Scope and opportunities in the areas of

Artificial Intelligence - Computer Vision

Dr. Sarwan Singh

Scientist – D, NIELIT Chandigarh

AI-Roadmap to Computer Vision

Step 1 :

- Python & Statistics : Basics of computer vision

Step 2 :

- Solving Image Classification : Basic Machine Learning Image pre-processing Techniques

Step 3 :

- Introduction to Keras & Neural Networks: Fundamentals of Neural Networks (keras)

Step 4 :

- Understanding CNNs, Transfer Learning : Convolutional Neural Networks(CNNs) Transfer learning

Step 5 :

- Object Detection Algorithms

Step 6 :

- Image Segmentation & Attention Models

Step 7 :

- Deep Learning Tools : PyTorch & TensorFlow

Step 8 :

- Generative Adversial Networks (GANs)

AI-Roadmap to Computer Vision

1. Introduction to computer vision
2. Getting started with Image Data
   - Reading, stacking
   - Understanding Pixels/ colorspace
   - Image processing
   - Edge/blur/ HOG…
3. CNN
   - Need, filter…
   - Neural network
   - Pooling, maths..
   - CNN Visualization
4. Transfer Learning
5. Pretrained Models
6. Inception
7. Resnets
   - Densnet
   - CSRnet
8. Object Detection
9. Image Generation
   - GAN

AI-Roadmap to Computer Vision

1. Python & Statistics : Basics of computer vision
2. Solving Image Classification : Basic Machine Learning Image pre-processing Techniques
3. Introduction to Keras & Neural Networks: Fundamentals of Neural Networks (keras)
4. Understanding CNNs, Transfer Learning : Convolutional Neural Networks(CNNs) Transfer learning
5. Object Detection Algorithms
6. Image Segmentation & Attention Models
7. Deep Learning Tools : PyTorch & TensorFlow
8. Generative Adversial Networks (GANs)

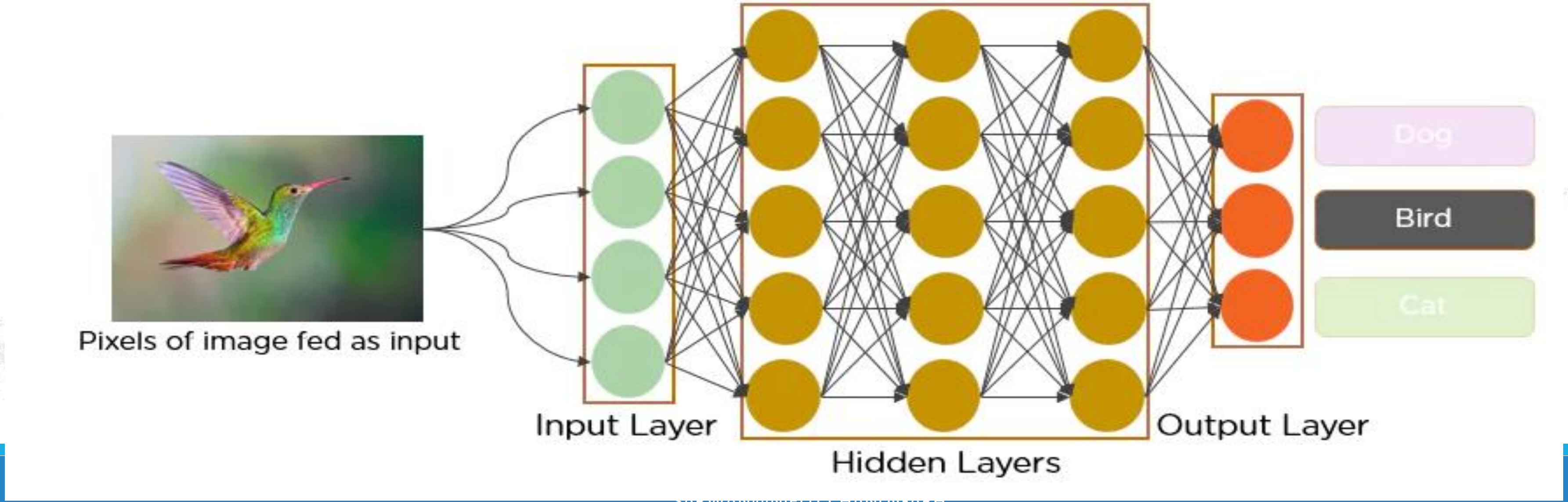
- What are Convolutional Neural Network (CNN) ?
- Artificial Neurons Role in CNN
- Feature Extraction in CNN
- Background of Convolutional neural networks (CNNs)
- What is a Pooling Layer?
- Limitations of Convolutional neural networks (CNNs)

What is a Convolutional Neural Network (CNN)

- A Convolutional Neural Network (CNN), also known as ConvNet, is a specialized type of deep learning algorithm mainly designed for tasks that necessitate object recognition, including image classification, detection, and segmentation.
- CNNs are employed in a variety of practical scenarios, such as autonomous vehicles, security camera systems, and others.

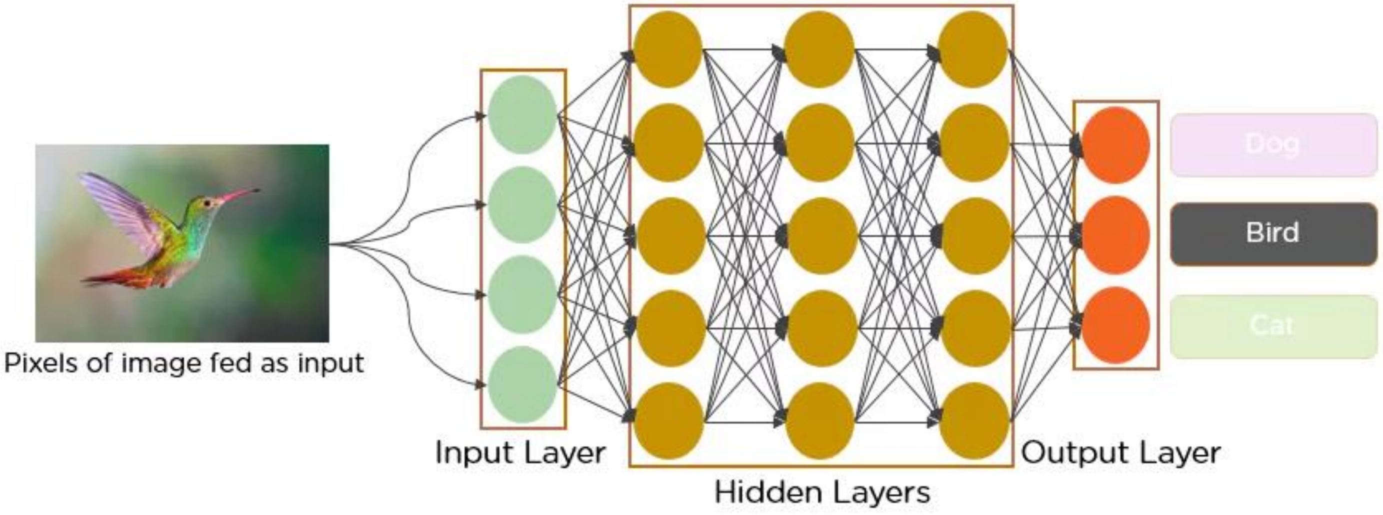
CNN – Introduction

- Since the 1950s, AI researchers have sought to develop systems that can understand visual data. This effort gave birth to the field of Computer Vision.
- In 2012, a significant breakthrough occurred when researchers from the University of Toronto developed AlexNet, an AI model that significantly outperformed previous image recognition algorithms.



CNN – Introduction

- AlexNet, created by Alex Krizhevsky, won the 2012 ImageNet contest with 85% accuracy, far surpassing the runner-up's 74%.
- This success was driven by CNNs, a type of neural network that mimics human vision.



Importance of CNN

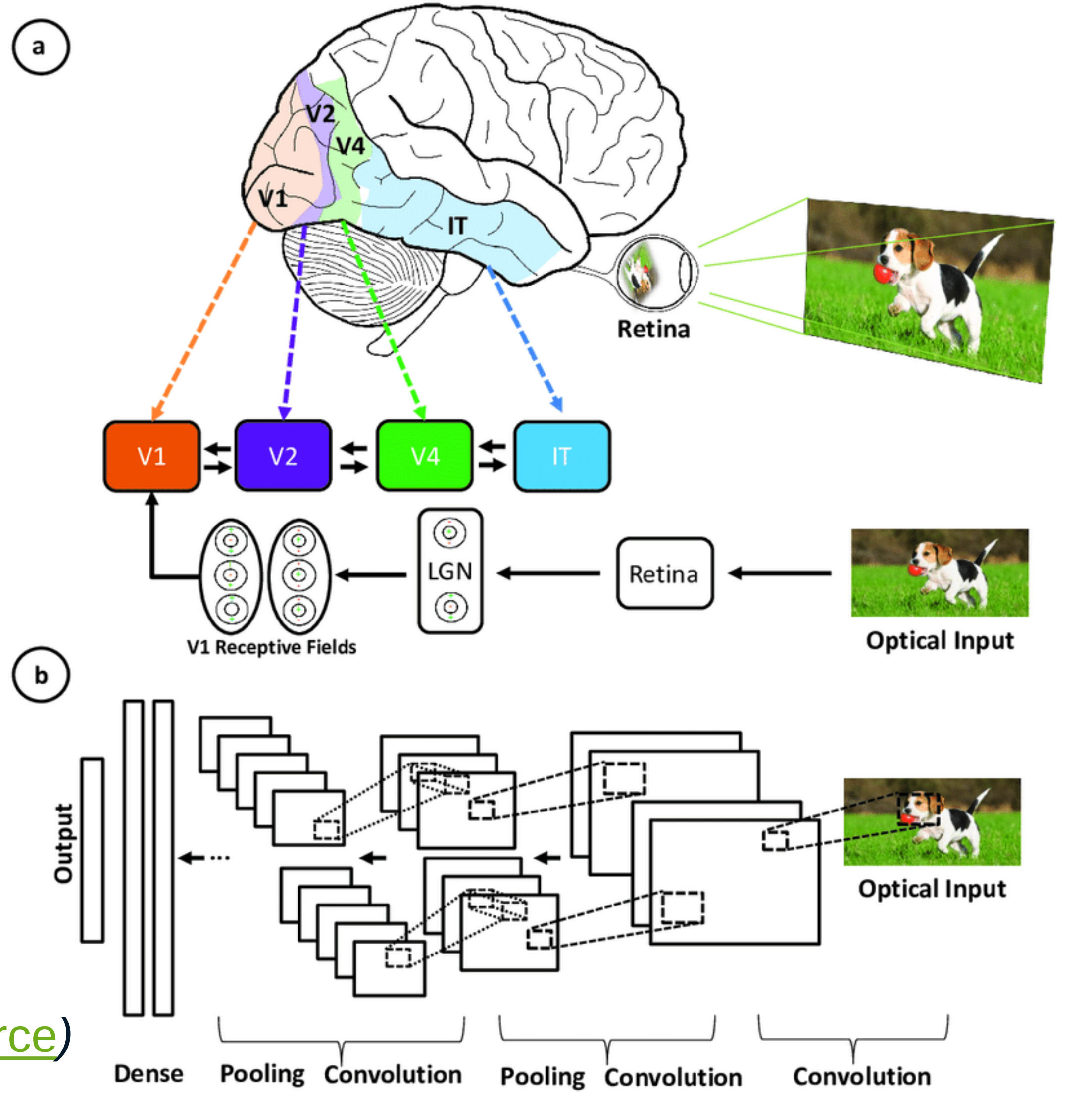
- CNNs are distinguished from classic machine learning algorithms such as **SVMs** and **decision trees** by their ability to autonomously extract features at a large scale, bypassing the need for manual feature engineering and thereby enhancing efficiency.
- The convolutional layers grant CNNs their translation-invariant characteristics, empowering them to identify and extract patterns and features from data irrespective of variations in position, orientation, scale, or translation.
- A variety of pre-trained CNN architectures, including VGG-16, ResNet50, Inceptionv3, and EfficientNet, have demonstrated top-tier performance. These models can be adapted to new tasks with relatively little data through a process known as fine-tuning.
- Beyond image classification tasks, CNNs are versatile and can be applied to a range of other domains, such as natural language processing, time series analysis, and speech recognition.

Inspiration Behind CNN and Parallels with the Human Visual System

- Convolutional neural networks were inspired by the layered architecture of the human visual cortex.

some key similarities and differences:

- **Hierarchical architecture:** Both CNNs and the visual cortex have a hierarchical structure, with simple features extracted in early layers and more complex features built up in deeper layers.

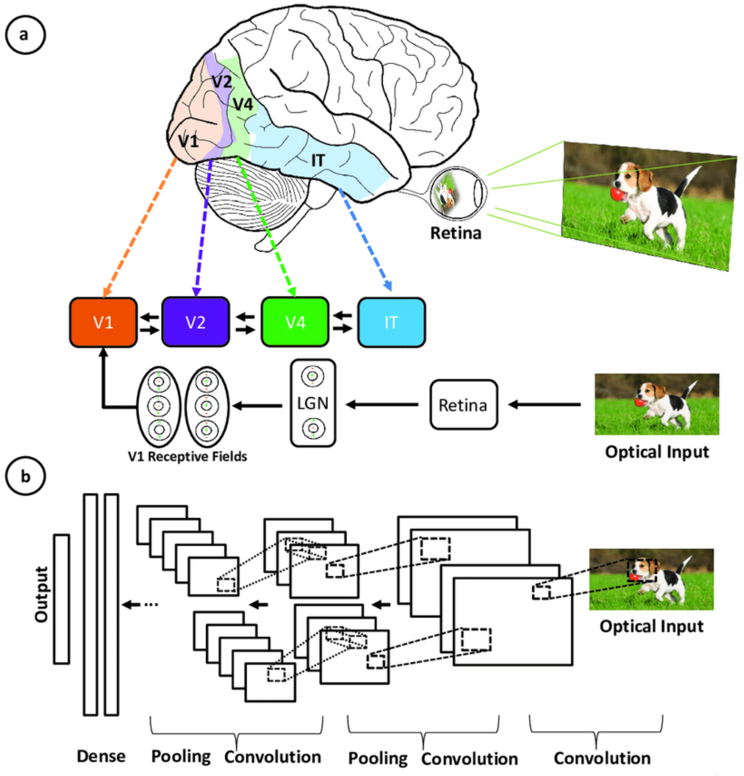


(source)

Inspiration Behind CNN and Parallels with the Human Visual System

Some key similarities and differences:

- **Local connectivity:** Neurons in the visual cortex only connect to a local region of the input, not the entire visual field. Similarly, the neurons in a CNN layer are only connected to a local region of the input volume through the convolution operation. This local connectivity enables efficiency.
- **Translation invariance:** Visual cortex neurons can detect features regardless of their location in the visual field. Pooling layers in a CNN provide a degree of translation invariance by summarizing local features.

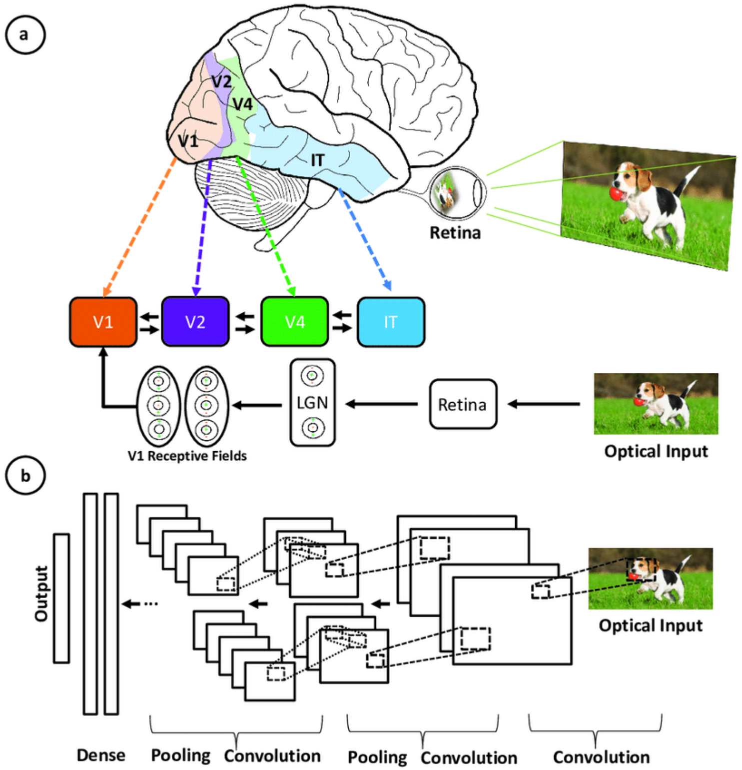


(source)

Inspiration Behind CNN and Parallels with the Human Visual System

some key similarities and differences:

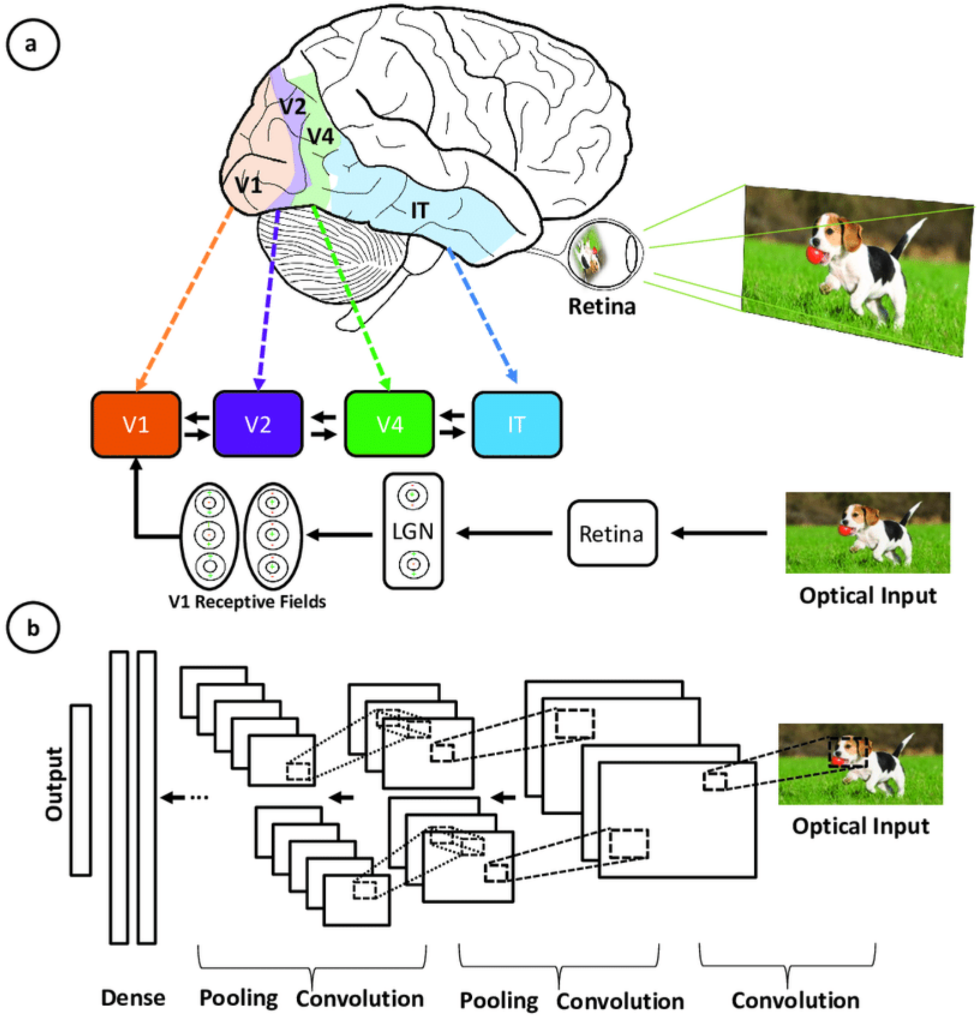
- **Multiple feature maps:** At each stage of visual processing, there are many different feature maps extracted. CNNs mimic this through multiple filter maps in each convolution layer.
- **Non-linearity:** Neurons in the visual cortex exhibit non-linear response properties. CNNs achieve non-linearity through activation functions like ReLU applied after each convolution.



(source)

Inspiration Behind CNN and Parallels with the Human Visual System

CNNs mimic the human visual system but are simpler, lacking its complex feedback mechanisms and relying on supervised learning rather than unsupervised, driving advances in computer vision despite these differences



(source)

CNN – Introduction

- Over the years, CNNs have become fundamental in computer vision tasks such as image classification, object detection, and segmentation.
- Modern CNNs are implemented using programming languages like Python and leverage advanced techniques to extract and learn features from images.
- Hyperparameters, optimization techniques, and regularization methods are crucial for training these models effectively.
- Since AlexNet, numerous improvements and new architectures like VGG, ResNet, and EfficientNet have been developed, pushing the boundaries of what CNNs can achieve.
- Today, CNNs are essential in many applications, from autonomous driving to medical image analysis.

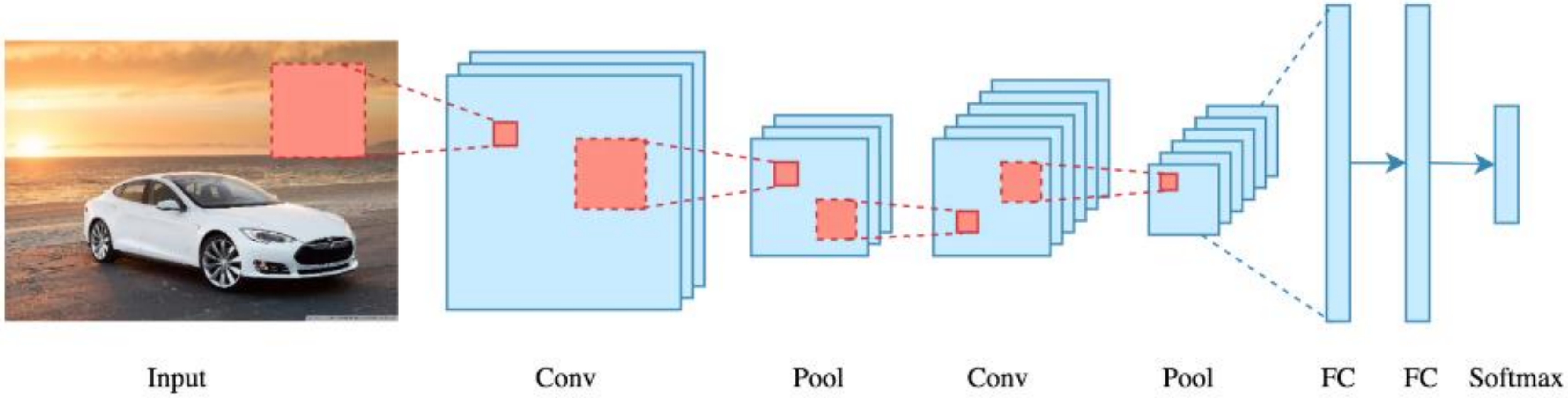
Key Components of a CNN

Convolutional Neural Network is made of four main parts :

- Convolutional layers
- Rectified Linear Unit (ReLU for short)
- Pooling layers
- Fully connected layers

Convolutional Neural Network (CNN)

- In deep learning, a Convolutional Neural Network (CNN/ConvNet) is a class of deep neural networks, most commonly applied to analyze visual imagery.
- The CNN architecture uses a special technique called Convolution instead of relying solely on matrix multiplications like traditional neural networks.
- Convolutional networks use a process called convolution, which combines two functions to show how one changes the shape of the other.



- The bottom line is that the role of the convolutional networks is to reduce the images into a form that is easier to process, without losing features that are critical for getting a good prediction.
- *The bottom line is that the role of the ConvNet is to reduce the images into a form that is easier to process, without losing features that are critical for getting a good prediction.*

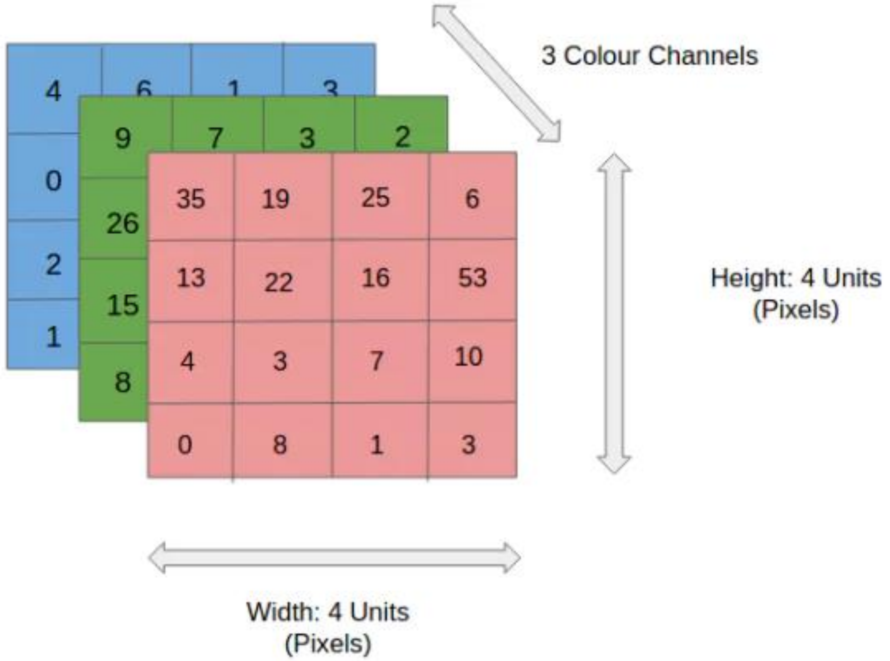
Background

- CNNs were first developed and used around the 1980s. The most that a Convolutional Neural Network (CNN) could do at that time was recognize handwritten digits.
- It was mostly used in the postal sector to read zip codes, pin codes, etc. The important thing to remember about any deep learning model is that it requires a large amount of data to train and also requires a lot of computing resources.
- This was a major drawback for CNNs at that period, and hence CNNs were only limited to the postal sectors and it failed to enter the world of machine learning.
- Backpropagation, the algorithm used to train neural networks, was also computationally expensive at the time.
- In 2012, Alex Krizhevsky realized that it was time to bring back the branch of deep learning that uses multi-layered neural networks.
- The availability of large sets of data, more specific ImageNet datasets with millions of labeled images, and an abundance of computing resources enabled researchers to revive CNNs.

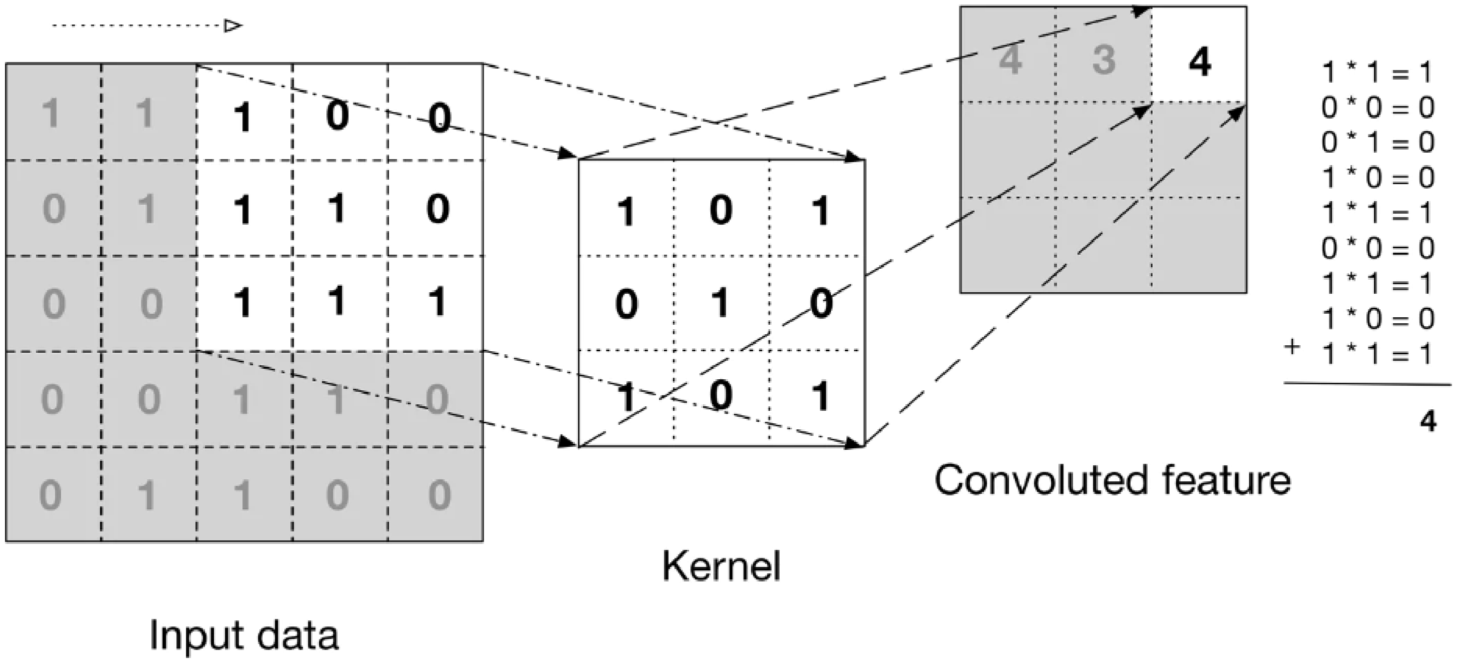
CNN Working

Image Storage

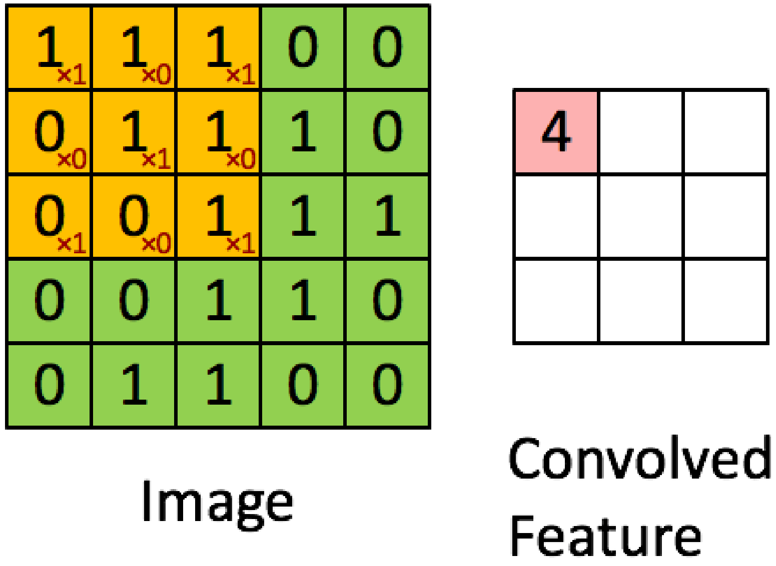
- An RGB image is nothing but a matrix of pixel values having three planes whereas a grayscale image is the same but it has a single plane.



How CNN works for grayscale images



- Image shows what a convolution is. We take a filter/kernel(3×3 matrix) and apply it to the input image to get the convolved feature.
- This convolved feature is passed on to the next layer.



How CNN works for color images

- The number of parameters in a CNN layer depends on the size of the receptive fields (filter kernels) and the number of filters. Each neuron in a CNN layer receives inputs from a local region of the previous layer, known as its receptive field.
- The receptive fields move over the input, calculating dot products and creating a convolved feature map as the output. Usually, this map then goes through a rectified linear unit (ReLU) activation function. Classic CNN architectures like LeNet and more modern ones like ResNet employ this fundamental principle.
- Convolutional neural networks are composed of multiple layers of artificial neurons.

0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #1

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #2

0	1	1
0	1	0
1	-1	1

Kernel Channel #3

$$308 \quad + \quad -498 \quad + \quad 164 \quad + 1 = -25$$

Bias = 1

Output

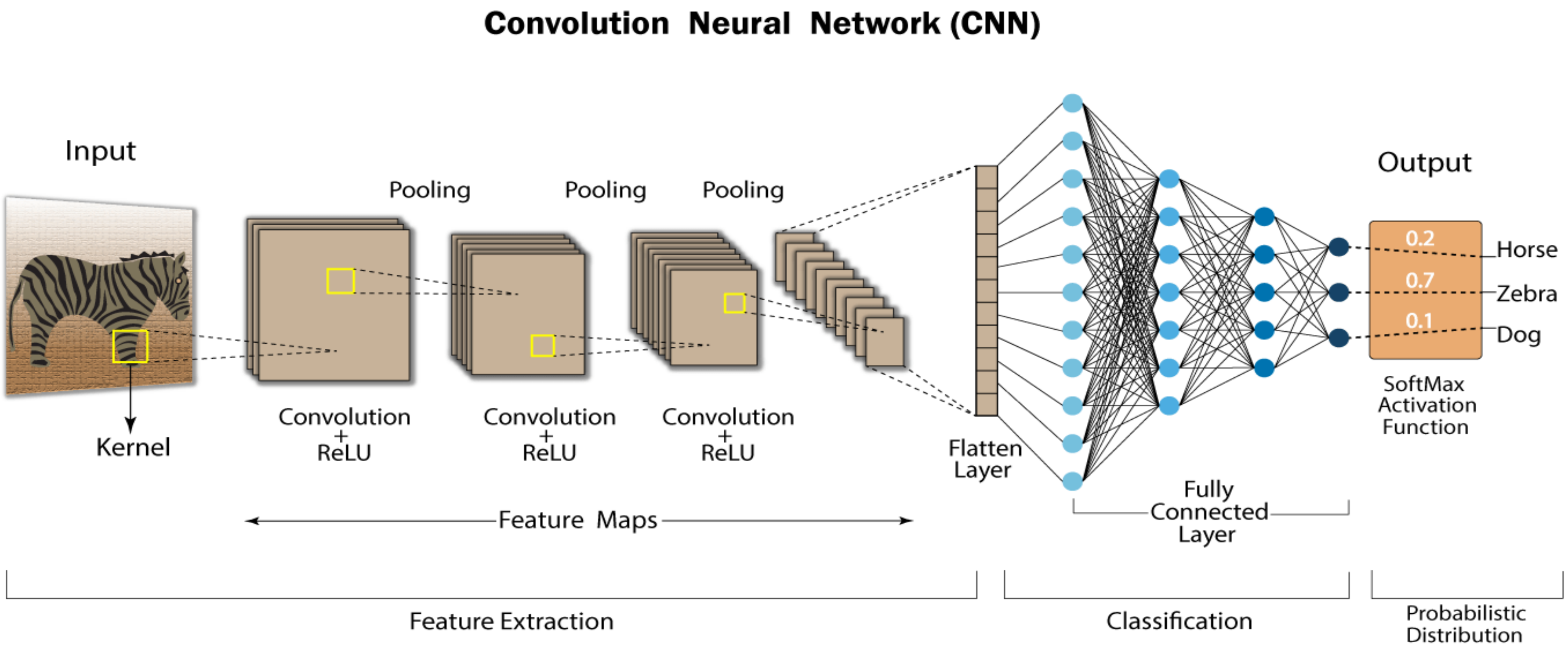
-25				...
				...
				...
				...
...	...	...	...	...

Convolutional Neural Networks – CNN

- ***Convolutional Neural Networks*** *also known as CNNs or ConvNets, are a type of* ***feed-forward artificial neural network*** *whose connectivity structure is inspired by the organization of the animal visual cortex.*
- Small clusters of cells in the visual cortex are sensitive to certain areas of the visual field. Individual neuronal cells in the brain respond or fire only when certain orientations of edges are present.
- Some neurons activate when shown vertical edges, while others fire when shown horizontal or diagonal edges.
- A convolutional neural network is a type of Artificial Neural Network(ANN) used in deep learning to evaluate visual information.
- These networks can handle a wide range of tasks involving images, sounds, texts, videos, and other media. Professor Yann LeCunn of Bell Labs created the first successful convolution networks in the late 1990s.

Convolutional Neural Networks – CNN

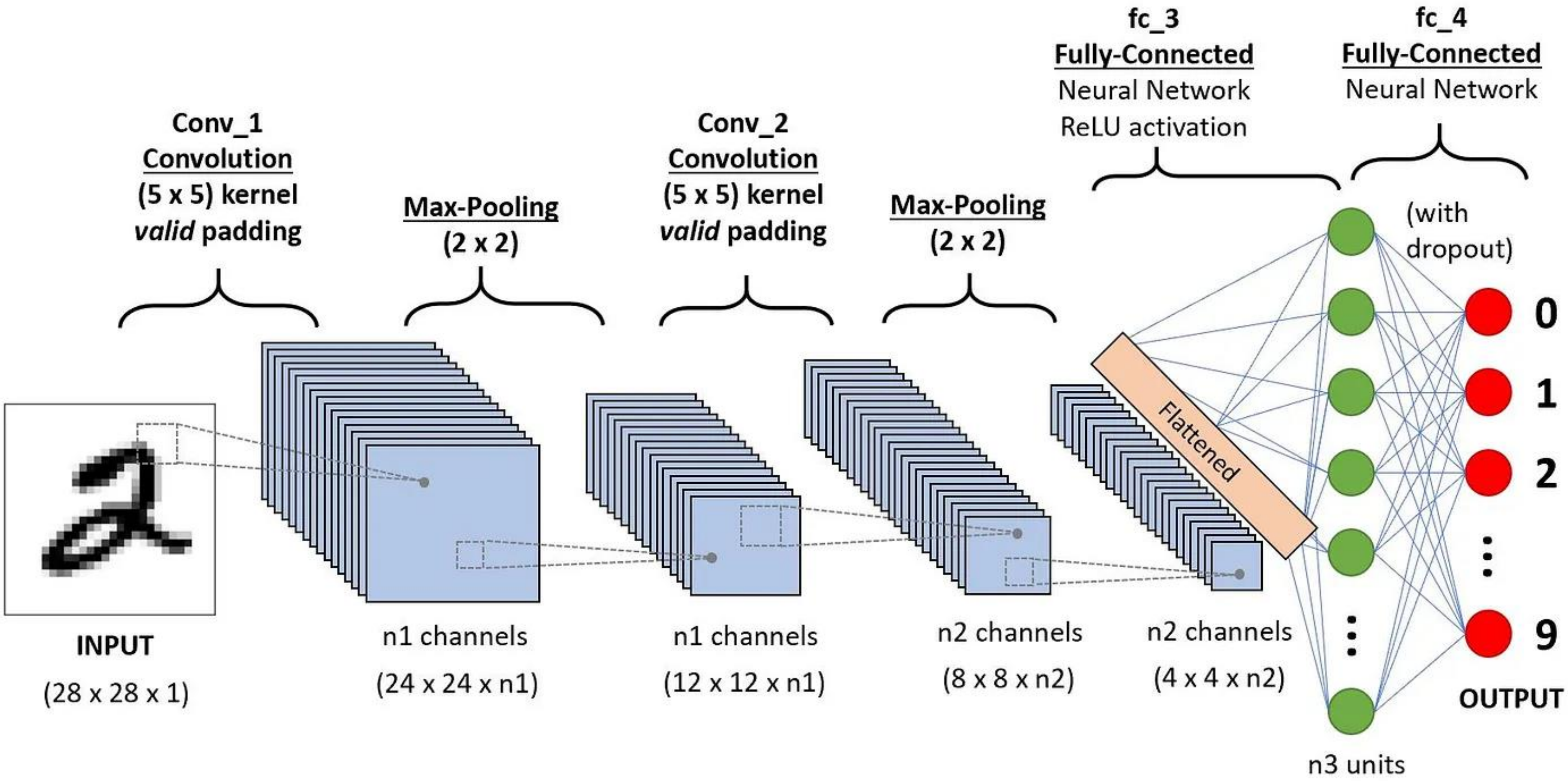
- ***CNNs*** have an input layer, an output layer, numerous hidden layers, and millions of parameters, allowing them to learn complicated objects and patterns.



Convolutional Neural Networks – CNN

- It uses convolution and pooling processes to sub-sample the given input before applying an activation function, where all of them are hidden layers that are partially connected, with the completely connected layer at the end resulting in the output layer.
- The output shape is similar to the size of the input image.

Architecture of the CNNs applied to Digit Recognition



source

Deep Learning Frameworks for CNNs

- **TensorFlow** is an open-source deep learning framework developed by Google and released in 2015. It offers a range of tools for machine learning development and deployment.
- **Keras** is a high-level neural network framework in Python that enables rapid experimentation and development. It's open-source and can be used within other frameworks like TensorFlow, CNTK, and Theano.

Deep Learning Frameworks for CNNs

- **Pytorch** - Released by Facebook's AI research division in 2017, it's designed for applications in Natural Language Processing and is noted for its dynamic computational graph and memory efficiency.

	Tensorflow	Pytorch	Keras
API Level	Both(High and Low)	Low	High
Architecture	Not easy to use	Complex, less readable	Simple, concise, readable
Datasets	Large datasets, high performance	Large datasets, high performance	Smaller datasets
Debugging	Difficult to conduct debugging	Good debugging capabilities	Simple network, so debugging is not often needed
Pretrained models?	Yes	Yes	Yes
Popularity	Second most popular of the three	Third most popular of the three	Most popular of the three
Speed	Fast, high-performance	Fast, high-performance	Slow, low performance
Written in	C++, CUDA, Python	Lua	Python