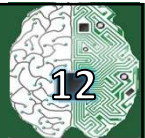
GoogLeNet



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								

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

GoogLeNet



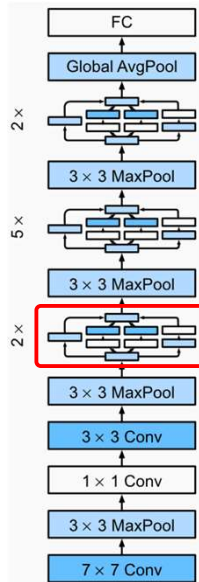
Layer	Input size		Layer			Output size				
	C	H / W	filters	kernel	stride	pad	C	H/W	memory (KB)	params (K)
conv	3	224	64	7	2	3	64	112	3136	9
max-pool	64	112		3	2	1	64	56	784	0
conv	64	56	64	1	1	0	64	56	784	4
conv	64	56	192	3	1	1	192	56	2352	111
max-pool	192	56		3	2	1	192	28	588	0

Total from 224 to 28 spatial resolution:
Memory: 7.5 MB
Params: 124K

Compare VGG-16:
Memory: 42.9 MB (5.7x)
Params: 1.1M (8.9x)

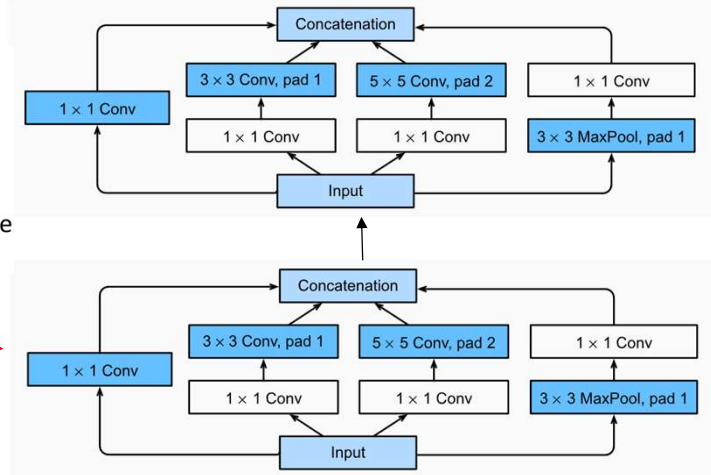
Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

GoogLeNet



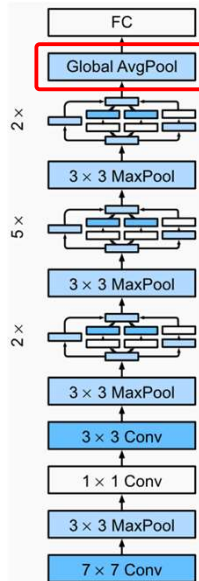
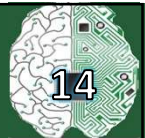
Inception module
Local unit with parallel branches

Local structure repeated many times throughout the network



Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

GoogLeNet



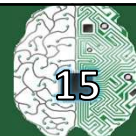
	Input size		Layer				Output size				
Layer	C	H/W	filters	kernel	stride	pad	C	H/W	memory (KB)	params (k)	flop (M)
avg-pool	1024	7		7	1	0	1024	1	4	0	0
fc	1024		1000				1000		0	1025	1

Compare with VGG-16:

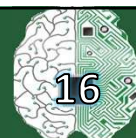
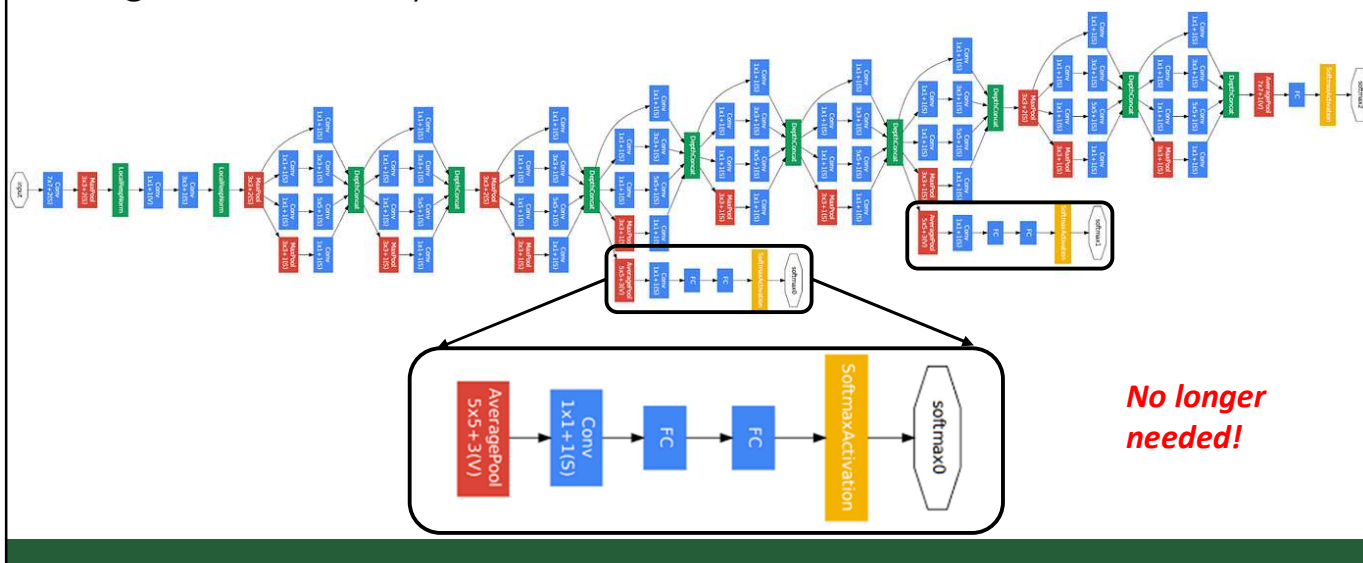
Layer	C	H/W	filters	kernel	stride	pad	C	H/W	memory (KB)	params (K)	flop (M)
flatten	512	7					25088		98		
fc6	25088			4096			4096		16	102760	103
fc7	4096			4096			4096		16	16777	17
fc8	4096			1000			1000		4	4096	4

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad

GoogLeNet



GoogLeNet: Auxiliary Classifiers



Questions?

Acknowledgement: Prof. Dr. A. Krizhevsky, Alex Krizhevsky, Ilya Sutskever, and Jeff Dean

Prof. Monidipa Das, Department of CSE, IIT (ISM) Dhanbad