CV: Facial Recognition

Task-3

What do one mean by Pre-process?

To do preliminary processing of something when there are several steps required to prepare data (it is a set of raw values when processed gives some information) for the user

The goal of this step is to make our data ready for ML model to make it easier to analyse and process computationally

What is Pre-processing in Image processing ?

1) They are the steps taken to format and clean images before they are used by model training and inference
2) It includes resizing, orienting and colour corrections
3) It may **decrease model training time** (training phase-a developer feeds their model the dataset so that it can "learn" everything it needs to about the type it will analyze) and **increase model inference speed** (how long it takes for a forward propogation, the model makes predictions based on live data to produce actionable results)In order to feed a dataset of images to a convolutional network, they must all be of the same size

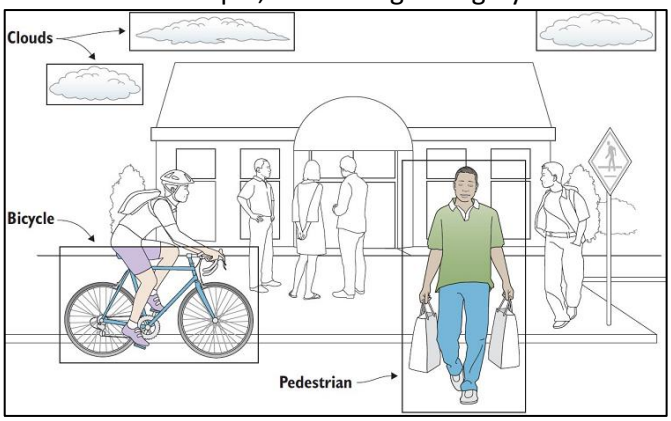
Image pre-processing steps are applied to training and test sets while image augmentation is only applied to the training data

DATASET:-

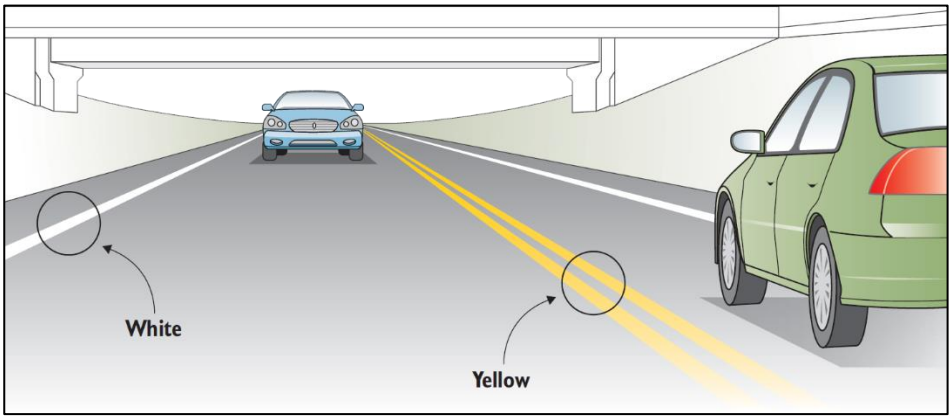
A) Steps Involved in data pre-processing of the given dataset:-

1) Convert colour images to grayscale to reduce computation complexity:-

- In certain problems it is useful to lose unnecessary information from the images to reduce space or computational complexity
- In many objects, colour isn't necessary to recognize and interpret an image
- Colour images contain more information than black and white images which can add unnecessary complexity and takes up more space in the memory
- Usually colour images are represented in three channels and converting it to grayscale reduces the number of pixels that need to be processed
- In the below example, converting it to grayscale will reduce the computational complexity



- There are a number of applications in which colours are important
- For example, the lane detection applications in self-driving cars. The car has to differentiate between the yellow and white lanes because they are treated differently. Converting it into images do not provide enough information.



- **If one is able to identity the object that one is looking for in a gray image then one probably has enough information to feed to their model. If not then one needs more information(colour images) about the images**

For the given data set this pre-processing technique is not required as the images are already in grayscale

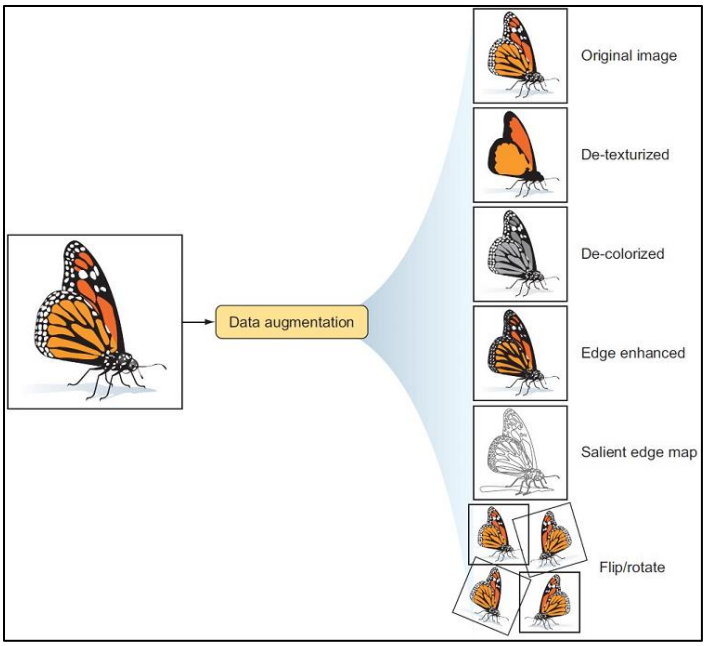
2) Standardize images:

Our images must be pre-processed and scaled to have identical widths and heights before fed to the learning algorithm

For the given data set this pre-processing technique is not required as the images are already standardized

3) Data Augmentation:

- They are manipulations applied to images to create different version of similar content in order to expose the model to a wider array of training examples
- Adjusting existing training data to generalize to other situations allows the model to learn from a wider array of situations
- It makes it more likely that your model recognizes objects when they appear in any form and shape
- This is particularly important when collected datasets may be small



For the given data set this pre-processing technique is required as the images, so that our model can recognize objects when they appear in any form

B) Techniques you would use for classifications:-

Single-label classification is the most common classification task in **supervised Image Classification**. A single label is present for each image in single-label classification. Therefore, the model outputs a single value or prediction for each image that it sees.

Given a set of images that are labelled with a single category, we are asked to predict these categories. There are variety of challenges associated with this task, including viewpoint variation, scale variation, image deformation, background clutter etc.

Researchers have come up with a data driven to approach to solve this. Instead of trying to specify what every one of the image categories look like directly in code, they provide the computer with many examples and learn about the visual appearance of each class. In other words, they first accumulate a training dataset of labelled images, then feed it to the computer in order for it to get familiar with the data .

The two types of classes are information classes and spectral classes.

Information classes:- are those categories of interest that the analyst is actually trying to identify in the imagery

Spectral classes:- are groups of similar pixels

The objective is to match the spectral classes in the data to the information classes of interest

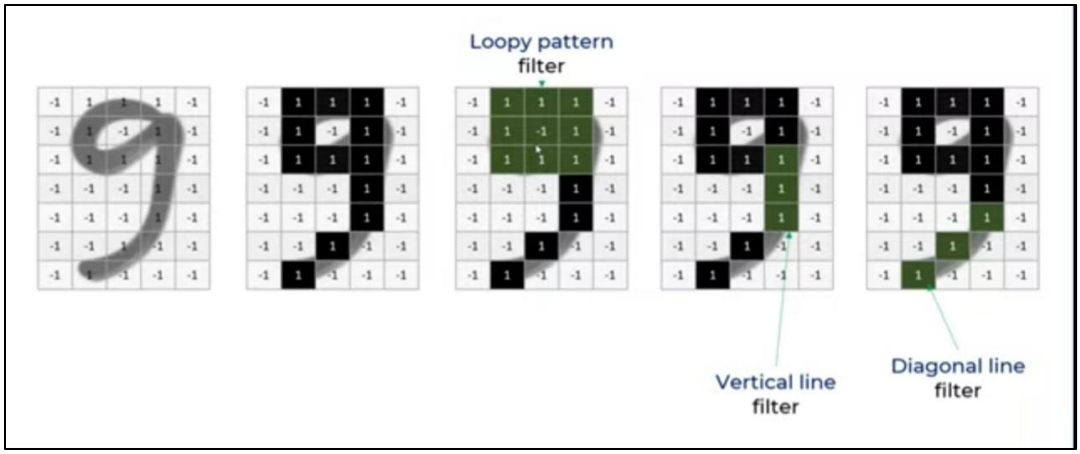
The complete image classification pipeline is:-

1) Our input is a training dataset that consists of N images, each labeled with one of K different classes
2) We use this training set to train a classifier to learn what every one of the classes looks like
3) We evaluate the quality of the classifier by asking it to predict labels for a new set of images that it has never seen before.
4) We will then compare the true labels of these images to the ones predicted by the classifier

Convolutional Neural Networks (CNNs)

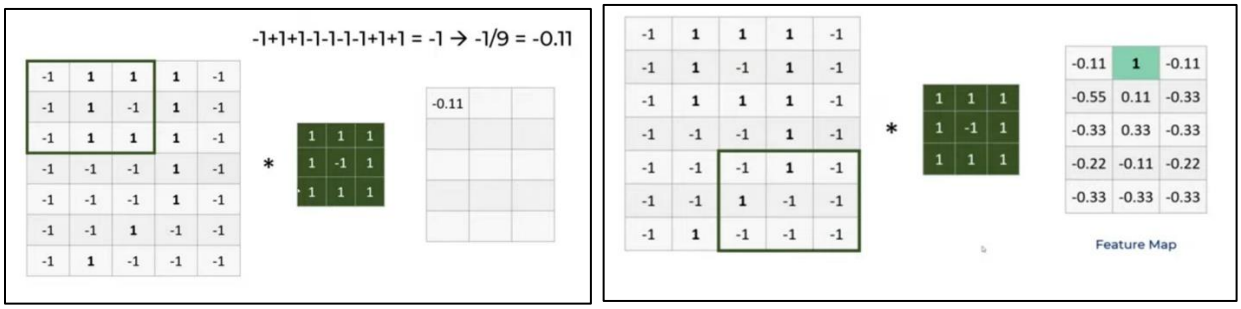
CNN has just enough weights to look at small patches of the image. It's like reading a book with a magnifying glass, one will read the entire page but it will read a small patch of the page at a time.

1) Image is transformed into a grid containing RGB values from 0 to 255.
2) We filter out small parts of the images we are used for testing. Filters are nothing but the features detectors



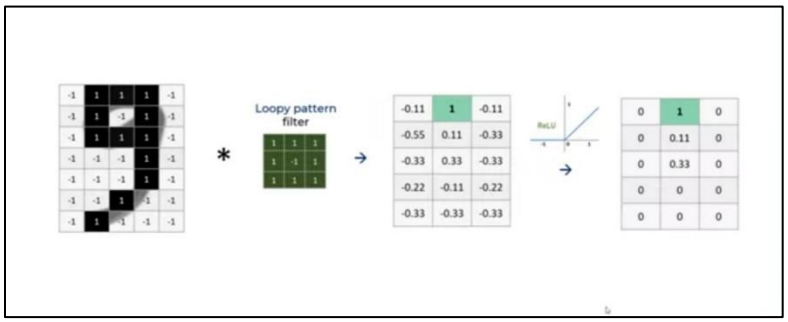
In the following example we divide the image into a grid of 1 and -1. The image is divided into 3 smaller parts or filters which is the loopy pattern filter, vertical line filter, diagonal line filter(may or may not be present).

3) Using these filters we formulate a feature map.
no of feature maps = no of filters
If the value is close to one or equal to one then the feature gets activated
they are stacked together to form a 3D volume



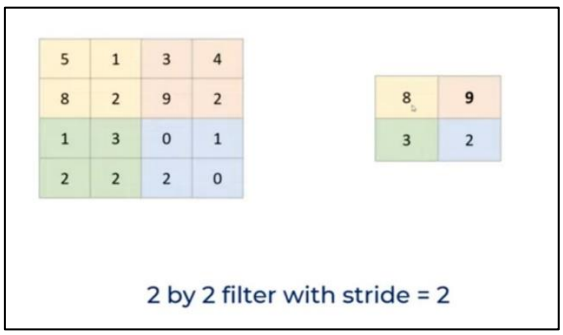
This feature map is for the loopy pattern feature. We take a grid of 3x3(can be of any size) and multiply by the grid of loopy pattern filter. We multiply first element of grid with the first element of the gid of the feature and the multiplication goes on. After the multiplying it we take the average of the values.

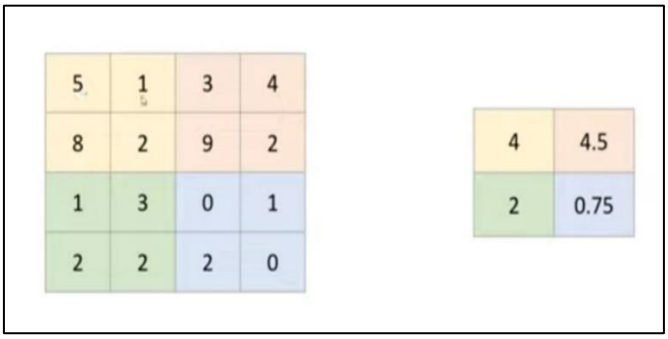
4) **<u>ReLU activation</u>**
In our feature map, numbers which are negative get replaced by 0 and the numbers above 0 remain the same



5) **<u>Pooling Layer</u>**
It is applied to the feature map to create a new feature map
How "quickly" its slides is called its stride length
eg:-a stride length of 2 means it moves by 2 pixels at a time until it spans the entire image

There are two types :- Max Pooling and Average Pooling



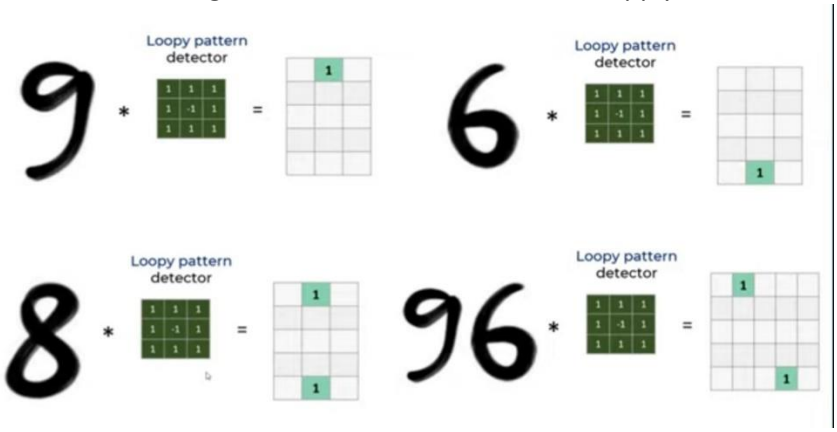


6) In the end, one has a fully connected dense network of neurons

C) Advantages of this algorithm over other algorithm

1) Advantages of convolution:-

- Reduces overfitting
  eg:-the image goes through convolution layer on a weight matrix 5x5x64. It generates 64 convolutions by sliding a 5x5 window. The model has 5x5x64=1600 parameters, which is remarkably fewer parameters than a fully connected network 256x256=65536
- Gives location invariant feature detection- it can detect the feature anywhere in the image
- Parameters sharing-if we filter out a feature we can apply it to the entire image



- The numbers of parameters is independent of the size of the original image

2) Advantages of ReLU:-

- Introduces nonlinearity
- Speeds up training, faster to compute

3) Advantages of Pooling:-

- Reduces dimensions and computation
- Reduces overfitting
- Makes the model tolerant towards small distortion and variations