Unit 5: Computer Vision

Title: Computer Vision	**Approach:** Practical Implementation
Summary: Computer Vision is a branch of artificial intelligence that enables machines to interpret and understand visual information from the real world. This unit provides an in-depth exploration of various methodologies, and applications in computer vision, equipping students with the skills necessary to analyze and process visual data.	
Objectives: 1. To introduce students to the basic principles and techniques of computer vision. 2. To familiarize students with common algorithms and tools used in image processing and analysis. 3. To enable students to apply computer vision techniques to solve real-world problems. 4. To foster critical thinking and problem-solving skills in the domain of computer vision.	
Learning Outcomes: 1. Understand the fundamental concepts and theories underlying computer vision. 2. Implement basic and advanced image processing techniques using programminglanguages such as Python. 3. Apply computer vision techniques to tasks such as object detection, image segmentation, and feature extraction. 4. Develop ideas to solve real-world problems leveraging computer vision technologies.	
Pre-requisites: Essential understanding of Artificial Intelligence	
Key-concepts: Image processing, feature extraction, object detection & recognition.	

5.1: Introduction

In the previous chapter, you studied the concepts of Artificial Intelligence for Data Sciences. It is a concept to unify statistics, data analysis, machine learning and their related methods to understand and analyse actual phenomena with data.

As we all know, artificial intelligence is a technique that enables computers to m1m1c human intelligence. As humans, we can see things, analyse them and then do the required action based on what we see.

But can machines do the same? Can machines have the eyes that humans have? If you answered yes, then you are right. The Computer Vision domain of Artificial Intelligence, enables machines to see• through images or visual data, process and analyse them on the basis of algorithms and methods to analyse actual phenomena with images.

Now before we get into the concepts of Computer Vision, let us experience this domain with the help of the following game:

* Emoji Scavenger Hunt: https://emojiscavengerhunt.withgoogle.com/

Go to the link and try to play the game Emoji Scavenger Hunt. The challenge here is to find 8 items within the time limit to pass. Did you manage to win?

_____

_____

What was the strategy that you applied to win this game?

_____

_____

_____

_____

_____

Was the computer able to identify all the items you brought in front of it?

_____

_____

_____

_____

_____

Did the lighting of the room affect the identifying of items by the machine?

_____

_____

_____

_____

_____

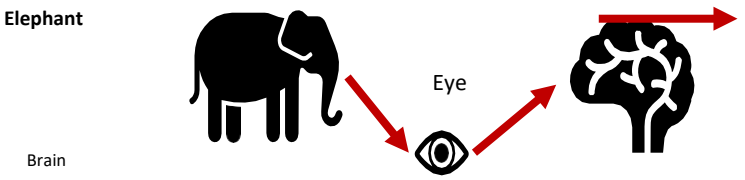
_____

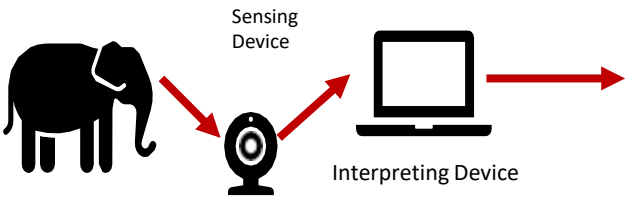
A Quick Overview of Computer Vision!

Computer vision is the process of extraction of information from images, text, videos, etc.

A system that can process, analyze and make sense of visual data in the same way as humans do.

Human Vision System

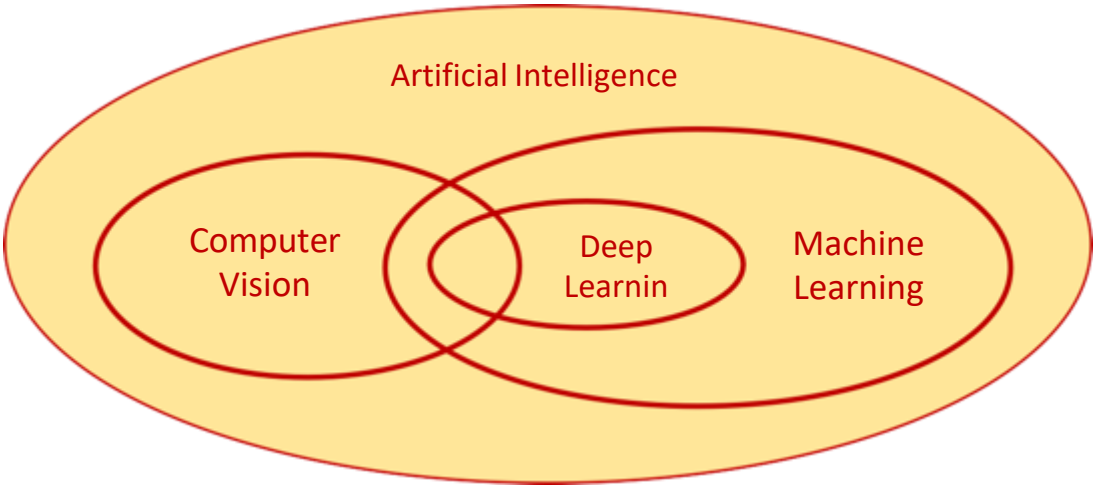




Computer Vision and Artificial Intelligence

Computer vision is a field of artificial intelligence (AI).

AI enables computers to think, and computer vision enables AI to see, observe and make sense of visual data(like images & videos).

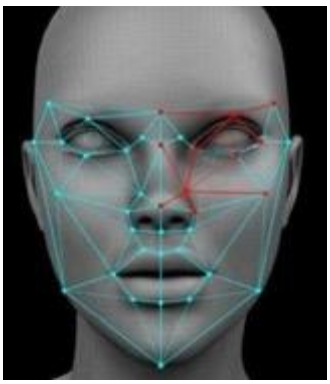


Computer Vision Vs. Image Processing

Computer Vision	Image Processing
• Computer vision deals with extracting information from the input images or videos toinfer meaningful information and understanding them to predict the visual input • Computer Vision is a superset of ImageProcessing. • Examples - Object detection, Handwritingrecognition, etc.	• Image processing is mainly focused on processing the raw input images to enhance them or preparing them to do other tasks • Image Processing is a subset of ComputerVision. • Examples - Rescaling image, Correcting brightness, Changing tones, etc.

5.1 Applications of Computer Vision

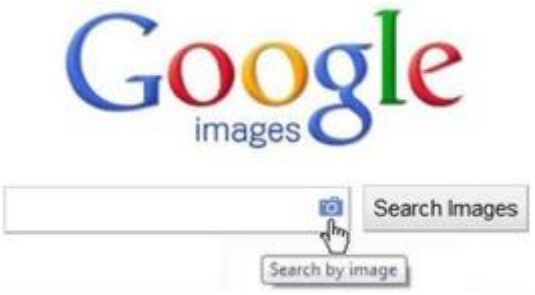
The concept of computer vision was first introduced in the 1970s. All these new applications of computer vision excited everyone. Having said that, computer vision technology advanced enough to make these applications available to everyone at ease today. However, in recent years the world witnessed a significant leap in technology that has put computer vision on the priority list of many industries. Let us look at some of them:



Facial Recognition*: With the advent of smart cities and smart homes, Computer Vision plays a vital role in making the home smarter. Security being the most important application involves the use of Computer Vision for facial recognition. It can be either guest recognition or log maintenance of the visitors. It also finds its application in schools for an attendance system based on facial recognition of students.

Face Filters*: Modern-day apps like Instagram and Snapchat have a lot of features based on the

usage of computer vision. The application of face filters is one among them. Through the camera, the machine or the algorithm is able to identify the facial dynamics of the person and applies the facial filter selected.



Google's Search by Image*: The maximum amount of searching for data on Google's search engine comesfrom textual data, but at the same time it has an interesting feature of getting search results through an image. This uses Computer Vision as it compares different features of the input image to the database of images and gives us the search result while at the same time analysing various features of the image.

Computer Vision in Retail*: The retail field has been one of the fastest-growing fields and at the same time is using Computer Vision for making the user experience more fruitful. Retailers can use Computer Vision techniques to track customers' movements through stores, analyse navigational routes and detect walking patterns.

Inventory Management is another such application. Through security camera image analysis, a Computer Vision algorithm can generate a very accurate estimate of the items available in the store. Also, it can analyse the use of shelf space to identify suboptimal configurations and suggest better item placement.

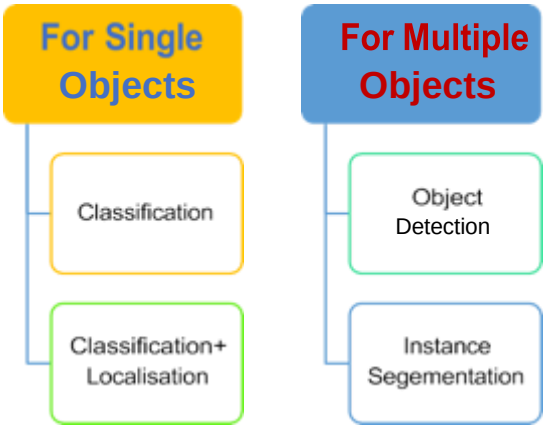
Self-Driving Cars: Computer Vision is the fundamental technology behind the development of autonomous vehicles. Most leading car manufacturers in the world are reaping the benefits of investing in artificial intelligence for developing on-road versions of hands-free technology. This involves the process of identifying the objects, getting navigational routes and also at the same time environment monitoring.

Medical Imaging*: For the last decades, computer supported medical imaging application has been a trustworthy help for physicians. It doesn't only create and analyse images, but also becomes an assistant and helps doctors with their interpretation. The application is used to read and convert 20 scan images into interactive 30 models that enable medical professionals to gain a detailed understanding of a patient's health condition.

Google Translate App*: All you need to do to read signs in a foreign language is to point your phone's camera at the words and let the Google Translate app tell you what it means in your preferredlanguage almost instantly. By using optical character recognition to see the image and augmented reality to overlay an accurate translation, this is a convenient tool that uses Computer Vision

5.2 Computer Vision Tasks

The various applications of Computer Vision are based on a certain number of tasks that are performed to get certain information from the input image which can be directly used for predictionor forms the base for further analysis. The tasks used in a computer vision application are:



Classification

The image Classification problem is the task of **assigning an input image one label from a fixed set ofcategories.** This is one of the core problems in CV that, despite its simplicity, has a large variety of practical applications.

Classification+ Localisation

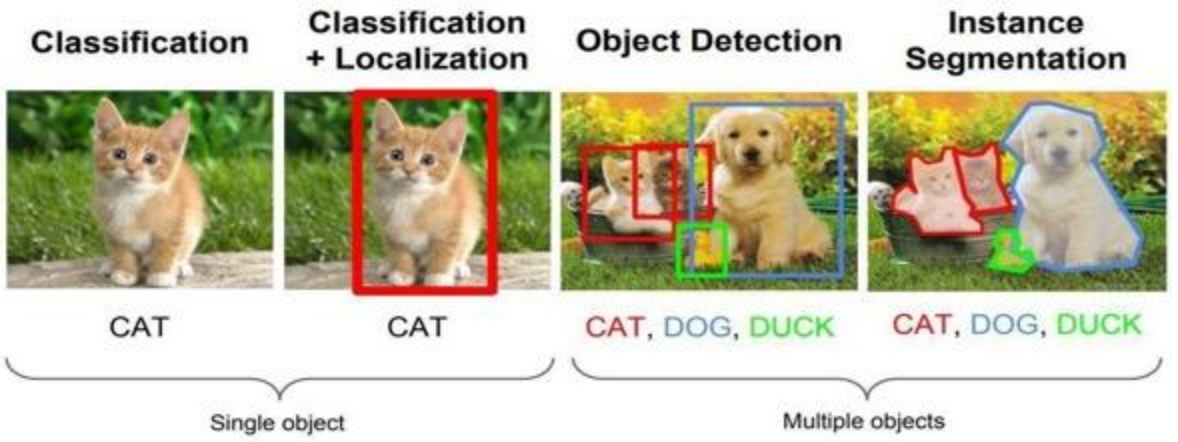
This is the task that involves both processes of **identifying what object is present** in the image and at the same time **identifying at what location** that object is present in that image. It is used only for single objects.

Object Detection

Object detection is the process of **finding instances of real-world objects such as faces, bicycles, and buildings in images or videos.** Object detection algorithms typically use extracted features andlearning algorithms to recognize instances of an object category. It is commonly used in applications such as image retrieval and automated vehicle parking systems.

Instance Segmentation

Instance Segmentation is the process of detecting instances of the objects, giving them a category, andthen giving each pixel a label based on that. A segmentation algorithm takes an image as input and outputs a collection of regions (or segments).



Basics of Images

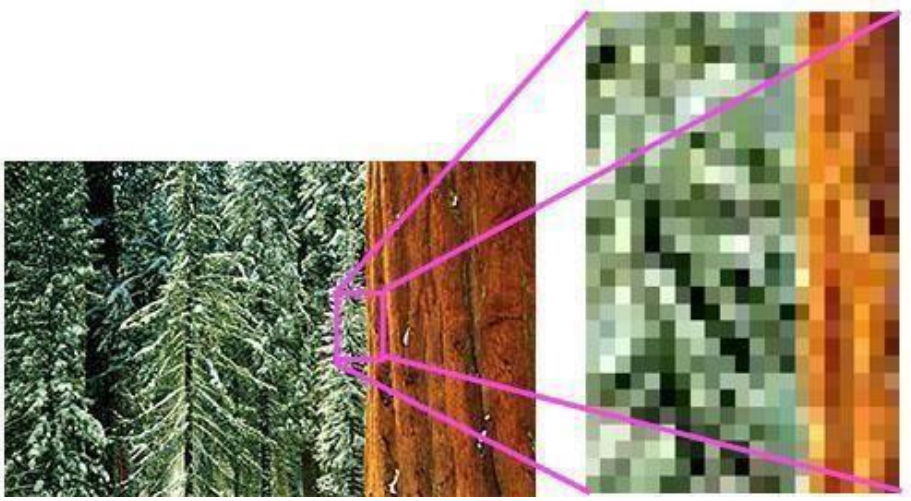
We all see a lot of images around us and use them daily either through our mobile phones or computer system. But do we ask some basic questions to ourselves while we use them on regular basis?

Don't know the answer yet? Don't worry, in this section, we will study the basics of an image:

Basics of Pixels

The word "pixel" means a picture element. Every photograph, in digital form, is made up of pixels. They are the smallest unit of information that make up a picture. Usually round or square, they are typically arranged in a 2-dimensional grid.

In the image below, one portion has been magnified many times over so that you can see its composition in pixels. As you can see, the pixels approximate the actual image. The more pixels you have, the more closely the image resembles the original.



Resolution

The number of pixels in an image is sometimes called the *resolution.* When the term is used to describepixel count, one convention is to express resolution as the width by the height, for example, a monitor resolution of 1280x1024. This means there are 1280 pixels from one side to the other, and 1024 from top to bottom.

Another convention is to express the number of pixels as a single number, like a 5 mega pixel camera (a megapixel is a million pixels). This means the pixels along the width multiplied by the pixels along the height of the image taken by the camera equals 5 million pixels. In the case of our 1280x1024 monitors, it could also be expressed as 1280 x 1024 = 1,310,720, or 1.31 mega pixels.

Pixel value

Each of the pixels that represent an image stored inside a computer has a *pixel value* that describes how bright that pixel is, and/or what colour it should be. The most common *pixel format* is the *byte image,* where this number is stored as an 8-bit integer giving a range of possible values from Oto 255.Typically, zero is to be taken as no colour or black and 255 is taken to be full colour or white. Why do we have a value of 255?
In computer systems, computer data is in the form of ones and zeros, which we call the binary system. Each bit in a computer system can have either a zero or a one. Since each pixel uses 1 byte of an image, which is equivalent to 8 bits of data. Since each bit can have two possible values which tell us that the 8 bits can have 255 possibilities of values that starts from 0 and ends at 255.

Number of bits	Different patterns	No. of patterns	No. of patterns
1	0 1	2^1	2
2	00 01 10 11	2^2	4
3	000 001 010 100 011 101 110 111	2^3	8

$2^8 = 256$

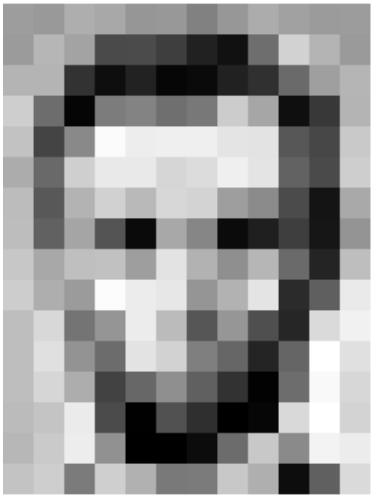
Here ^, represents exponent (2 raised to the power 8)

Grayscale Images

Grayscale images are images that have a range of shades of gray without apparent colour. The darkest possible shade is black, which is the total absence of colour or zero value of pixel. The lightest possible shade is white, which is the total presence of colour or 255 value of a pixel. Intermediate shades of gray are represented by equal brightness levels of the three primary colours.

A grayscale has each pixel of size 1 byte having a single plane of 2d array of pixels. The size of a grayscale image is defined as the Height x Width of that image.

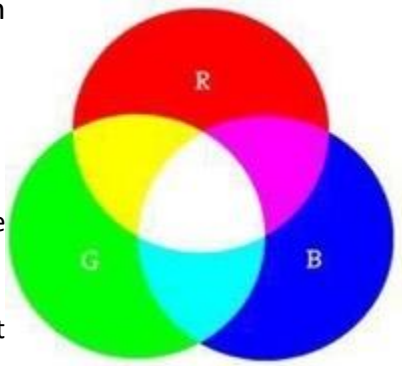
Let us look at an image to understand grayscale images.



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

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

Here is an example of a grayscale image. As you check, the value of pixels is within the range of 0-255.The computers store the images we see in the form of these numbers.



RGB Images

All the images that we see around us are coloured images. These images are made up of three primary colours Red, Green, and Blue.

All the colours that are present can be made by combining different intensities of red, green, and blue.

Let us experience!

Go to this online link https://www.w3schools.com/colors/colors_rgb.asp. On the basis of this online tool,try and answer all the below mentioned questions.

1) What is the output colour when you put R=G=B=255?

_____

2) What is the output colour when you put R=G=B=0?

_____

3) How does the colour vary when you put either of the three as 0 and then keep on varying theother two?

_____

_____

_____

4) How does the output colour change when all the three colours are varied in same proportion?

_____
_____
_____
_____

5) What is the RGB value of your favourite colour from the colour palette?

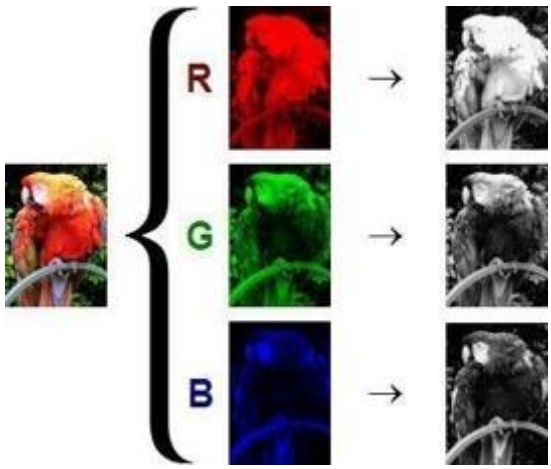
_____

Were you able to answer all the questions? If yes, then you would have understood how every colour we see around is made.

Now the question arises, how do computers store RGB images? Every RGB image is stored in the form of three different channels called the R channel, G channel, and the B channel.

Each plane separately has many pixels with each pixel value varying from Oto 255. All the three planeswhen combined form a colour image. This means that in an RGB image, each pixel has a set of three different values which together give colour to that particular pixel.

For Example,



As you can see, each colour image is stored in the form of three different channels, each having different intensity. All three channels combine to form a colour we see.

In the above given image, if we split the image into three different channels, namely Red (R), Green {G) and Blue (B), the individual layers will have the following intensity of colours of the individual pixels. These individual layers when stored in the memory looks like the image on the extreme right. The images look in the grayscale image because each pixel has a value intensity of O to 255 and as studied earlier, 0 is considered as black or no presence of colour and 255 means white or full presenceof colour. These three individual RGB values when combined form the colour of each pixel.

Therefore, each pixel in the RGB image has three values to form the complete colour.

Task:

Go to the following link www.piskelapp.com and create your pixel art. Try and make a GIF using the online app for your pixel art.

5.3 No-Code AI Tools:

Introduction to Lobe

- Lobe.ai is an Auto-ML tool, which means that it is a no-code AI tool
- It works with image classification and allows a set of images with labels and will

automatically find the most optimal model to classify the images

Introduction to Teachable Machine

- Teachable Machine is an AI, Machine Learning, and Deep Learning tool that was developed by Google in 2017
- It runs on top of tensorflow.js which was also developed by the same company
- It is a web-based tool that allows training of a model based on different images, audio, orposes given as input through webcam or pictures

Activity Time: Build a Smart Sorter
Purpose: Using CV is to automate and enhance sorting processes through computer vision technology.

- Form groups of 4 members
- Find images of Bottles, Cans and Paper online or from around.
- Visit the No-code AI tool
- Ensure to build 3 different classes [Bottles, Cans and Paper].
- Train the model
- Finally, test the classifier!

Orange Data Mining Tool:

Let's work on a real-world Classification Model: Coral Bleaching (Use Case Walkthrough)

Activity Time: Build a classification model
Purpose: Developing a classification model for early identification of coral bleaching to safeguard marine ecosystems.

What Are Coral Reefs?

- Coral can be found in tropical ocean waters around the world.
- Coral reefs are large underwater structures composed of the skeletons of marine invertebrates called coral.
- Corals are an integral part of aquatic life.

What causes Coral Bleaching?

Coral bleaching has caused unbalanced scenarios in aquatic life. So, detecting bleached corals at early stages prevents aquatic life from disaster

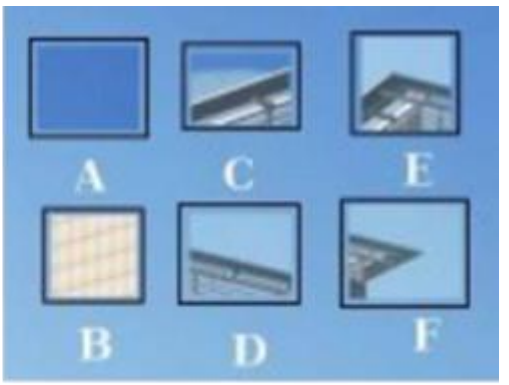
Use Case Walkthrough - Steps involved in project development & required dataset can be checked using the links and QR code given below:

Short link: https://bit.ly/orange_computer_vision
Long Link: https://drive.google.com/drive/folders/1ppJ4d-8yOFJ2G22rHHpjNrK0ejdIAe5Q?usp=sharing
or scan the QR code below

5.4 Image Features

In computer vision and image processing, a **feature** is a piece of information that is relevant for solving the computational task related to a certain application. Features may be specific structures in the image such as points, edges, or objects.

For example:
Imagine that your security camera is capturing an image. At the top of the image, we are given six smallpatches of images. Our task is to find the exact location of those image patches in the image.
Take a pencil and mark the exact location of those patches in the image.



Were you able to find the exact location of all the patches?

_____

Which one was the most difficult to find?

_____

_____

Which one was the easiest to find?

_____

_____

Let's Reflect:

Let us take individual patches into account at once and then check the exact location of those patches. **For Patch A and B:** The patch A and Bare flat surfaces in the image and are spread over a lot of area. They can be present at any location in a given area in the image.
For Patch C and D: The patches C and D are simpler as compared to A and B. They are edges of a building and we can find an approximate location of these patches but finding the

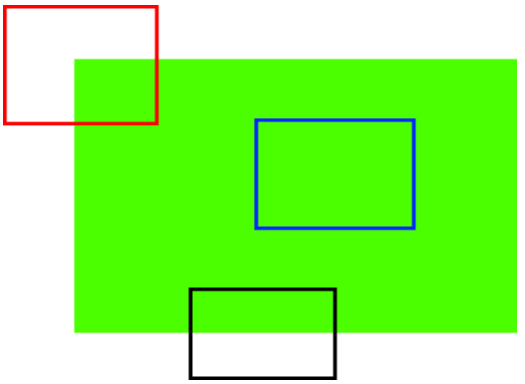
exact location isstill difficult. This is because the pattern is the same everywhere along the edge.

For Patch E and F: The patches E and Fare the easiest to find in the image. The reason is that E and F are some corners of the building. This is because at the corners, wherever we move this patch it will look different.

Conclusion

In image processing, we can get a lot of features from the image. It can be either a blob, an edge, or a corner. These features help us to perform various tasks and then get the analysis done based on the application. Now the question that arises is which of the following are good features to be used? As you saw in the previous activity, the features having the corners are easy to find as they can be found only at a particular location in the image, whereas the edges are spread over a line or an edge look the same all along. This tells us that the corners are always good features to extract from an image followed by the edges.

Let's look at another example to understand this. Consider the images given below and apply the concept of good features for the following.



In the above image how would we determine the exact location of each patch?

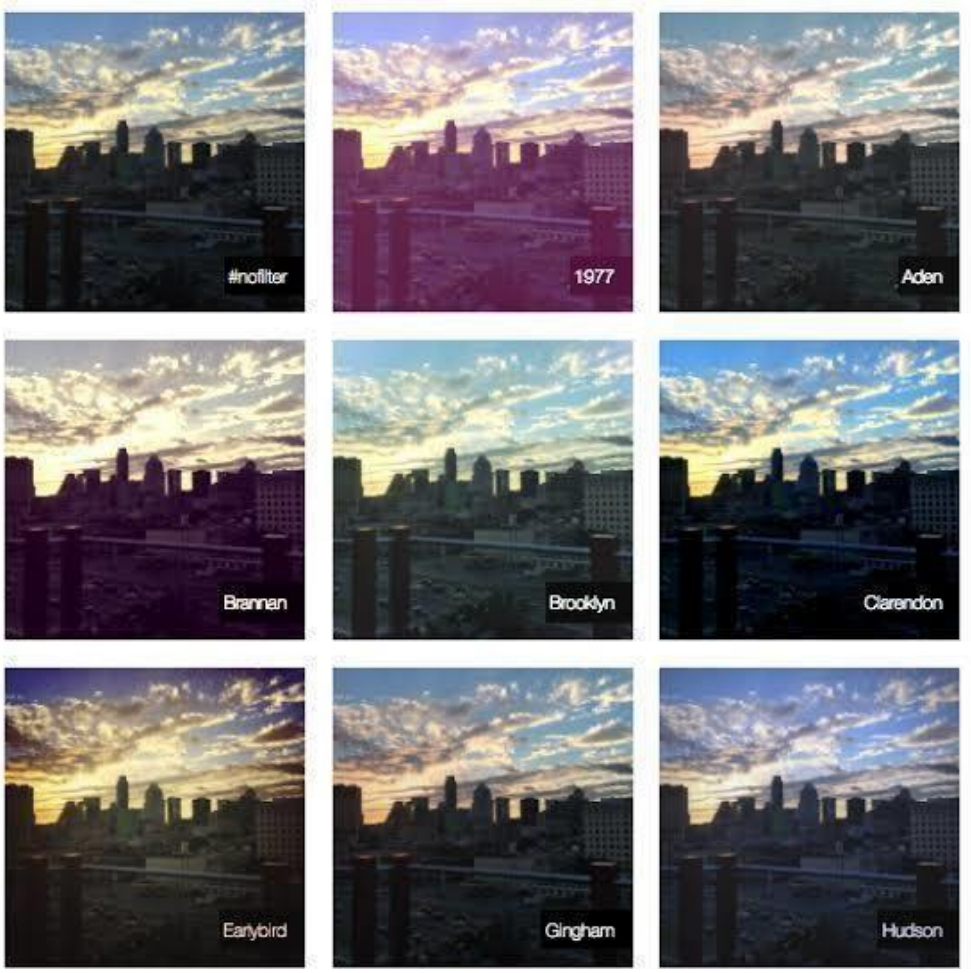
The blue patch is a flat area and difficult to find and track. Wherever you move the blue patch it looks the same. The black patch has an edge. Moved along the edge (parallel to edge), it looks the same. The red patch is a corner. Wherever you move the patch, it looks different, therefore it is unique. Hence, corners are considered to be good features in an image.

5.5 Convolution

We have learnt that computers store images in numbers and that pixels are arranged in a particular manner to create the picture we can recognize. These pixels have values varying from Oto 255 and the value of the pixel determines the color of that pixel.

But what if we edit these numbers, will it bring a change to the image? The answer is yes. As we change the values of these pixels, the image changes. This process of changing pixel values is the base of image editing.

We all use a lot of images editing software like photoshop and at the same time use apps like Instagram and Snapchat, which apply filters to the image to enhance the quality of that image.



As you can see, different filters applied to an image change the pixel values evenly throughout the image. How does this happen? This is done with the help of the process of convolution and the convolution operator which is commonly used to create these effects.

Before we understand how the convolution operation works, let us try and create a theory for the convolution operator by experiencing it using an online application.

Task

Go to the link https://setosa.io/ev/image-kernels/and scroll down to check the convolution matrix applied on the image.

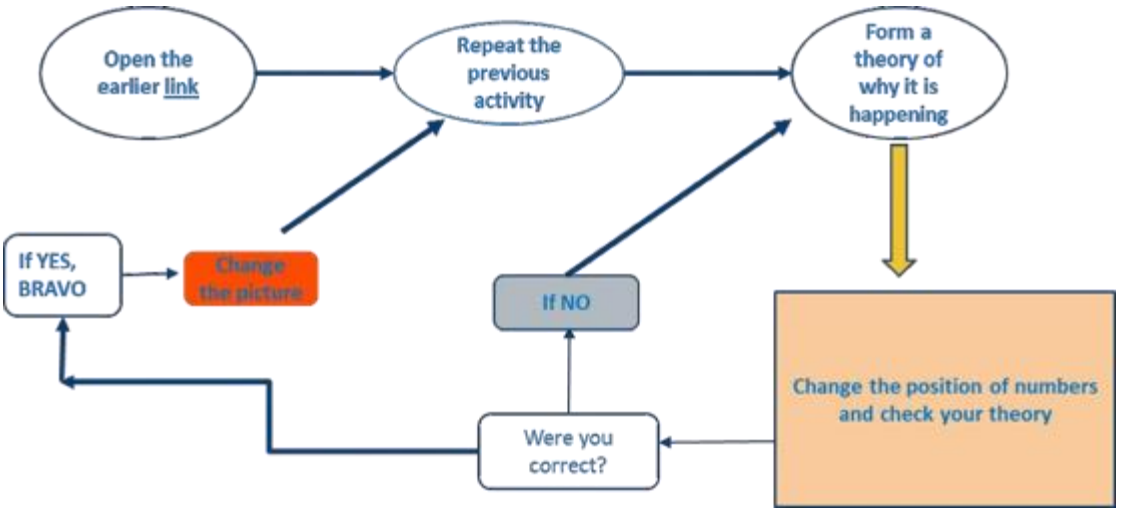
Try changing the values in the convolution operator and observe the modifications in the outputimage. Try these steps

5.6.1 Change all to positive values

5.6.2 Change all to negative values

5.6.3 Have a mixture of negative and positive values

Let us follow the following steps to understand how a convolution operator works. The steps to befollowed are:



Try experimenting with the following values to come up with a theory:

5.6.3.1 Make 4 numbers negative. Keep the rest as 0.5.6.3.2Now make one of them positive.

5.6.3.3 Observe what happens.

5.6.3.4 Now make the second positive.

What theory do you propose for convolution based on the observation?

_____

_____

It is time to test the theory. Change the location of the four numbers and follow the above mentioned steps. Does your theory hold true?

If yes, change the picture and try whether the theory holds true or not. If it does not hold true, modify your theory and keep trying until it satisfies all the conditions.

Let's Discuss

What effect did you apply?

How did different kernels affect the image?

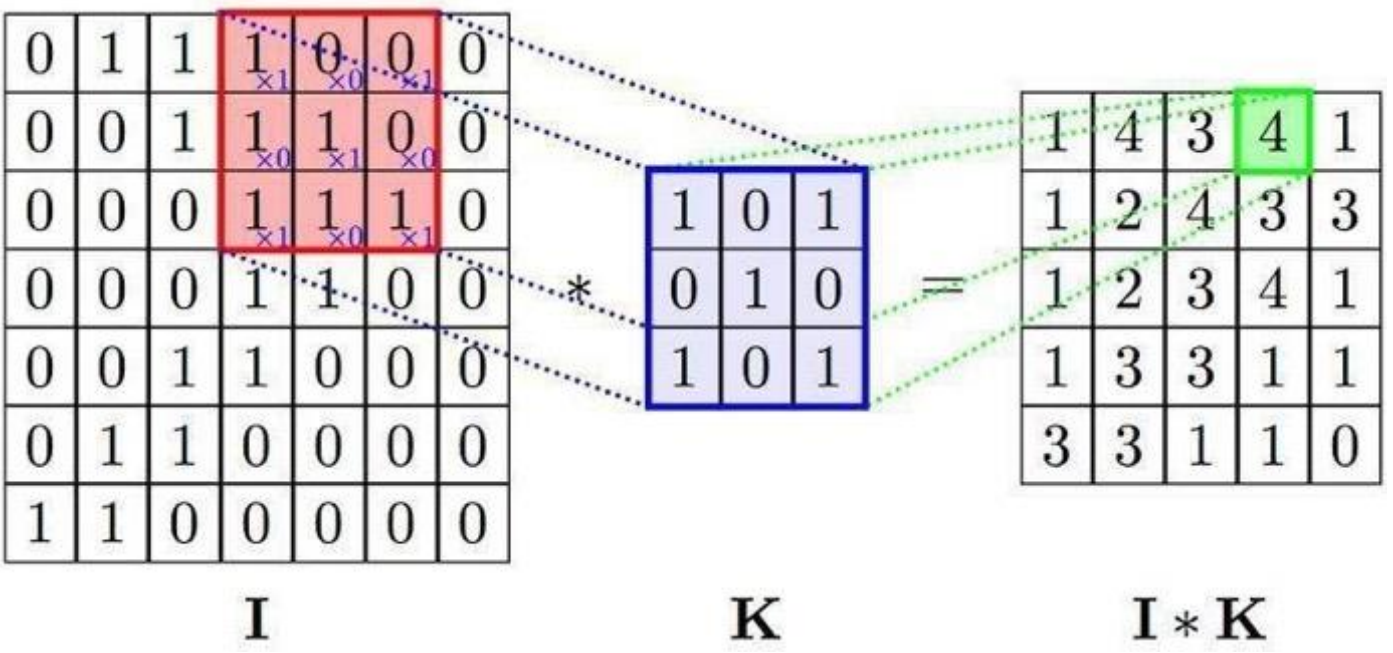
Why do you think we apply these effects?

How do you think the convolution operator works?

Convolution: Explained

Convolution is a simple mathematical operation that is fundamental to many common image processing operators. Convolution provides a way of multiplying together two arrays of numbers, generally of different sizes, but of the same dimensionality, to produce a third array of numbers of the same dimensionality.

An (image) convolution is simply an element-wise multiplication of image arrays and another array called the kernel followed by sum.



As you can see here,

I= Image Array

K = Kernel Array

I * K = Resulting array after performing the convolution operator

Note: The Kernel is passed over the whole image to get the resulting array after convolution.

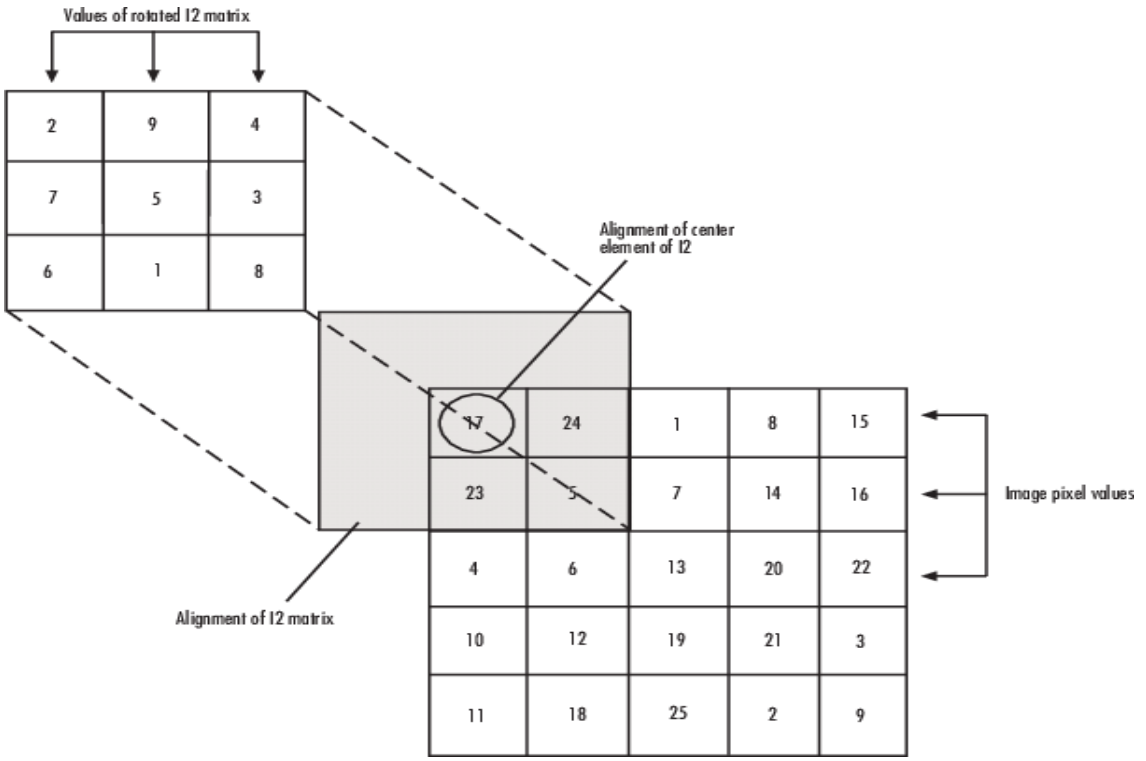
What is a Kernel?

A Kernel is a matrix, which is slid across the image and multiplied with the input such that the outputis enhanced in a certain desirable manner. Each kernel has a different value for different kinds of effectsthat we want to apply to an image.

In Image processing, we use the convolution operation to extract the features from the images which can le later used for further processing especially in Convolution Neural Network (CNN), which we will study later in the chapter.

In this process, we overlap the centre of the image with the centre of the kernel to obtain the convolution output. In the process of doing it, the output image becomes smaller as the overlapping is done at the edge row and column of the image. What if we want the output image to be of the exact size of the input image, how can we achieve this?

To achieve this, we need to extend the edge values out by one in the original image while overlapping the centres and performing the convolution. This will help us keep the input and output image of the same size. While extending the edges, the pixel values are considered zero.



Let's try

In this section we will try performing the convolution operator on paper to understand how it works.
Fill the blank places of the output images by performing the convolution operation.

150	0	255	240	190	25	89	255
100	179	25	0	200	255	67	100
155	146	13	20	0	12	45	0
100	175	0	25	25	15	0	0
120	156	255	0	78	56	23	0
115	113	25	90	0	80	56	155
135	190	115	116	178	0	145	165
123	255	255	0	255	255	255	0

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

Write Your Output Here:

Summary

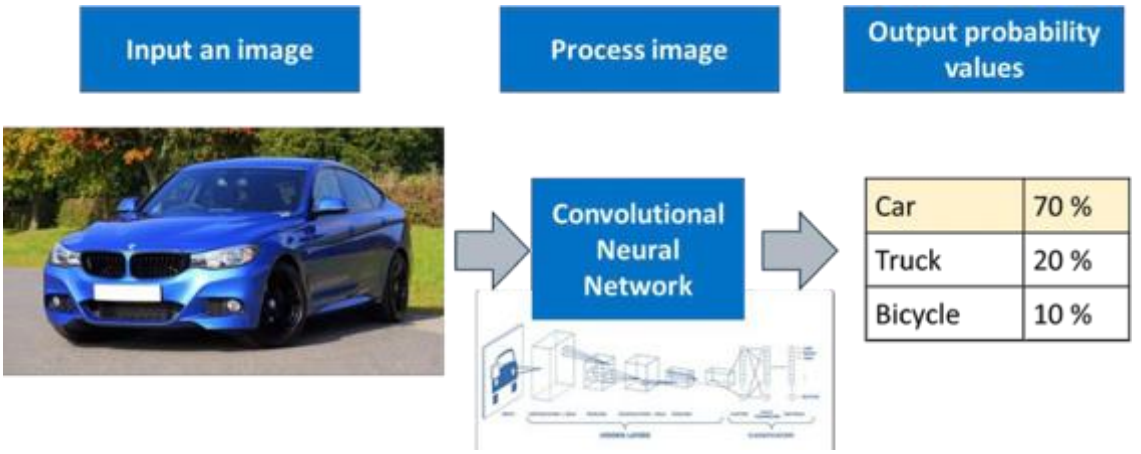
1. Convolution is a common tool used for image editing.
2. It is an element-wise multiplication of an image and a kernel to get the desired output.
3. In computer vision applications, it is used in Convolutional Neural Network (CNN) to extractimage features.

5.6 Convolution Neural Networks (CNN)

Introduction

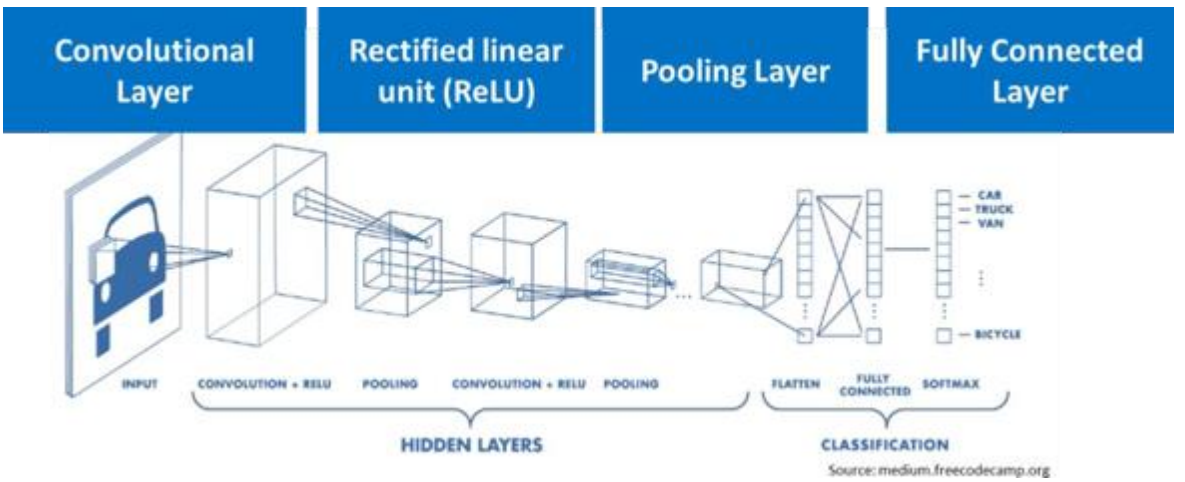
A **Convolutional Neural Network (CNN)** is a Deep Learning algorithm that can take in an input image,assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other.

The process of deploying a CNN is as follows:



In the above diagram, we give an input image, which is then processed through a CNN and then gives prediction based on the label given in the particular dataset.

The different layers of a Convolutional Neural Network (CNN) are as follows:



A convolutional neural network consists of the following layers:

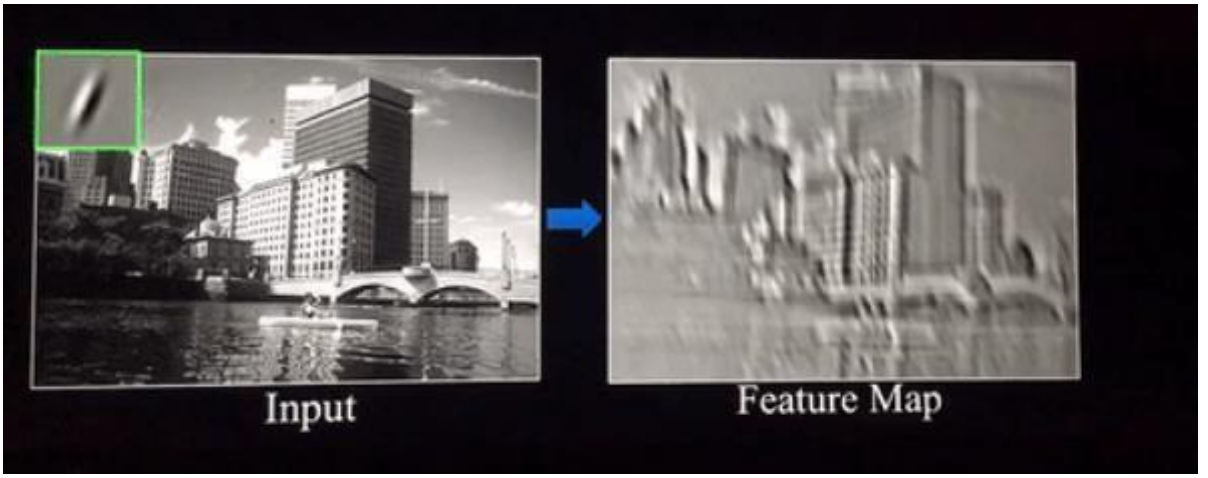
5.7.1 Convolution Layer
5.7.2 Rectified linear unit (ReLU)
5.7.3 Pooling Layer
5.7.4 Fully Connected Layer

5.7.1 Convolution Layer

It is the first layer of a CNN. The objective of the Convolution Operation is to extract the **high-level features** such as edges, from the input image. CNN need not be limited to only one Convolutional Layer. Conventionally, the first Convolution Layer is responsible for capturing the Low-Level features such as edges, colour, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network that has a wholesome understanding of images inthe dataset.

It uses convolution operation on the images. In the convolution layer, several kernels are used to produce several features. The output of this layer is called the feature map. A feature map is also called an activation map. We can use these terms interchangeably.
There are several uses we derive from the feature map:

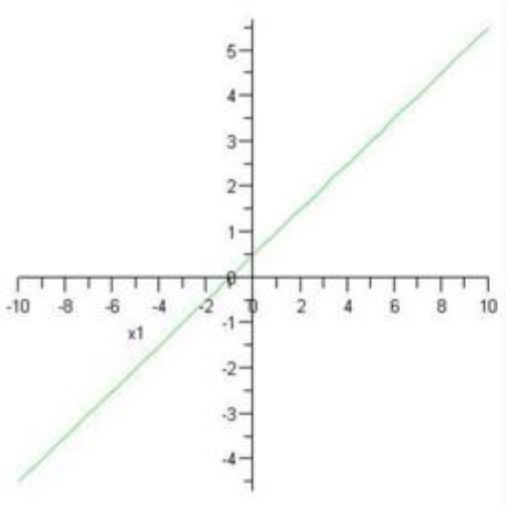
- We **reduce the image size** so that it can be processed more efficiently.
- We only focus on the features of the image that can help us in processing the image further. For example, you might only need to recognize someone's eyes, nose, and mouth to recognize the person. You might not need to see the whole face.

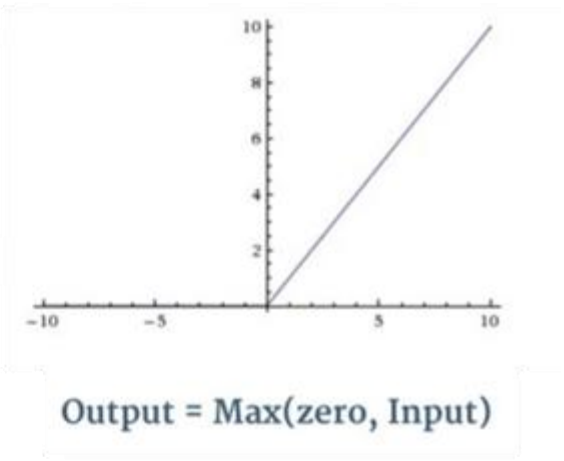


5.7.2 Rectified Linear Unit Function

The next layer in the Convolution Neural Network is the Rectified Linear Unit function or the ReLU layer. After we get the feature map, it is then passed onto the ReLU layer. This layer simply gets rid of all the negative numbers in the feature map and lets the positive number stay as it is.

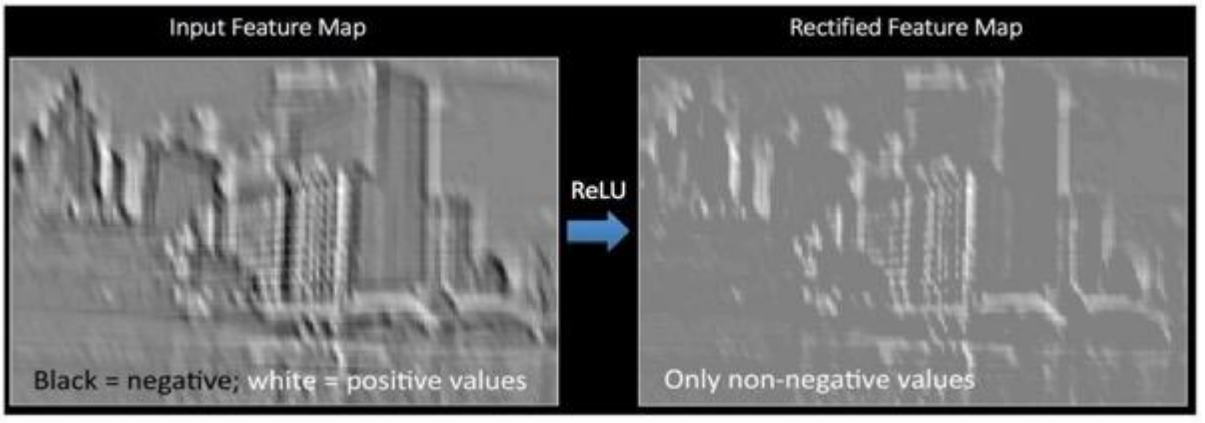
The process of passing it to the ReLU layer introduces non - linearity in the feature map. Let us see it through a graph.





If we see the two graphs side by side, the one on the left is a linear graph. This graph when passed through the ReLU layer gives the one on the right. The ReLU graph starts with a horizontal straight line and then increases linearly as it reaches a positive number.

Now the question arises, why do we pass the feature map to the ReLU layer? It is to make the colour change more obvious and more abrupt?



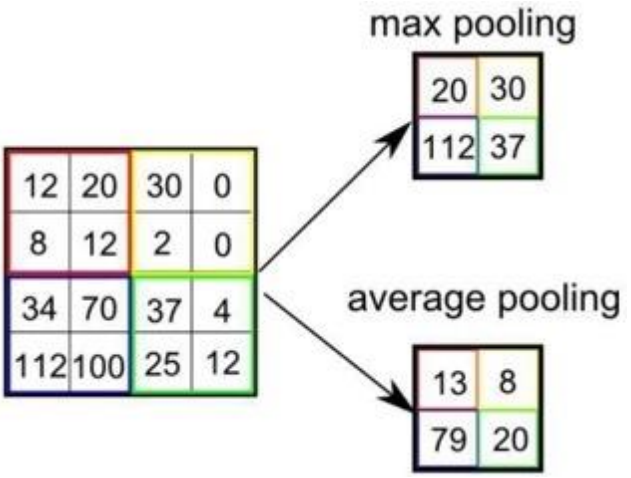
As shown in the above-convolved image, there is a smooth grey gradient change from black to white. After applying the ReLu function, we can see a more abrupt color change which makes the edges more obvious and acts as a better feature for the further layers in a CNN as it enhances the activation layer.

5.7.3 Pooling Layer

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature while still retaining the important features.

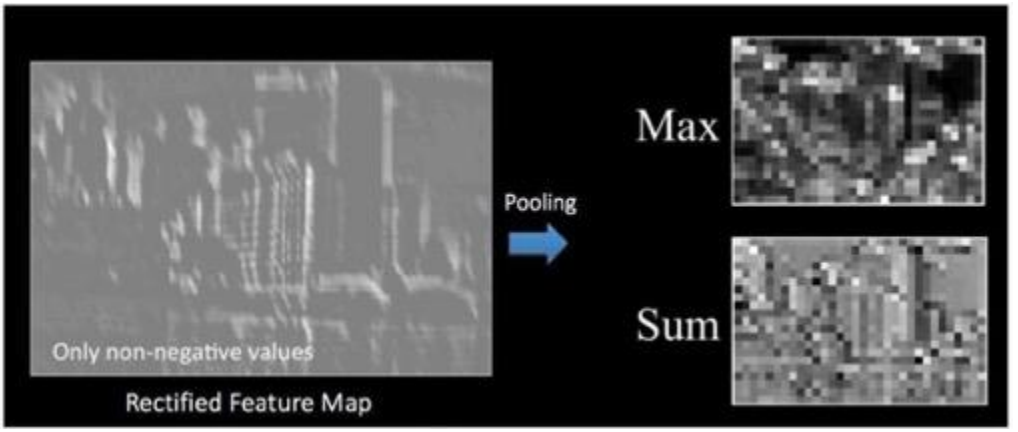
Two types of pooling can be performed on an image.

1. Max Pooling: Max Pooling returns the maximum value from the portion of the image covered by the Kernel.
2. Average Pooling: Max Pooling returns the maximum value from the portion of the imagecovered by the Kernel.



The pooling layer is important in the CNN as it performs a series of tasks which are as follows:

1. Makes the image smaller and more manageable
2. Makes the image more resistant to small transformations, distortions, and translations in theinput image.

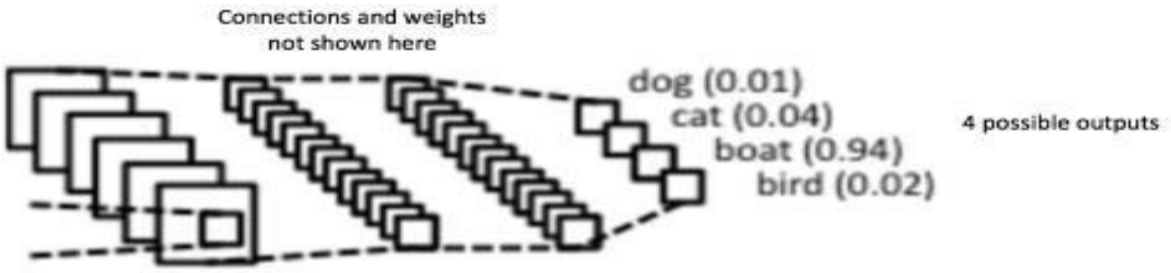


A small difference in the input image will create a very similar pooled image.

5.7.4 Fully Connected Layer

The final layer in the CNN is the Fully Connected Layer (FC layer). The objective of a fully connected layer is to take the results of the convolution/pooling process and use them to classify the image into a label (in a simple classification example).

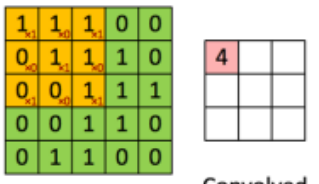
The output of convolution/pooling is flattened into a single vector of values, each representing a probability that a certain feature belongs to a label. For example, if the image is of a cat, features representing things like whiskers or fur should have high probabilities for the label "cat".

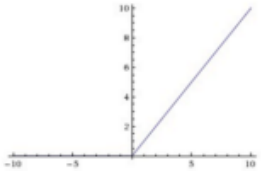


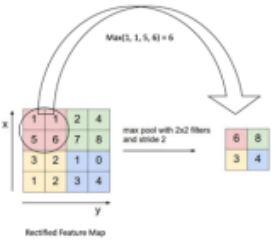
Let's Summarize:

Convolutional Neural Network

Convolutional Layer	Rectified linear unit (ReLU)	Pooling Layer	Fully Connected Layer







Car	70 %
Truck	20 %
Bicycle	10 %

Reduce size, improve feature, give probability value

Write the whole process of how a CNN works on the basis of the above diagram.

Test Yourself:

1. What is the primary objective of the Convolution Layer in a Convolutional Neural Network (CNN)?
   A) To flatten the input image
   B) To assign importance to various aspects/objects in the image
   C) To reduce the spatial size of the input image
   D) To perform element-wise multiplication of image arrays

2. Which of the following tasks is an example of computer vision?
   A) Rescaling an image
   B) Correcting brightness levels in an image
   C) Object detection in images or videos
   D) Changing tones of an image

3. How is resolution typically expressed?
   A) By the number of pixels along the width and height, such as 1280x1024
   B) By the brightness level of each pixel, ranging from 0 to 255
   C) By the total number of pixels, such as 5 megapixels
   D) By the arrangement of pixels in a 2-dimensional grid

4. What is the core task of image classification?
   A) Identifying objects and their locations in images
   B) Segmenting objects into individual pixels
   C) Assigning an input image one label from a fixed set of categories
   D) Detecting instances of real-world objects in images

5. What is the function of the Rectified Linear Unit (ReLU) layer in a CNN?
   A) To reduce the image size for more efficient processing
   B) To assign importance to various aspects/objects in the input image
   C) To get rid of negative numbers in the feature map and retain positive numbers
   D) To perform the convolution operation on the input image

6. Object detection and handwriting recognition are examples of tasks commonly associated with:
   A) Computer vision
   B) Image processing
   C) Both computer vision and image processing
   D) Neither computer vision nor image processing

7. What does the pixel value represent in an image?
   A) Width of the pixel
   B) Brightness or color of the pixel
   C) Height of the pixel
   D) Resolution of the pixel

8. In the byte image format, what is the range of possible pixel values?
   A) 0 to 10
   B) 0 to 100
   C) 0 to 1000
   D) 0 to 255

9. In a grayscale image, what does the darkest shade represent?
   A) Total presence of color
   B) Zero value of pixel
   C) Lightest shade of gray
   D) Maximum pixel value

10. In an RGB image, what does a pixel with an intensity value of 0 represent?
    A) Full presence of color
    B) No presence of color
    C) Maximum brightness level
    D) Minimum brightness level

11. **Assertion:** Object detection is a more complex task than image classification because it involvesidentifying both the presence and location of objects in an image.

    Reasoning: Object detection algorithms need to not only classify the objects present in an imagebut also accurately localize them by determining their spatial extent.

    Select the appropriate option for the statements given above:
    A) Both A and R are true and R is the correct explanation of A
    B) Both A and R are true and R is not the correct explanation of A
    C) A is true but R is false
    D) A is False but R is true

12. **Assertion:** Grayscale images consist of shades of gray ranging from black to white, where each pixel is represented by a single byte, and the size of the image is determined by its height multipliedby its width.

    Reasoning: Grayscale images are represented using a three intensities per pixel, typically rangingfrom 0 to 255.

    Select the appropriate option for the statements given above:
    A) Both A and R are true and R is the correct explanation of A
    B) Both A and R are true and R is not the correct explanation of A
    C) A is true but R is false
    D) A is False but R is true

Reflection Time:

1. Imagine you have a smartphone camera app that can recognize objects. When you point your camera at a dog, the app identifies it as a dog, analyzing patterns and features in the image. Behind the scenes, the app's software processes the image, detecting edges, shapes, and colors, then compares these features to a vast database to make accurate identifications."
   Identify the technology used in the above scenario and explain the way it works.

2. Enlist two smartphone apps that utilize computer vision technology? How have these apps improved your efficiency or convenience in daily tasks?

3. How an RGB image is different from a grayscale image?

4. Determine the color of a pixel based on its RGB values mentioned below: (i) R=0, B=0, G=0
   (ii) R=255, B=255, G=255
   (iii) R=0, B=0, G=255 (iv) R=0, B=255, G=0

5. Briefly describe the purpose of the convolution operator in image processing.

6. What are the different layers in Convolutional Neural Network? What features are likely to be detected by the initial layers of a neural network and how is it different from what is detected by the later layers?

7. "Imagine you're a researcher tasked with improving workplace safety in a manufacturing environment. You decide to employ computer vision technology to enhance safety measures."

8. How would you utilize computer vision in two distinct applications to promote safety within the manufacturing plant, ensuring both the physical well-being of employees and the efficiency of operations?
   Provide detailed explanations for each application, including the specific computer vision techniques or algorithms you would employ, and how they would contribute to achieving your safety goals.

9. Explain the distinctions between image classification, classification with localization, object detection, and instance segmentation in computer vision tasks. Provide examples for each to support your answer.

10. “Agriculture is an industry where precision and efficiency are crucial for sustainable production. Traditional farming methods often rely on manual labor and visual inspection, which can be time- consuming and error-prone. However, advancements in computer vision technology offer promising solutions to optimize various agricultural processes.

Agricultural drones equipped with high-resolution cameras and computer vision algorithms are increasingly being used to monitor crop health, detect diseases, and assess crop yields.”

Answer the following questions based on the case study mentioned above:
How does the integration of computer vision technology with drones improve efficiency in agricultural practices compared to traditional methods?
What are some key indicators or parameters that computer vision algorithms can analyze to assess crop health and detect diseases?

11. You are tasked with developing a computer vision system for a self-driving car company. The system needs to accurately detect and classify various objects on the road to ensure safe navigation. Imagine you're working on improving the object detection algorithm for the self-driving car's computer vision system. During testing, you notice that the system occasionally misclassifies pedestrians as cyclists, especially in low-light conditions.
How would you approach addressing this issue? What steps would you take to enhance the accuracy of pedestrian detection while ensuring the system's overall performance and reliability on the road?

Reflection Time:

1. What is the primary objective of the Convolution Layer in a Convolutional Neural Network (CNN)?

A) To flatten the input image

B) To assign importance to various aspects/objects in the image

C) To reduce the spatial size of the input image

D) To perform element-wise multiplication of image arrays

2. Which of the following tasks is an example of computer vision?

A) Rescaling an image

B) Correcting brightness levels in an image

C) Object detection in images or videos

D) Changing tones of an image

3. How is resolution typically expressed?

A) By the number of pixels along the width and height, such as 1280x1024

B) By the brightness level of each pixel, ranging from 0 to 255

C) By the total number of pixels, such as 5 megapixels

D) By the arrangement of pixels in a 2-dimensional grid

4. What is the core task of image classification?

A) Identifying objects and their locations in images

B) Segmenting objects into individual pixels

C) Assigning an input image one label from a fixed set of categories

D) Detecting instances of real-world objects in images

5. What is the function of the Rectified Linear Unit (ReLU) layer in a CNN?

A) To reduce the image size for more efficient processing

B) To assign importance to various aspects/objects in the input image

C) To get rid of negative numbers in the feature map and retain positive numbers

D) To perform the convolution operation on the input image

6. Object detection and handwriting recognition are examples of tasks commonly associated with:

A) Computer vision

B) Image processing

C) Both computer vision and image processing

D) Neither computer vision nor image processing

7. What does the pixel value represent in an image?

A) Width of the pixel

B) Brightness or color of the pixel

C) Height of the pixel

D) Resolution of the pixel

8. In the byte image format, what is the range of possible pixel values?

A) 0 to 10

B) 0 to 100

C) 0 to 1000

D) 0 to 255

9. In a grayscale image, what does the darkest shade represent?

A) Total presence of color

B) Zero value of pixel

C) Lightest shade of gray

D) Maximum pixel value

10. In an RGB image, what does a pixel with an intensity value of 0 represent?

A) Full presence of color

B) No presence of color

C) Maximum brightness level

D) Minimum brightness level

11. Assertion: Object detection is a more complex task than image classification because it involves identifying both the presence and location of objects in an image.

Reasoning: Object detection algorithms need to not only classify the objects present in an image but also accurately localize them by determining their spatial extent.

Select the appropriate option for the statements given above:

A) Both A and R are true and R is the correct explanation of A

B) Both A and R are true and R is not the correct explanation of A

C) A is true but R is false

D) A is False but R is true

12. Assertion: Grayscale images consist of shades of gray ranging from black to white, where

each pixel is represented by a single byte, and the size of the image is determined by its height multiplied by its width.

Reasoning: Grayscale images are represented using a three intensities per pixel, typically ranging from 0 to 255.

Select the appropriate option for the statements given above:

A) Both A and R are true and R is the correct explanation of A

B) Both A and R are true and R is not the correct explanation of A

C) A is true but R is false

D) A is False but R is true

Answers of MCQs

1. B) To assign importance to various aspects/objects in the image
2. **C) Object detection in images or videos**
3. A) By the number of pixels along the width and height, such as 1280x1024
4. C) Assigning an input image one label from a fixed set of categories
5. C) To get rid of negative numbers in the feature map and retain positive numbers
6. A) Computer vision
7. B) Brightness or colour of the pixel
8. D) 0 to 255
9. B) Zero value of pixel
10. B) No presence of colour
11. A) Both A and R are true and R is the correct explanation of A
12. C) A is true but R is false

Reflection Time:

1. Imagine you have a smartphone camera app that can recognize objects. When you point your camera at a dog, the app identifies it as a dog, analyzing patterns and features in the image. Behind the scenes, the app's software processes the image, detecting edges, shapes, and colors, then compares these features to a vast database to make accurate identifications." Identify the technology used in the above scenario and explain the way it works.

   Ans: Technology Used: Computer Vision using AI (specifically Convolutional Neural Networks - CNNs)

   Explanation: The app uses computer vision to understand what the camera sees. It processes the image using a CNN model that:

   Detects edges, shapes, colors (feature extraction)
   Compares features with a trained model (from a large image database)
   Classifies the object (like identifying a "dog")

2. Enlist two smartphone apps that utilize computer vision technology? How have these apps improved your efficiency or convenience in daily tasks?

Ans: Here are two smartphone apps that use computer vision technology and how they help in daily life:

1. Google Lens
   - Use: Identifies objects, text, landmarks, plants, and translates text from images.
   - Efficiency: Saves time by quickly finding information without typing or manually searching.
2. Face Unlock (built-in in many smartphones)
   - Use: Unlocks your phone by recognizing your face.
   - Convenience: Eliminates the need to enter a PIN or pattern every time, making access faster and hands-free.

3. How an RGB image is different from a grayscale image?

Ans: Refer Text book page no 157

4. Determine the color of a pixel based on its RGB values mentioned below:

(i)	R=0, B=0, G=0	Black Colour
(ii)	R=255, B=255, G=255	White Colour
(iii)	R=0, B=0, G=255	Green Colour
(iv)	R=0, B=255, G=0	Blue Colour

5. Briefly describe the purpose of the convolution operator in image processing.

Ans: The convolution operator in image processing helps detect patterns like edges, textures, and shapes by applying a filter (kernel) to an image. It slides over the image, performing mathematical operations to highlight important features. This process is key in image recognition, sharpening, and blurring in Computer Vision and CNNs.

6. What are the different layers in Convolutional Neural Network? What features are likely to be detected by the initial layers of a neural network and how is it different from what is detected by the later layers?

Ans: A Convolutional Neural Network (CNN) has multiple layers that process images step by step. The initial layers detect basic features like edges, lines, and colours. As the image moves deeper into the network, the later layers recognize complex patterns, such as shapes, objects, and faces. The Convolutional Layer extracts features, the Pooling Layer reduces data size, and the Fully Connected Layer makes final decisions. Early layers focus on small details, while deeper layers understand the whole object.

7. “Imagine you're a researcher tasked with improving workplace safety in a manufacturing environment. You decide to employ computer vision technology to enhance safety measures.”

   How would you utilize computer vision in two distinct applications to promote safety within the manufacturing plant, ensuring both the physical well-being of employees and the efficiency of operations? Provide detailed explanations for each application, including the specific computer vision techniques or algorithms you would employ, and how they would contribute to achieving your safety goals.

Ans: I would use Computer Vision to enhance workplace safety by implementing real-time monitoring for detecting hazards like unsafe worker behaviour, improper PPE usage, and machine malfunctions. AI-powered cameras can identify risks, send alerts, and automate compliance checks, reducing accidents and ensuring a safer manufacturing environment.

8. Explain the distinctions between image classification, classification with localization, object detection, and instance segmentation in computer vision tasks. Provide examples for each to support your answer.

Ans: 1. Image Classification

- What it does: Predicts the category of the main object or scene in an image.
- No location information is provided—only the label.
- Example: An image is labelled as "cat" without saying *where* in the image the cat is.
- Use case: Classifying handwritten digits (MNIST dataset).

2. Classification with Localization

- What it does: Predicts the category of the object and provides its bounding box coordinates.
- Only one main object is usually considered.
- Example: An image of a dog is classified as "dog" with a bounding box drawn around it.
- Use case: Detecting a product in an image and highlighting its position.

3. Object Detection

- What it does: Identifies multiple objects in an image, predicts their categories, and provides bounding boxes for each.
- Example: In a street photo, detecting and locating "cars," "traffic lights," and "pedestrians" all at once.
- Use case: Self-driving cars detecting road elements.

4. Instance Segmentation

- What it does: Detects and classifies each object instance and provides a pixel-level mask outlining its exact shape.
- Example: In a crowd photo, segmenting each person separately with their precise boundaries, not just a rectangle.
- Use case: Medical imaging to segment tumours from scans.

9. "Agriculture is an industry where precision and efficiency are crucial for sustainable production. Traditional farming methods often rely on manual labour and visual inspection, which can be time- consuming and error-prone. However, advancements in computer vision technology offer promising solutions to optimize various agricultural processes. Agricultural drones equipped with high-resolution cameras and computer vision algorithms are increasingly being used to monitor crop health, detect diseases, and assess crop yields."

   Answer the following questions based on the case study mentioned above:

How does the integration of computer vision technology with drones improve efficiency in agricultural practices compared to traditional methods?

What are some key indicators or parameters that computer vision algorithms can analyse to assess crop health and detect diseases?

a. Improved Efficiency with Computer Vision Drones: Computer Vision-equipped drones improve efficiency by automating crop monitoring, reducing the need for manual inspections. They cover large fields quickly, providing real-time data on crop health and detecting issues early, leading to faster decision-making and improved yields.

b. Key Indicators for Crop Health & Disease Detection : Computer Vision algorithms analyse factors like leaf colour changes, plant height, wilting, pest infestations, and nutrient deficiencies. They also use NDVI (Normalised Difference Vegetation Index) and thermal imaging to assess plant stress and disease outbreaks accurately.

10. You are tasked with developing a computer vision system for a self-driving car company. The system needs to accurately detect and classify various objects on the road to ensure safe navigation. Imagine you're working on improving the object detection algorithm for the self-driving car's computer vision system. During testing, you notice that the system occasionally misclassifies pedestrians as cyclists, especially in low-light conditions. How would you approach addressing this issue? What steps would you take to enhance the accuracy of pedestrian detection while ensuring the system's overall performance and reliability on the road?

Ans: I would improve pedestrian detection by:

1. **Collecting more diverse training data** — especially low-light images of pedestrians and cyclists.
2. **Using data augmentation** — adjust brightness, contrast, and shadows in training images.
3. **Enhancing the model** — fine-tune or use a better architecture for low-light performance.
4. **Adding extra sensors** — like infrared or LiDAR for better night detection.

This makes the system more accurate and reliable in all conditions.