Computer Vision

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 in order to understand and analyse actual phenomena with data.

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

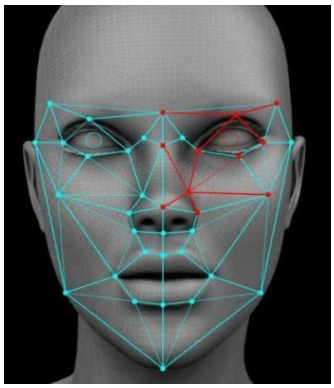
But can machines do the same? Can machines have the eyes that humans have? If you answered Yes, then you are absolutely right. The Computer Vision domain of Artificial Intelligence, enables machines to see through images or visual data, process and analyse them on the basis of algorithms and methods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

Applications of Computer Vision

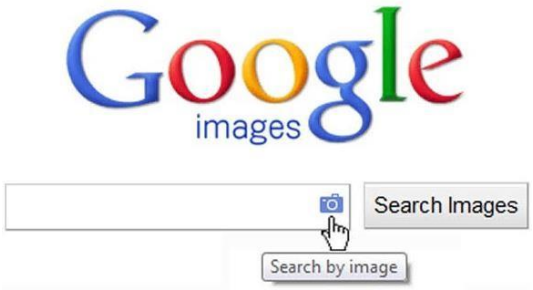
The concept of computer vision was first introduced in the 1970s. All these new applications of computer vision excited everyone. Having said that, the computer vision technology advanced enough to make these applications available to everyone at ease today. However, in recent years the world witnessed 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 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**: *The 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.*

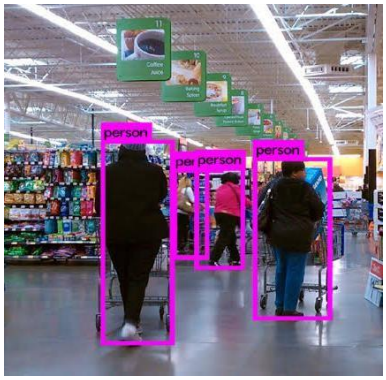


Google's Search by Image*: The maximum amount of searching for data on Google's search engine comes from textual data, but at the same time it has an interesting feature of getting search results through an image. This uses Computer Vision as it compares different features of the input image to the database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

* Images shown here

Computer Vision in Retail*: The retail field has been one of the fastest growing field 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 developing 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, computersupported 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 2D scan images into interactive 3D models

* Images shown here are the property of individual organisations and are used here for reference purpose only.

that enable medical professionals to gain a detailed understanding of a patient's health condition.

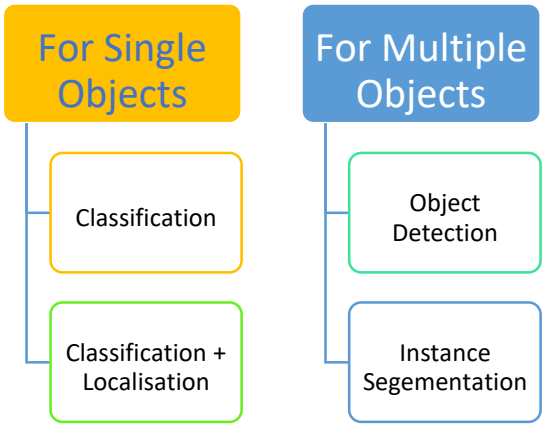
Google Translate App*: All you need to do to read signs in a foreign language is to point your phone's camera at the words and let the Google Translate app tell you what it means in your preferred language almost instantly. By using optical character recognition to see the image and augmented reality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

Computer Vision: Getting Started

Computer Vision is a domain of Artificial Intelligence, that deals with the images. It involves the concepts of image processing and machine learning models to build a Computer Vision based application.

Computer Vision Tasks

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



Classification

Image Classification problem is the task of **assigning an input image one label from a fixed set of categories**. 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 which 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

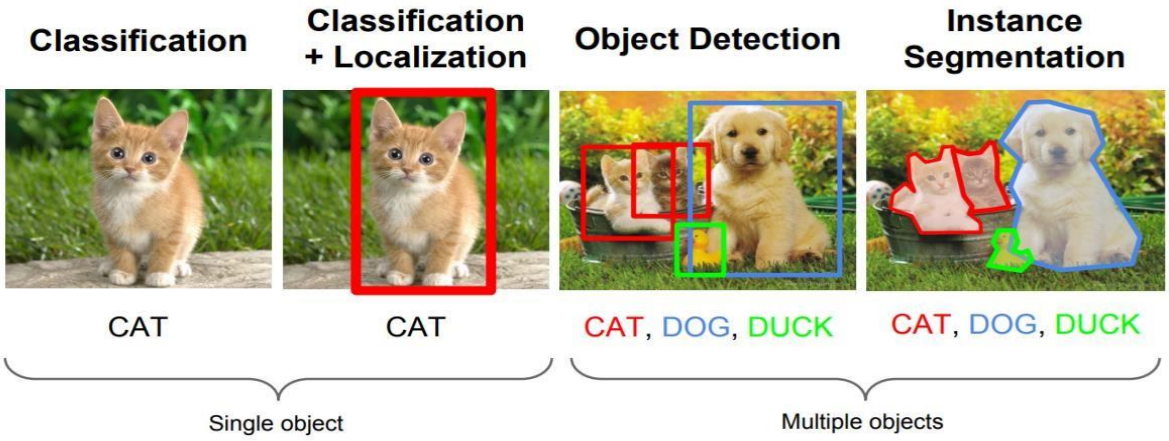
Object detection is the process of **finding instances of real-world objects such as faces, bicycles, and buildings in images or videos**. Object detection algorithms typically use extracted features and

* Images shown here are the property of individual organisations and are used here for reference purpose only.

learning algorithms to recognize instances of an object category. It is commonly used in applications such as image retrieval and automated vehicle parking systems.

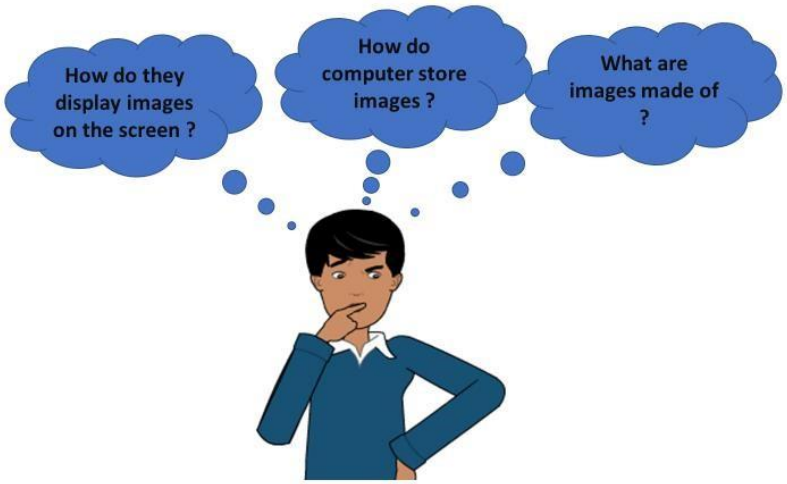
Instance Segmentation

Instance Segmentation is the process of detecting instances of the objects, giving them a category and then giving each pixel a label on the basis of 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 such a regular basis.

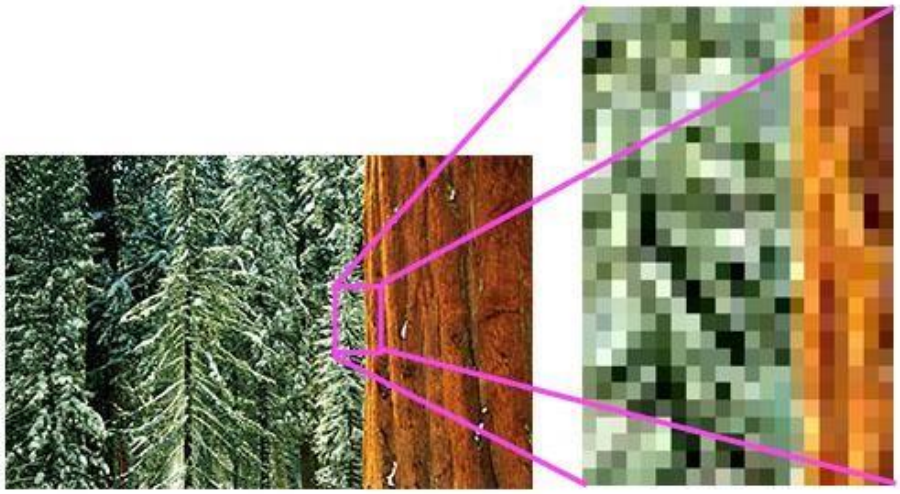


Don’t know the answer yet? Don’t worry, in this section we will study about the basics of an image:

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 individual 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 describe pixel count, one convention is to express resolution as the width by the height, for example a monitor resolution of 1280×1024. 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 1280×1024 monitors, it could also be expressed as 1280 x 1024 = 1,310,720, or 1.31 megapixels.

Pixel value

Each of the pixels that represents an image stored inside a computer has a *pixel value* which 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 0 to 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 the 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 tells us that the 8 bit can have 255 possibilities of values which 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

* Images shown here are the property of individual organisations and are used here for reference purpose only.

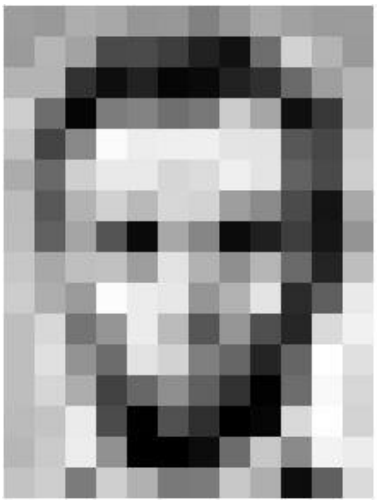
here

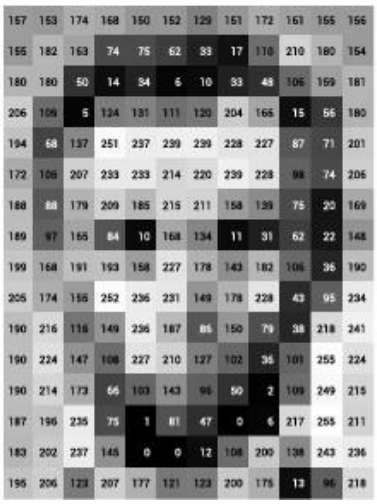
Grayscale Images

Grayscale images are images which 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 about grayscale images.



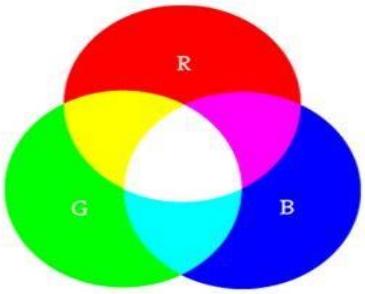


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	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 are within the range of 0255.The computers store the images we see in the form of these numbers.

RGB Images



All the images that we see around 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.

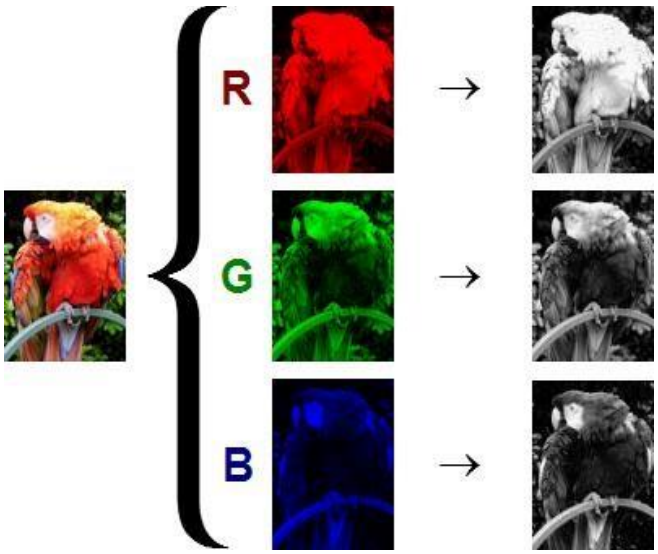
1. What is the output colour when you put R=G=B=255 ?
2. What is the output colour when you put R=G=B=0 ?
3. How does the colour vary when you put either of the three as 0 and then keep on varying the other two?
4. How does the output colour change when all the three colours are varied in same proportion?
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 a number of pixels with each pixel value varying from 0 to 255. All the three planes when combined together form a colour image. This means that in a 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 together 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 0 to 255 and as studied earlier, 0 is considered as black or no presence of colour and 255 means white or full presence of colour. These three individual RGB values when combined together form the colour of each pixel.

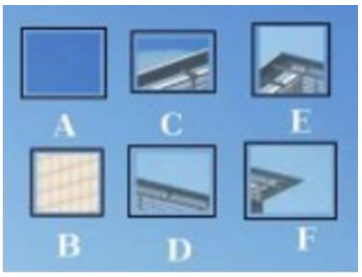
Therefore, each pixel in the RGB image has three values to form the complete colour.

Image Features

In computer vision and image processing, a **feature** is a piece of information which 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 small patches 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.



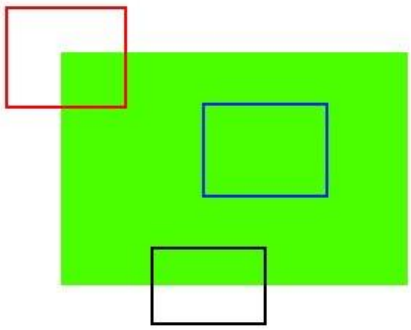
1. Were you able to find the exact location of all the patches?
2. Which one was the most difficult to find?
3. Which one was the easiest to find?

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 B are 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 is still difficult. This is because the pattern is the same everywhere along the edge.
For Patch E and F: The patches E and F are the easiest to find in the image. The reason being 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

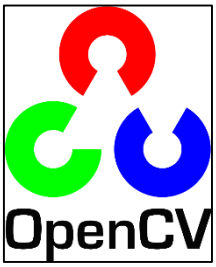
In image processing, we can get a lot of features from the image. It can be either a blob, an edge or a corner. These features help us to perform various tasks and then get the analysis done on the basis of the application. Now the question that arises is which of the following are good features to be 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 which 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.

Introduction to OpenCV

Now that we have learnt about image features and its importance in image processing, we will learn about a tool we can use to extract these features from our image for further processing.



OpenCV or Open Source Computer Vision Library is that tool which helps a computer extract these features from the images. It is used for all kinds of images and video processing and analysis. It is capable of processing images and videos to identify objects, faces, or even handwriting. In this chapter we will use OpenCV for basic image processing operations on images such as resizing, cropping and many more.

To install OpenCV library, open anaconda prompt and then write the following command:

```
pip install opencv-python
```

Now let us take a deep dive on the various functions of OpenCV to understand the various image processing techniques. Head to Jupyter Notebook for introduction to OpenCV given on this link: http://bit.ly/cv_notebook

ASSIGNMENT QUESTIONS

1. **What is the use of computer vision in AI?**
2. **What is Computer Vision?**
3. **Face lock in smart phone is feature of Computer Vision. Briefly Explain the feature.**
4. **Explain the tasks used in computer vision for single object.**
5. **What do you understand by GrayScale image?**
6. **Write three differences between Computer Vision (CV) and Human Vision System(HVS).**
7. **What is OpenCV Computer Vision Library?**
8. **What is Pixel? Give any two important features of a Pixel in digital Image.**