

# Today's Agenda

## 1) VGG Network

CNN

Stats }  
ML Revision }

1) Why CNN and not ANN?

{ Image, Audio, Video }

ANN → { hidden layer }  
↓  
Dense, Linear

Mel spectrograms

Why CNN?

1) Spatial Information

1	3	5
7	8	0
2	1	3

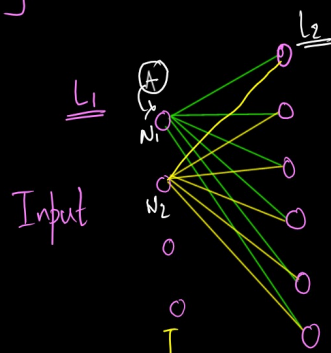
\* { Autofocus }

Image → Object

↳ Person  
↳ Hill  
↳ Bird

$$y = \underline{Wx} + \text{constant}$$

Structure / Integrity



Increase in Complexity

Hidden Dense Linear  
128 → No of Neurons L1  
512 → No of Neurons L2

= \*65536 \*

learnable Parameters  
Training Time

Car

Dense Layer

DNN

GPU

① ANN

Graphics Card

Nvidia

AMD

Dense ↑  
CNN ↓

↳ Complex  
↳ Lose Structure Integrity  
↳ Increase the Parameters

Normal English  
Academic English

Hidden, Linear, Language  
↓  
Framework

{ GRE/SAT }

↳ 2500/3000

Dense  $\rightarrow$  TF, PyT  $\rightarrow$  Linear

CNN  $\rightarrow$  Image, video, Audio  $\rightarrow$  unstructured data

Input Image  
64x64

Kernel (Filter, Feature Extractor, Edge Detector)  
3x3

Horizontal  
Vertical Feature  
Diagonal Feature

Grabs Pattern  
Type Conv

How many features are there in any image?

$3 \times 3 = 9$

unique

Many

16 kernels = 16 features  
32 kernels = 32 features

Raw Image

32 ① Low level Features  
64 ② Mid level Features  
128 ③ High level Features

Image  $\times$  Kernel = Channel / Feature Map

Image  $\times$  32 (Kernel) = Feature Maps (32)

\* Total no of kernels = Total no of FMs

98  
2012  
{ LeNet, Alexnet }

US Postal Services  $\rightarrow$  Pin Code Index  
Zipcode Identification

Deep learning - 2012  
UGuGu

No of layers  $\uparrow$  Accuracy  $\downarrow$

## Selection of Kernel

$$\sim 7 \times 7, \sim 5 \times 5$$

$$\underline{3 \times 3}$$

$$64, 64, 3$$

$$(3 \times 3)^{32} = 62 \times 62 \times 32$$

$$64 \times 64 \times 3$$

$$\sim \sim$$

$$(7 \times 7) 64 = \underline{58 \times 58 \times 64}$$

$$3 \times 3 = 9$$

$$5 \times 5 = 25$$

$$3 \times 3 = \begin{matrix} 2 \text{ pixels } \uparrow \\ 2 \text{ pixels } \downarrow \end{matrix}$$

$$\underline{5 \times 5}$$

$$= \underline{60 \times 60 \times 128}$$

25 Parameters

$$I: I = 64 \times 64$$

$$N: K = 128$$

$$\underline{3 \times 3}$$

$$= 62 \times 62 \times 128$$

$$\rightarrow 3 \times 3$$

$$9 + 9 = 18$$

$$= \underline{60 \times 60 \times 128}$$

$$3 \times 3 = 2 \text{ pixels}$$

$$5 \times 5 = 4$$

$$7 \times 7 = 6$$

$$\underline{7 \times 7}$$

$$\underline{3 \times 3}$$

$$7 \times 7 = (3 \times 3)^3$$

$$\underline{49} = 27$$

$$\underline{\underline{L.P}}$$

## Max Pooling

1) Purpose  $\rightarrow$  To Decrease the no of layers in CNN

