

NPTEL Week-12 Live Session

on Machine Learning and Deep Learning - Fundamentals and Applications (noc24_ee146)

A course offered by: Prof. Manas Kamal Bhuyan, IIT Guwahati

Convolutional neural networks quiz,
Colab implementation



By

Arka Roy

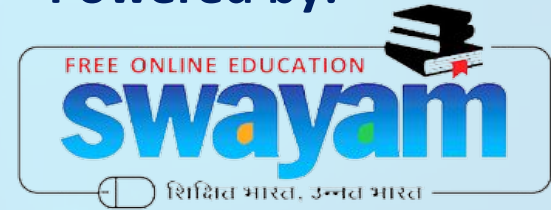
NPTEL PMRF TA

Prime Minister's Research Fellow

Department of Electrical Engineering, IIT Patna

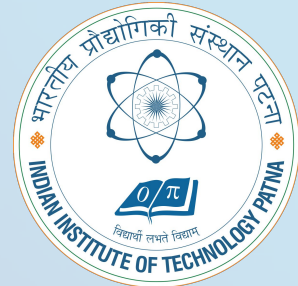
Web: <https://sites.google.com/view/arka-roy/home>

Powered by:



PMRF

Prime Minister's Research Fellows
Ministry of Education
Government of India



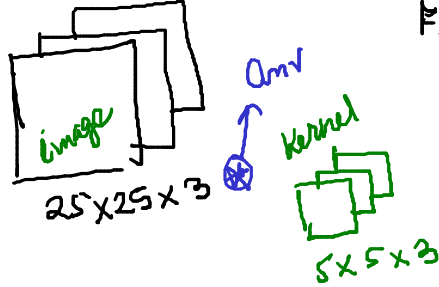
1) What should be the response dimension if an image of dimension $3 \times 25 \times 25$ is convolved with a filter of dimension $3 \times 5 \times 5$?

Channel ~~Height~~ ~~Width~~.

padding = 0; stride = 1

image dim = $25 \times 25 \times 3$

Filter dimension = $5 \times 5 \times 3$.



$$\begin{aligned} \text{output's shape} &= \left(\frac{\text{Input shape} - \text{Kernel size} + (2 \times \text{Pd})}{\text{stride}} \right) + 1 \\ &= \frac{25 - 5 + 0}{1} + 1 \\ &= 21 \end{aligned}$$

output size
= $21 \times 21 \times 1$

2) What should be the response dimension if an image of dimension $3 \times 25 \times 25$ is padded with width of 2 and convolved with a filter of dimension $3 \times 5 \times 5$?

Padding = 2; Stride = 1.

$$\begin{aligned} \text{output shape} &= \frac{25 - 5 + (2 \times 2)}{1} + 1 \\ &= 25 - 5 + 4 + 1 \\ &= 25. \end{aligned}$$

overall convolution o/p dimension
= $25 \times 25 \times 1$
Padding = "Same"

☒ $1 \times 21 \times 21$

☐ $3 \times 25 \times 25$

☐ $1 \times 25 \times 25$

☐ $25 \times 25 \times 1$

☐ $1 \times 21 \times 21$

☐ $3 \times 25 \times 25$

☒ $1 \times 25 \times 25$

☐ $25 \times 25 \times 1$

An RGB input image has been converted into a matrix of size $257 \times 257 \times 3$ and a kernel/filter of size $7 \times 7 \times 3$ with a stride of 2 and padding = 3 is used for 2D convolution. What will be the size of the output of convolution?

- ☒ a. 129x129x1
- b. 128x128x1
- c. 254x254x3
- d. 256x256x1

$$\begin{aligned} \text{o/p dimension} &= \frac{257 - 7 + (2 \times 3)}{2} + 1 \\ &= 128 + 1 \\ &= 129 \end{aligned}$$

overall output shape after convolution

$$= \underbrace{129 \times 129 \times 1}$$

Primary reason for adding pooling layers is?

a. Increase feature map size ~~X~~

~~b.~~ Reduce computations for subsequent layers

~~c.~~ To produce activations that summarize filter response in local windows.

~~d.~~ Both b and c

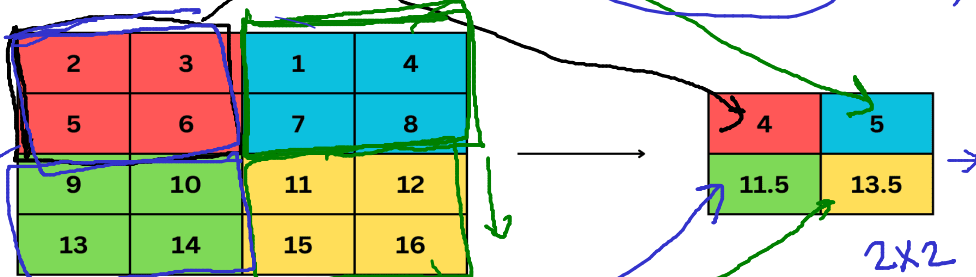
Kernel of Pooling

Computation in coming layer is also reduced

i) Reduces the size of the activation map.

ii) Within the kernel of the pooling block we summarize the information by taking either average or maximum value. (MP)

Average Pooling

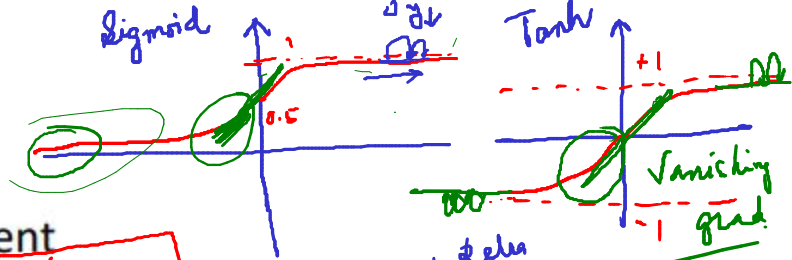


$$\frac{1}{4} (2+3+5+6) = 4$$

Take average of all values in the window

Advantage of ReLU over Sigmoid and TanH is

- a. Low computational requirements
- b. Alleviates vanishing gradient to some extent
- c. Backpropagation is simpler
- d. All of above



$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

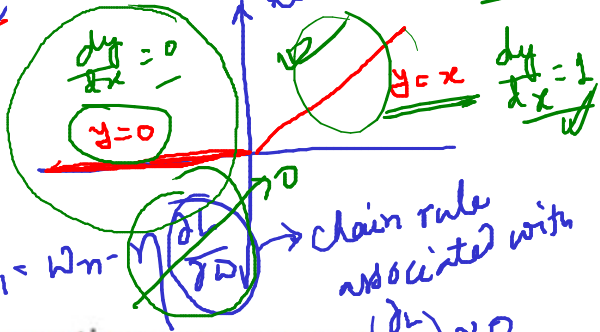
lets x is very high.

$$\sigma(x) \approx 1$$

$$\sigma'(x) = 1 \times (1 - 1) = 0$$

$$\sigma(x) \approx 0$$

$$\sigma'(x) = 0 \times (1 - 0) = 0$$



chain rule associated with

$$W_{n+1} = W_n - \eta \frac{\partial L}{\partial W}$$

$$\left(\frac{\partial L}{\partial W}\right) \approx 0$$

Learning will stop.

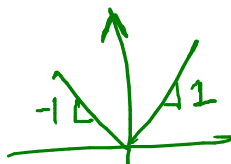
(Vanishing gradient)

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

ReLU: - $y = 0 : x < 0$

$$\frac{dy}{dx} = 0$$

$$y = |x|$$

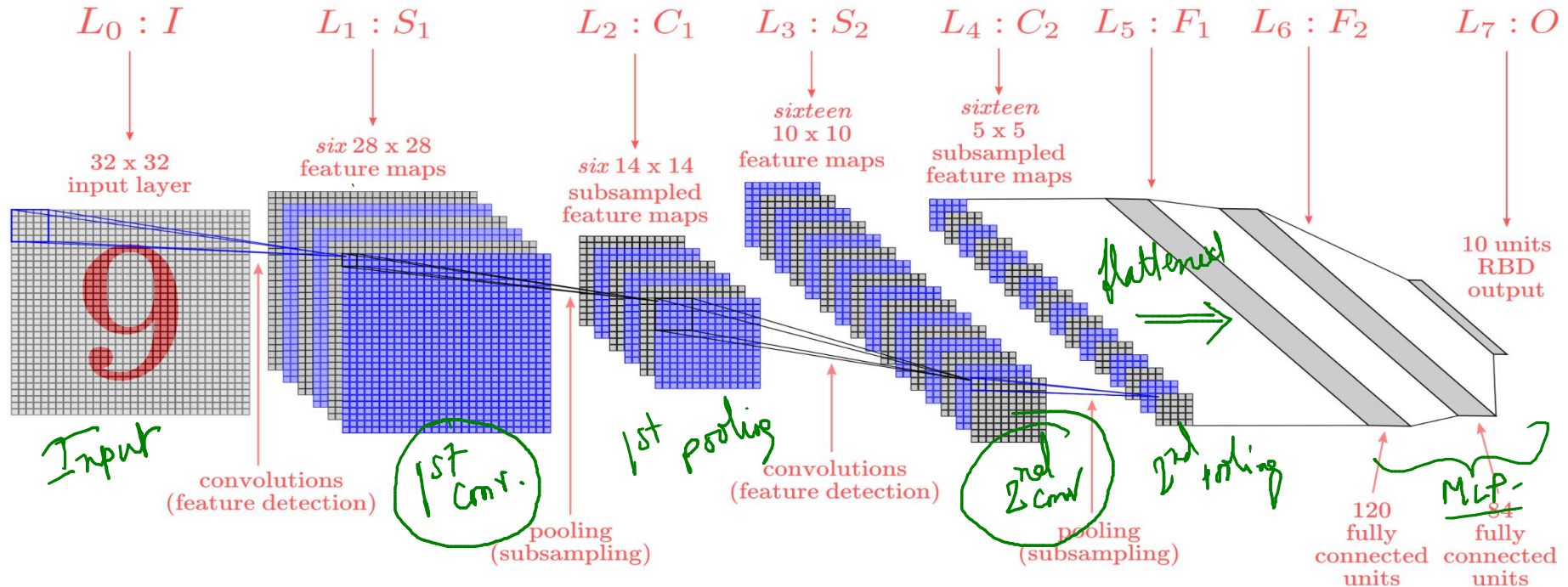


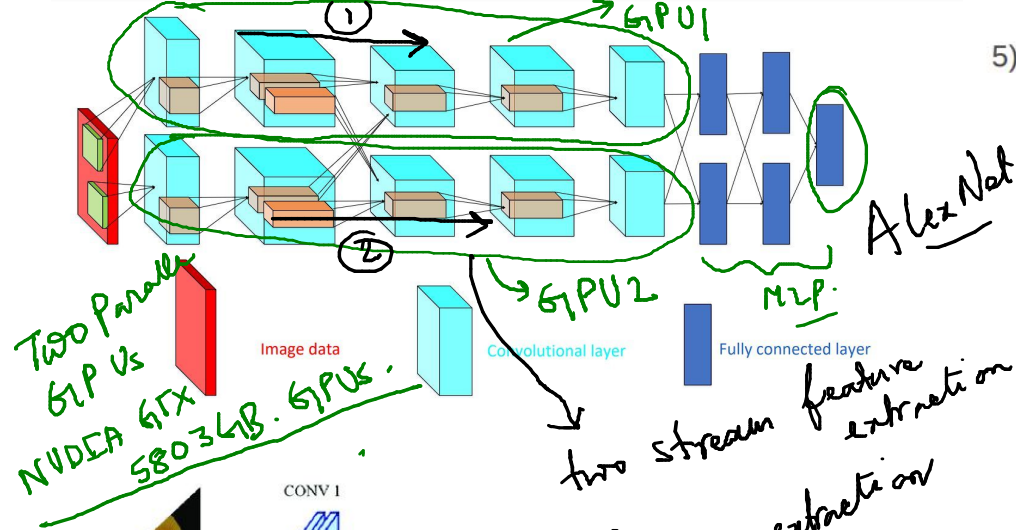
- ☒ Sigmoid
- ☒ Tanh
- ☒ ReLU
- ☒ Absolute

9) What of the following activation function has no gradient for negative responses

How many convolution layers are there in a LeNet-5 architecture?

- ☒ a. 2
- b. 3
- c. 4





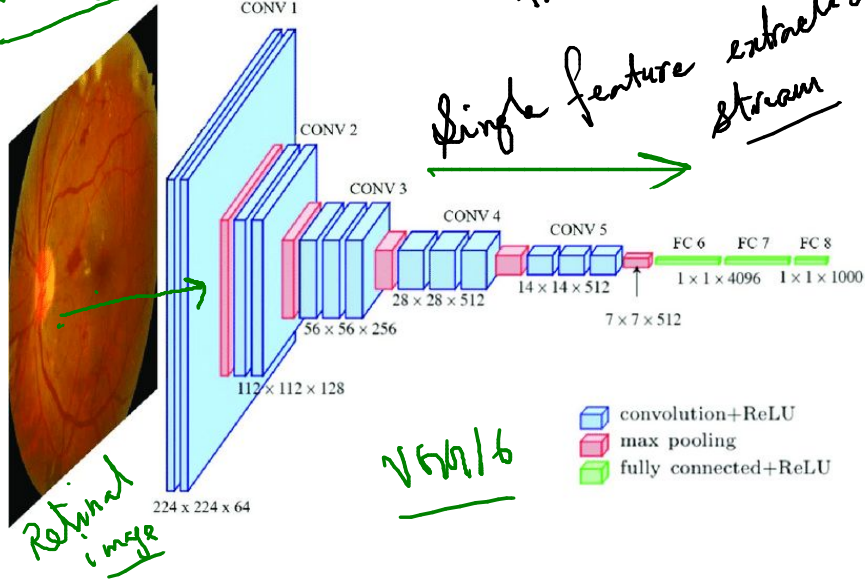
5) What is the architectural difference between VGG and Alexnet

VGG has single feature extraction stream and Alexnet has two

Deconvolution

Bilinear interpolation

None of the above



8) If CNN filters (no bias) of size 3×3 are employed for transforming an image with 3 channels into 64 responses. How many weights need to be learned

☐ 1378

☒ 1728

☐ 1738

☐ 2000

Kernel size = (3×3) [No bias] \rightarrow applied on 3 channel image.

Kernel number = 64.

Effectively Kernel dimension = $(3 \times 3) \times 3$

Total parameter = $(3 \times 3 \times 3) \times 64 = 1728$

single for all 64 case

learnable weights
64 such filters
Kernel = $3 \times 3 \times 3$
= 27.

CNN \rightarrow learnable layer
learnable weights

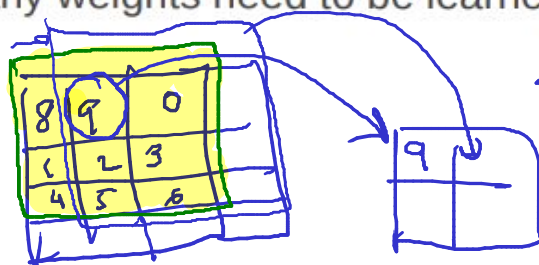
9) If max pooling kernels of size 3×3 are employed for transforming an image with 3 channels into 3 responses. How many weights need to be learned

☐ 27

☐ 18

☐ 9

☒ 0



pooling structure
does not need
any learning
mechanism

1) In inception block the input information is processed through how many parallel channels?

☐ 2

☐ 3

☒ 4

☐ None of the above

2) In inception block the processed input information is combined using which operation?

☒ Sum

☒ Multiplication

☒ Concatenation

☒ None of the above

