Introduction to Computer Vision with Neural Networks

Presented by - Astha Garg

01 Computer Vision

02 Convolution Neural Networks

03 Introduction to Neural Networks

04 Transfer Learning

COMPUTER VISION

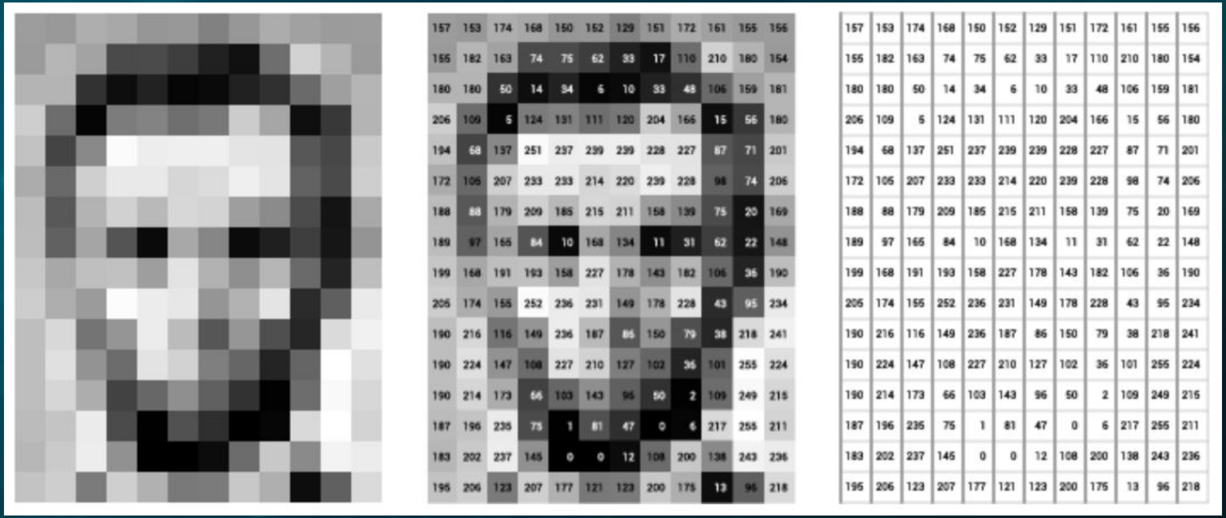
01

WHAT IS COMPUTER VISION?

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs — and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand.

HOW DOES COMPUTER VISION WORK?

Machines interpret images very simply: as a series of pixels, each with their own set of color values.



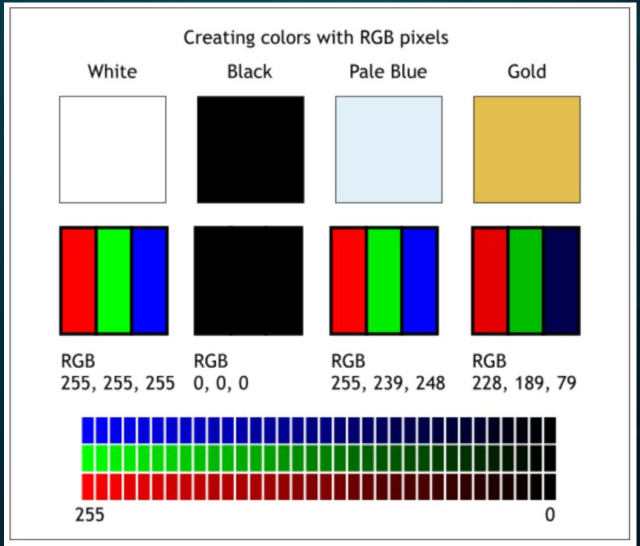
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

HOW DOES COMPUTER VISION WORK?

Computers usually read color as a series of 3 values – red, green, and blue (RGB) – on that same 0 – 255 scale.

For some perspective on how computationally expensive this is, consider this tree:

- Each color value is stored in 8 bits.
- 8 bits x 3 colors per pixel = 24 bits per pixel.
- A normal sized 1024 x 768 image x 24 bits per pixel = almost 19M bits, or about 2.36 megabytes.



SOME ESTABLISHED COMPUTER VISION TASKS

- **Image classification** - It is able to accurately predict that a given image belongs to a certain class.
- **Object detection** - Examples include detecting damages on an assembly line or identifying machinery that requires maintenance.
- **Object tracking** - Follows or tracks an object once it is detected. This task is often executed with images captured in sequence or real-time video feeds.
- **Content-based image retrieval** - Uses computer vision to browse, search and retrieve images from large data stores, based on the content of the images rather than metadata tags associated with them.

SOME USE CASES

Healthcare
Analysis of ultrasound images, MRI, and CT scans

Automotive Industry
Real-time monitoring by an image recognition algorithm for autonomous vehicles.

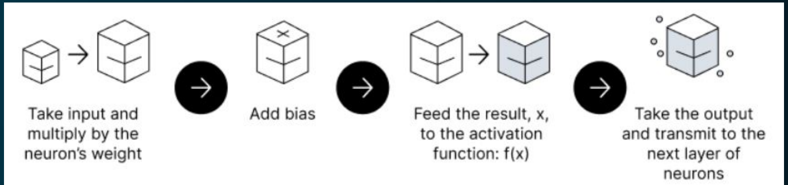
Retail - Customer Behavior Tracking
Evaluate video material and study consumer behaviour

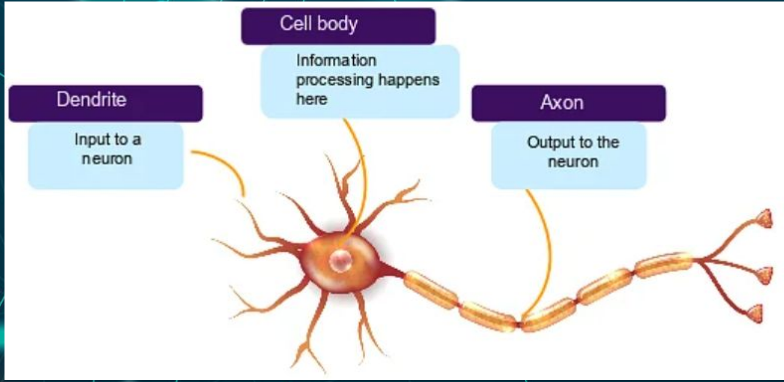
Agriculture
Analysis of data generated using drones, satellite images, and remote sensors

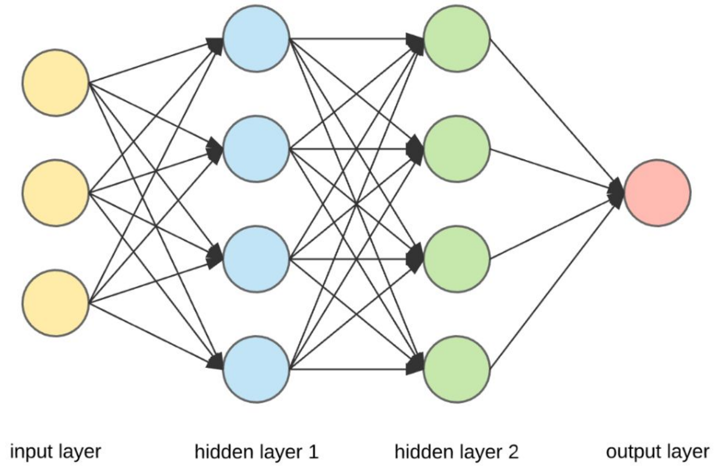
NEURAL NETWORKS

02

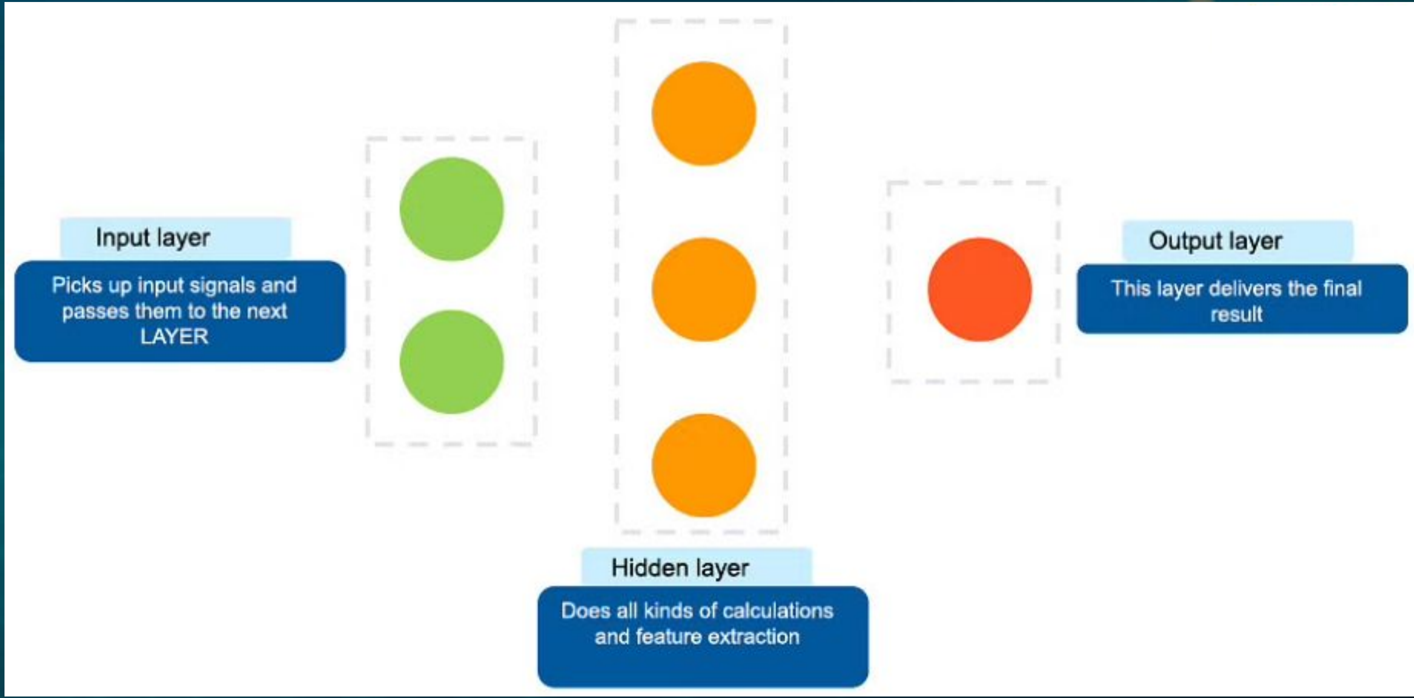
WHAT IS A NEURAL NETWORK?



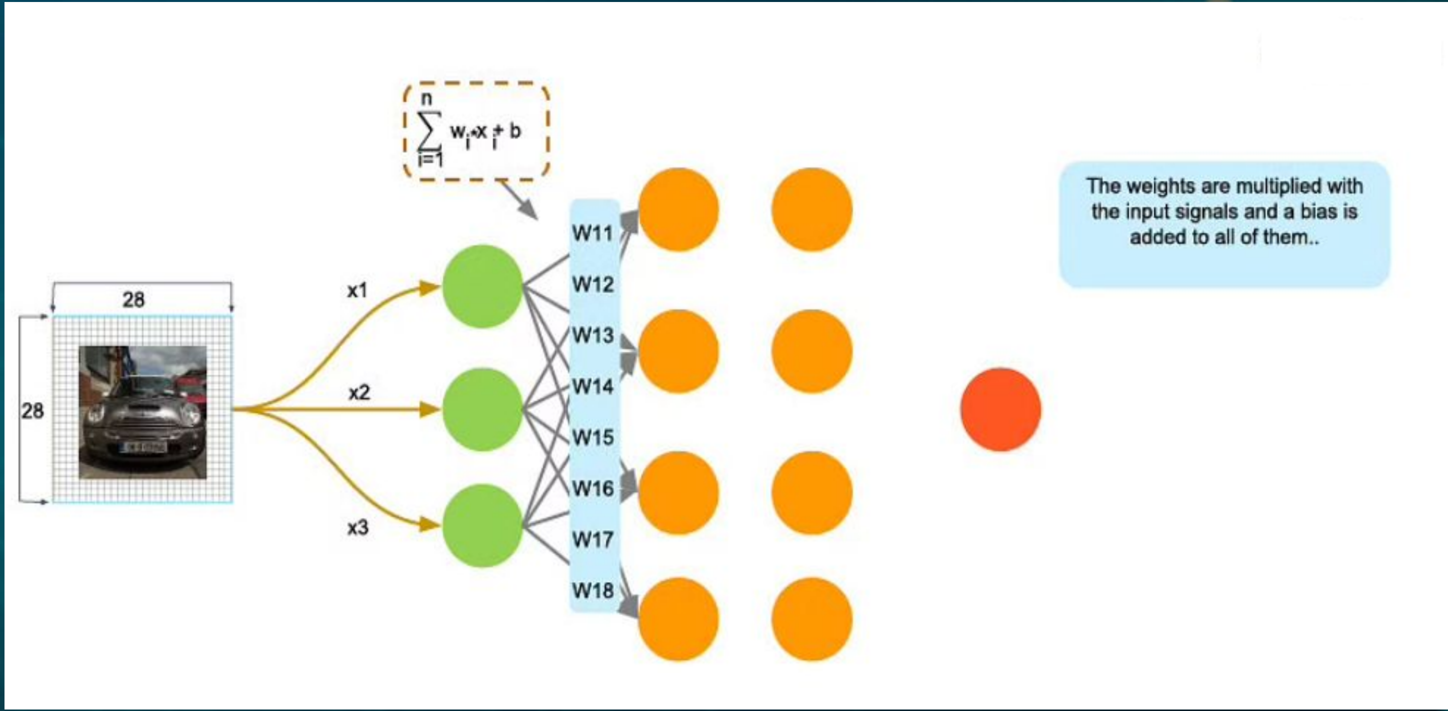




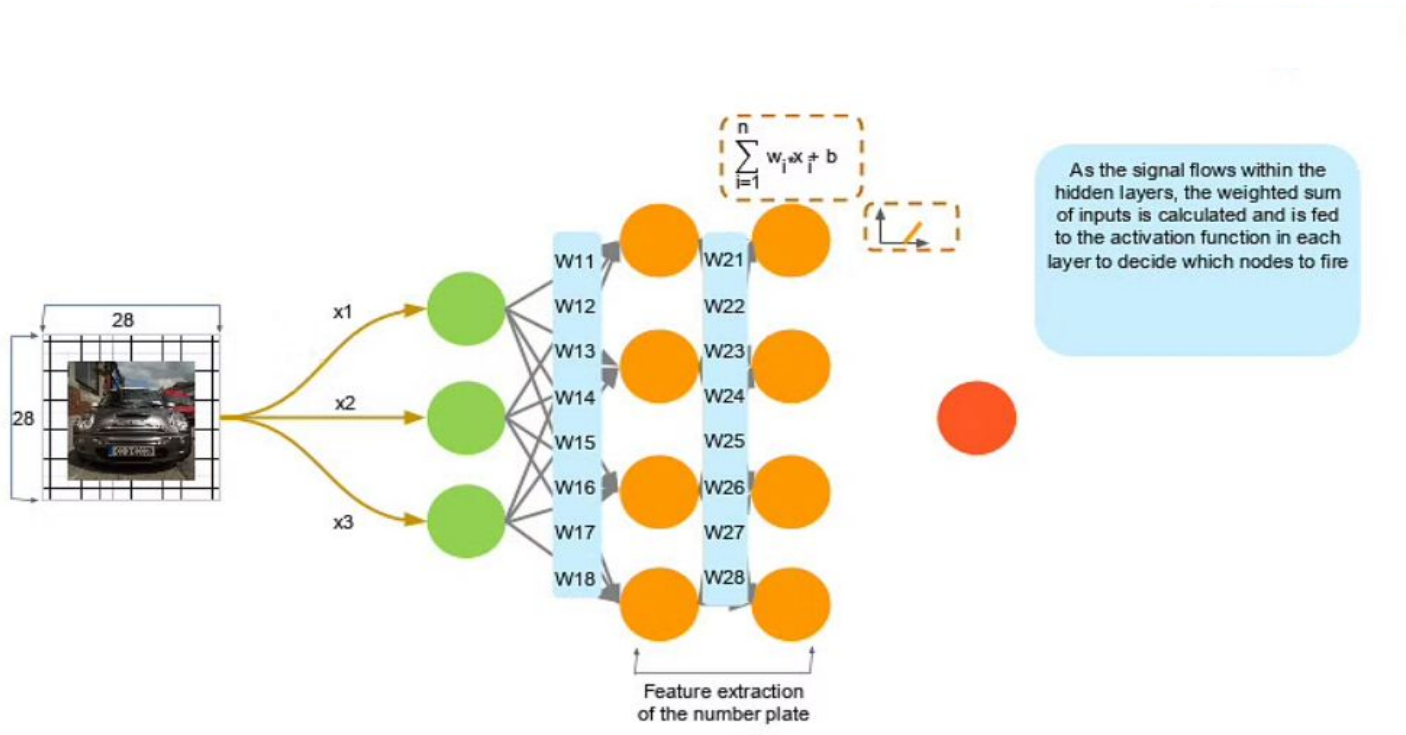
BASICS OF NEURAL NETWORKS



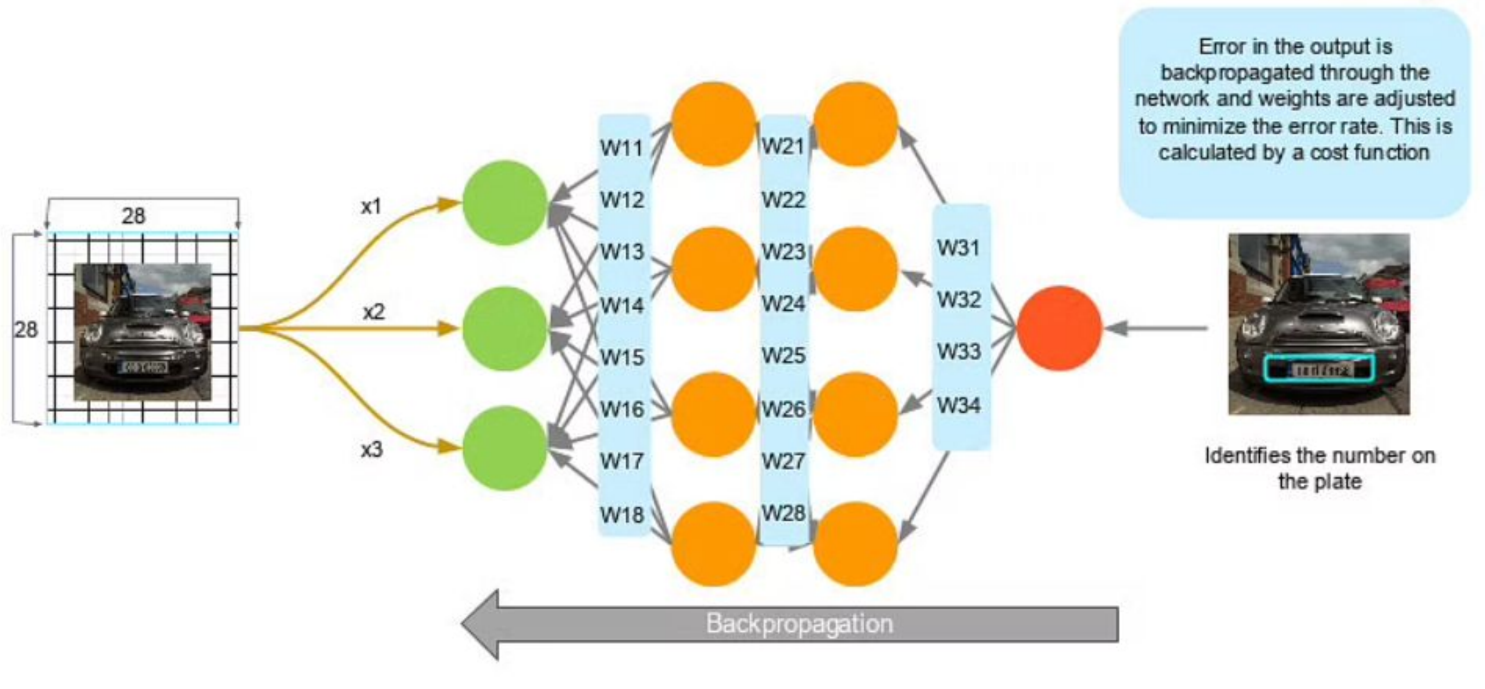
BASICS OF NEURAL NETWORKS



BASICS OF NEURAL NETWORKS



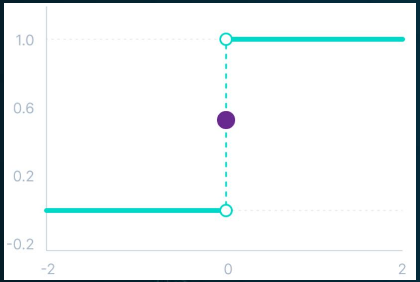
BASICS OF NEURAL NETWORKS



TYPES OF ACTIVATION FUNCTIONS

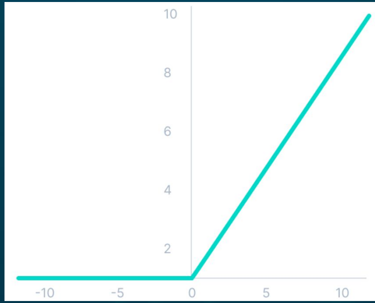
Binary - Step Function

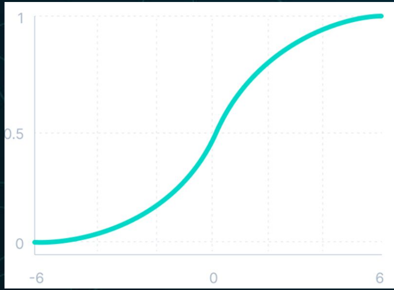
$$f(x) = \begin{cases} 0 & for\ x < 0 \\ 1 & for\ x \geqslant 0 \end{cases}$$



ReLU Function

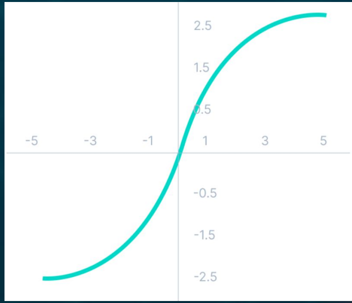
$$f(x) = max\ (0, x)$$





Sigmoid Function

$$f(x) = \frac{1}{1 + e^{-x}}$$



$$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$$

Hyperbolic Tangent Function

TYPES OF NEURAL NETWORKS

- Feed Forward Neural Network
- Multi - layered Perceptron Neural Network
- Convolution Neural Network
- Radial Basis Neural Network
- Recurrent Neural Network
- Sequence to Sequence Model
- Modular Neural Network

[1] https://www.mygreatlearning.com/blog/types-of-neural-networks/

BASIC IMAGE CLASSIFICATION *HANDS-ON*

CONVOLUTION NEURAL NETWORKS

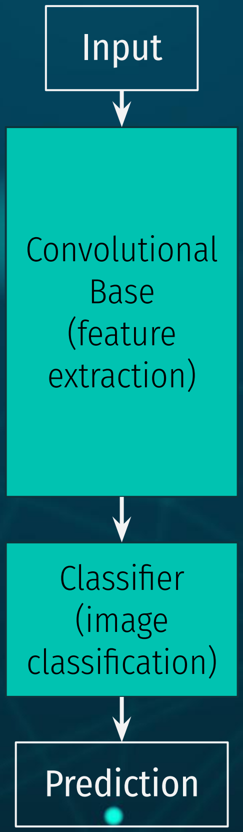
03

BASICS OF CNN

A convolution neural network has multiple hidden layers that help in extracting information from an image.

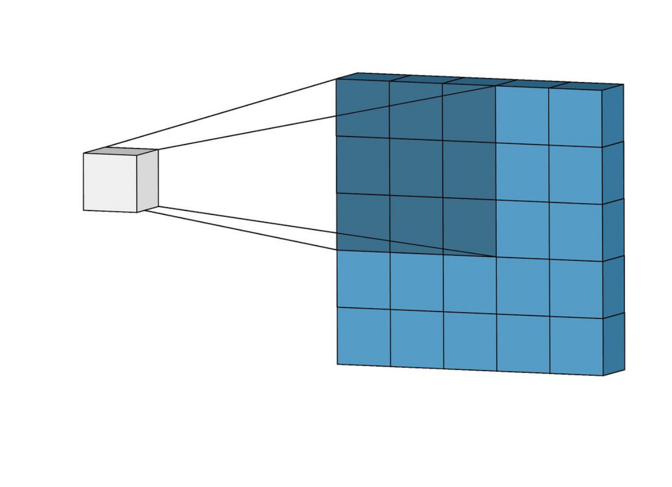
The important parts of CNN architecture are:

1. **Convolutional base**
2. **Classifier**

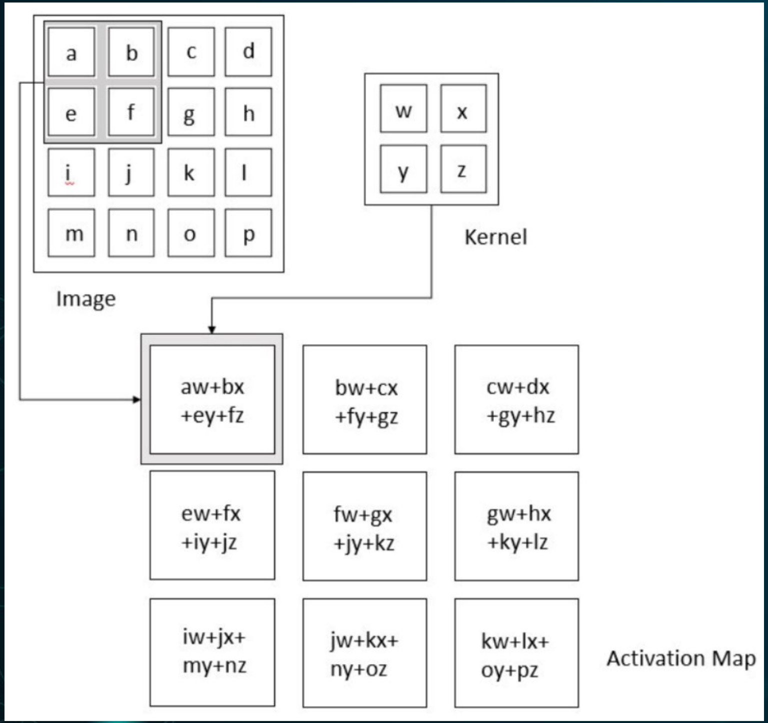


CONVOLUTION LAYER

This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field. The kernel is spatially smaller than an image but is more in-depth. This means that, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels.



CONVOLUTION LAYER



During the forward pass, the kernel slides across the height and width of the image-producing the image representation of that receptive region. This produces a two-dimensional representation of the image known as an activation map that gives the response of the kernel at each spatial position of the image. The sliding size of the kernel is called a stride.

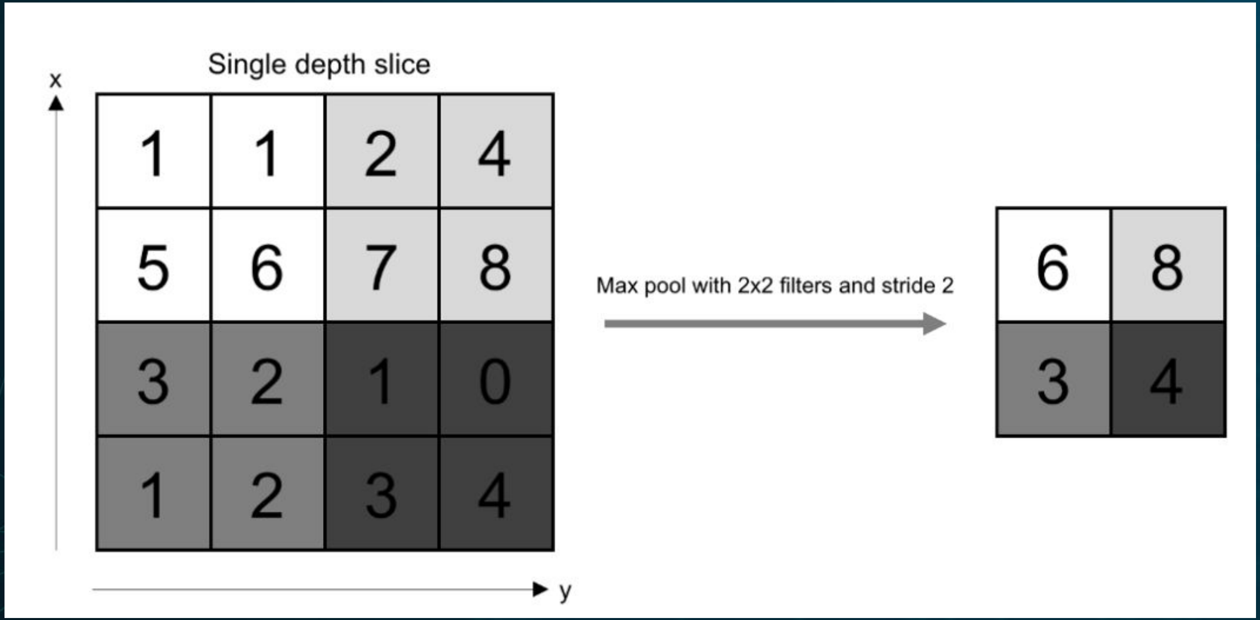
POOLING LAYER

The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs.

This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights.

The pooling operation is processed on every slice of the representation individually.

POOLING LAYER

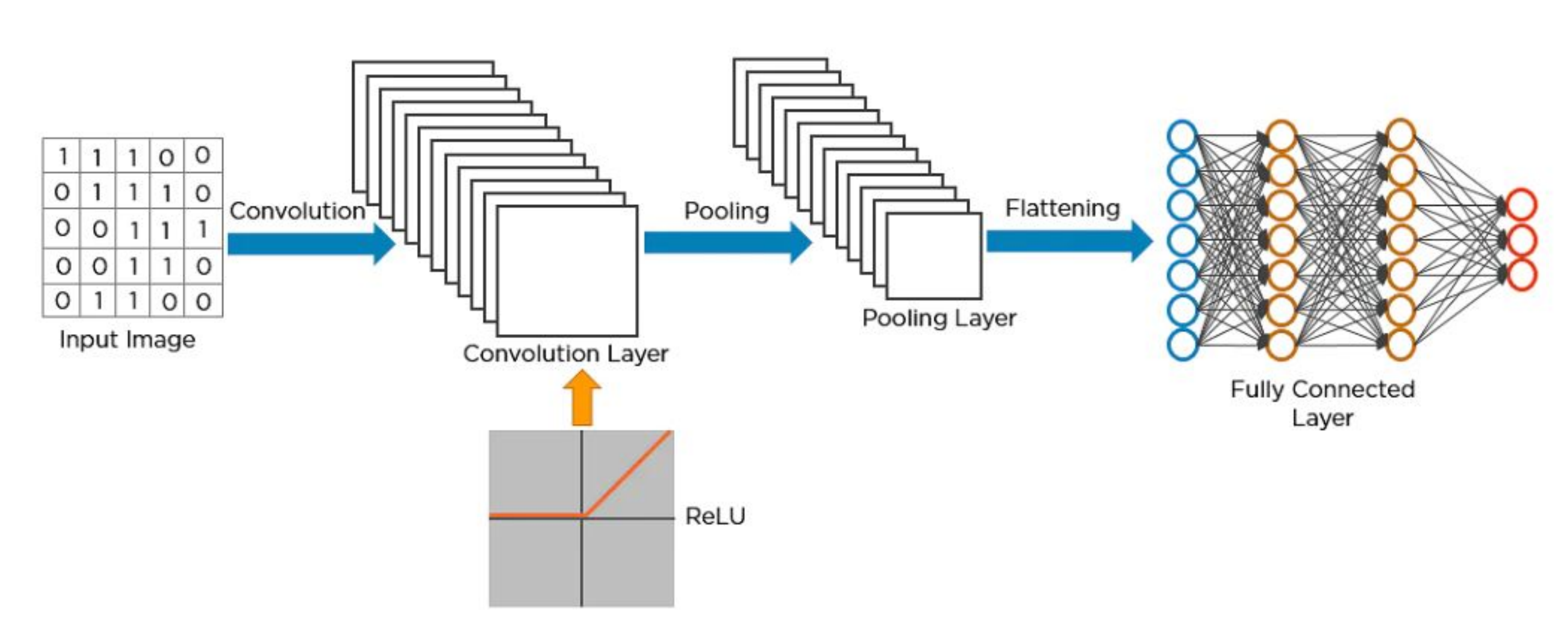


FULLY CONNECTED LAYER

Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.

The FC layer helps to map the representation between the input and the output.

PUTTING IT ALL TOGETHER

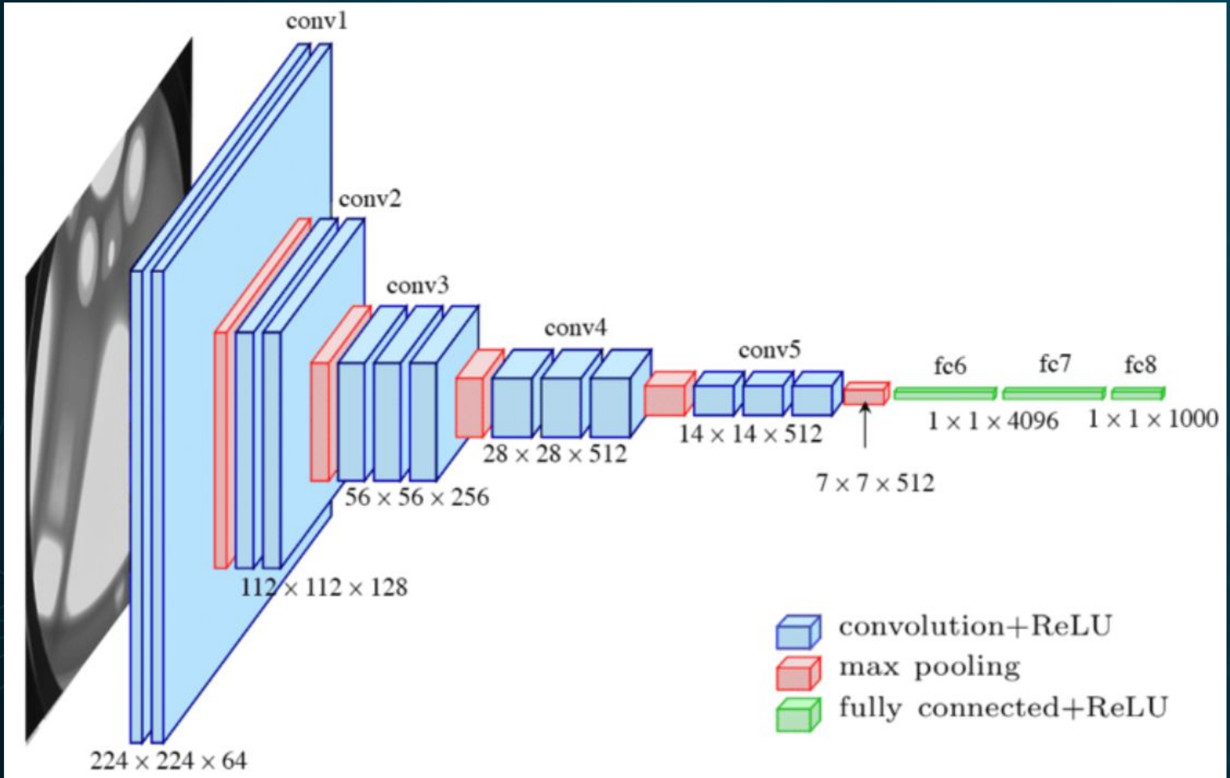


CNN
HANDS-ON

TRANSFER LEARNING

04

PRE-TRAINED MODELS



REPURPOSING A PRE-TRAINED MODEL

Train the entire model

In this case, you use the architecture of the pre-trained model and train it according to your dataset. You're learning the model from scratch, so you'll need a large dataset (and a lot of computational power).

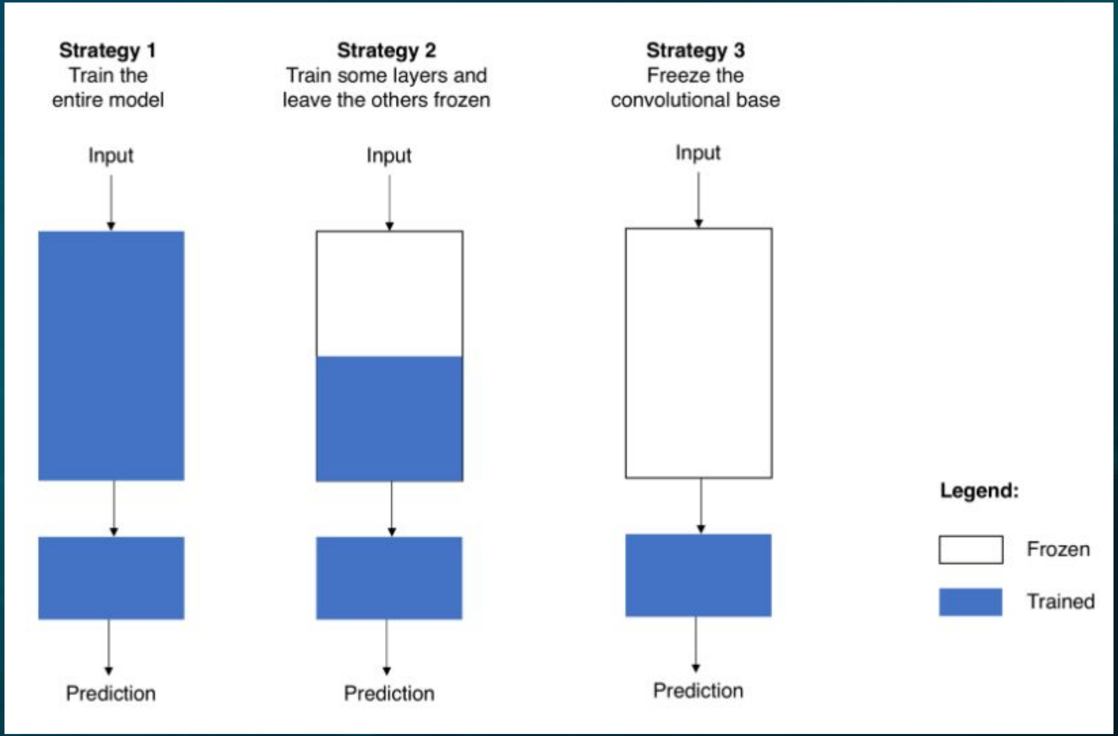
Train some layers

Here, we play with that dichotomy by choosing how much we want to adjust the weights of the network. (a frozen layer does not change during training.

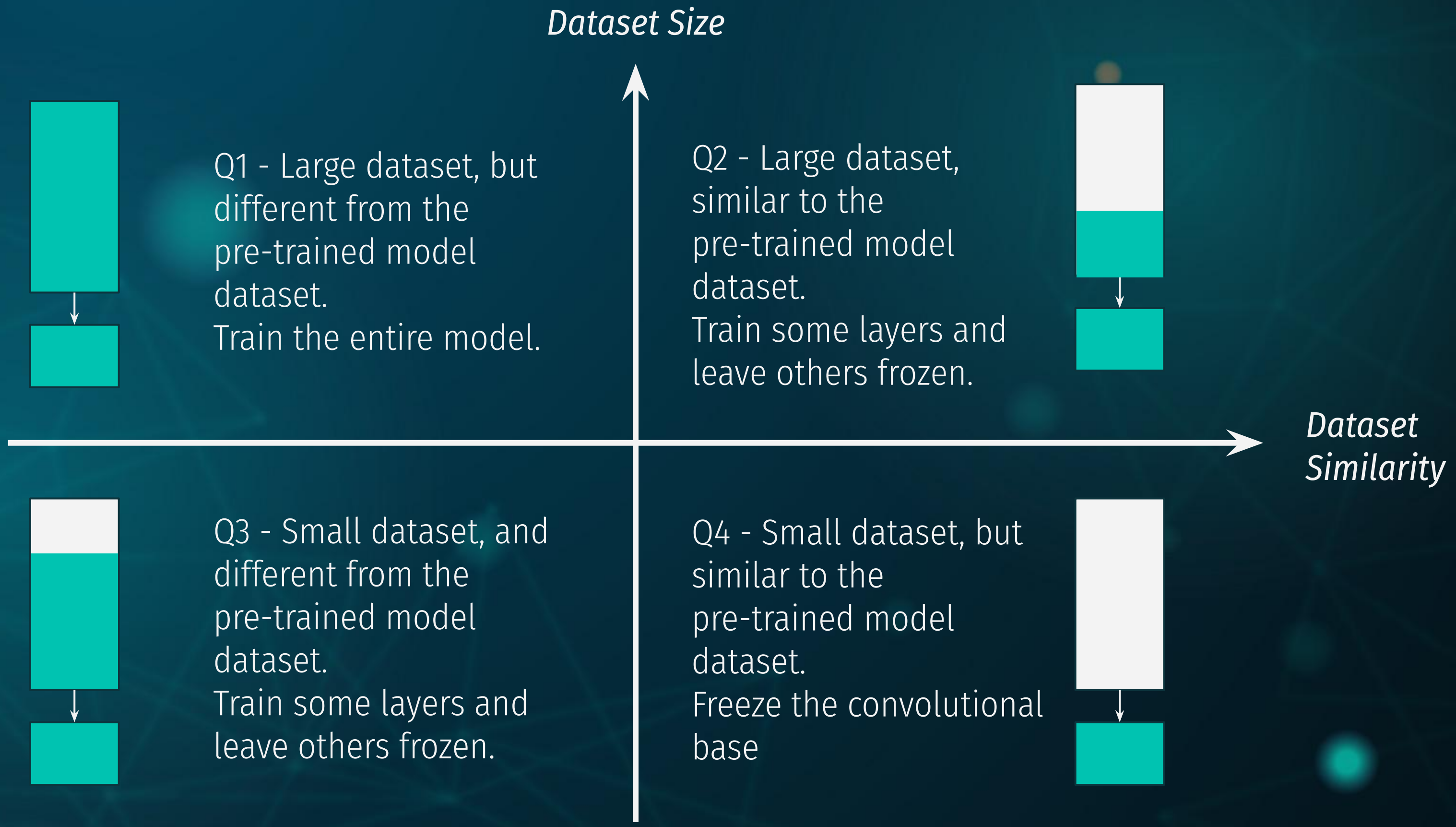
Freeze the convolutional base

The main idea is to keep the convolutional base in its original form and then use its outputs to feed the classifier. You're using the pre-trained model as a fixed feature extraction mechanism

REPURPOSING A PRE-TRAINED MODEL



REPURPOSING A PRE-TRAINED MODEL



REPURPOSING A PRE-TRAINED MODEL

You will follow the general machine learning workflow.

1. Examine and understand the data
2. Build an input pipeline
3. Compose the model
   - Load in the pretrained base model (and pretrained weights)
   - Stack the classification layers on top
4. Train the model
5. Evaluate model

USING PRE-TRAINED MODELS *HANDS-ON*

REFERENCES

- https://www.datarobot.com/blog/introduction-to-computer-vision-what-it-is-and-how-it-works/
- https://towardsdatascience.com/a-friendly-introduction-to-computer-vision-with-artificial-neural-networks-d2a38acc047c
- https://priyalwalpita.medium.com/convolutional-neural-networks-for-artificial-vision-455be7c85d15
- https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network
- https://docs.microsoft.com/en-us/learn/modules/intro-computer-vision-tensorflow/6-transfer-learning
- https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/
- https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751#:~:text=Transfer%20learning%20is%20a%20popular,when%20solving%20a%20different%20problem.
- https://www.tensorflow.org/tutorials/images/transfer_learning
- https://www.kdnuggets.com/2022/01/transfer-learning-image-recognition-natural-language-processing.html

THANKS!

Do you have any questions?