

TERM PAPER

ARTIFICIAL INTELLIGENCE

September 30, 2021

Image Recognition and Monitoring using python

Roll Number - 19111026

BME/ 5th sem / B.Tech

1 Introduction

Image recognition in python gives an input image to a Neural network (the most popular neural network used for image recognition is Convolution Neural Network).

This is the main focus of our article that will be discussed in detail shortly.
The task is split mainly into two categories:

1. Classification of the image to a single category /multiple categories.
2. Identification of certain objects in an Image (This can be done only for the purpose of detection, segmentation, object tracking in videos, etc..)

Though final Tasks are different but the algorithm used in the neural network is the same.

2 Layers in recognition

The Input image consists of pixels. If it is a grayscale Image (B/W Image), it is displayed as a 2D array, and each pixel takes a range of values from 0 to 255. If it is RGB Image (coloured Image), it is transformed into a 3D array where each layer represents a colour.

2.1 Convolutional layer:

Purpose: Detect certain features in the image.

Operation: The convolution of Input Image and feature detector (or filter) is used to detect certain features in the image. Convolution occurs in the same manner as digital signal processing. Convolution occurs in the same manner as digital signal processing. Feature detector values can be predetermined if you know what features to extract from the image, or values can be initialized randomly, and the network training process determines the best filter values that fit our model.

Output: The output of this layer is called a feature map. The size of the feature map is less than the size of the image. This has the advantage of making the computation process easier. A point to elaborate is that part of image information is lost due to decreased output size. However, this doesn't cause a problem because the feature map's values are different from the original image as they represent the locations where the highest detection of the filter is performed.

2.2 Relu rectifier:

Purpose: increase non-linearity of images so they can be easily separable. Normally, images are highly non-linear because there are many details related to intensity, borders, etc. The convolutional layer can result in linear feature maps, so this step is highly crucial.

Operation: A relu rectifier is applied to the feature map

Output: The output of this layer is a feature map with higher non-linearity.

2.3 Maximum Pooling layer:

Purpose: Distinguish features if they are distorted. The main purpose is to detect features even if there is a slight difference in the feature itself.

Operation: Maximum pooling finds the maximum value of a certain window. The maximum pooling Layer shifts to the left by a certain number of steps called strides.

Output: Output of this layer is pooled feature map. Pooled feature map has multiple advantages. The output size is always smaller. Maximum values are still present, and these are the locations of highest similarity with the featured filter. In addition, more than 75% of image information that isn't related to features or is useless are removed. In addition, the Feature map becomes prominent to distortion if the feature value is shifted from its location.

Convolutional and MaxPool layers can be repeated more than once according to our machine learning problem. Then, We add MLP to the existing CNN. The main purpose of this step is to increase the number of feature attributes to make better class predictions.

2.4 Flattening

Numbers are taken row by row, column by column and put in a single column. The main purpose of this step is to convert matrix output from the previous layer to a format that can be accepted by ANN.

2.5 Fully Connected Layer

This is an artificial neural network where input is the flattened layer, followed by a group of fully connected layers—finally, the output layer according to categories that we have or objects that need to be detected.

3 PROGRAMMING LANGUAGES AND LIBRARIES MAINLY USED IN IMAGE RECOGNITION

There are many programming languages that are used in today's world for image recognition and monitoring systems. According to ranking they are:

- 1)Python - highly used for its fast performance and immense amount of libraries
- 2)C/C++ - second most used language for image recognition due its speed.
- 3)Java
- 4)Matlab - mainly used in monitoring systems

libraries mainly used for the process are:

- 1)Open-CV
- 2)Sci-kit image
- 3)Sci-py
- 4)Pillow/Pil
- 5)Numpy
- 6)Mahotas
- 7)SimpleITK
- 8)Pgmagick

Other Libraries used in the process are: 1)pandas

- 2)Matplotlib
- 3)Seaborn
- 4)os

4 How does image recognition work?

The following steps are done in the process of image recognition

4.1 Step 1

First, a great number of characteristics, called features are extracted from the image. An image is actually made of “pixels”. Each pixel is represented by a number or a set of numbers — and the range of these numbers is called the color depth (or bit depth). In other words, the color depth indicates the maximum number of potential colors that can be used in an image. In an (8-bit) greyscale image (black and white) each pixel has one value that ranges from 0 to 255. Most images today use 24-bit color or higher. An RGB color image means the color in a pixel is the combination of red, green and blue. Each of the colors ranges from 0 to 255. So a pixel contains a set of three values RGB(102, 255, 102) refers to color #66ff66. An image 800 pixel wide, 600 pixels high has $800 \times 600 = 480,000$ pixels = 0.48 megapixels (“megapixel” is 1 million pixels). An image with a resolution of 1024×768 is a grid with 1,024 columns and 768 rows, which therefore contains $1,024 \times 768 = 0.78$ megapixels.

4.2 Step 2

Once each image is converted to thousands of features, with the known labels of the images we can use them to train a model. Figure (B) shows many labeled images that belong to different categories such as “dog” or “fish”. The more images we can use for each category, the better a model can be trained to tell an image whether is a dog or a fish image. Here we already know the category that an image belongs to and we use them to train the model. This is called supervised machine learning.

4.3 Step 3

Training the data requires labelled data which is then converted into pixels and the fed into deep neural networks for processing. The huge networks in the middle can be considered as a giant filter. The images in their extracted forms enter the input side and the labels are in the output side. The purpose here is to train the networks such that an image with its features coming from the input will match the label in the right.

4.4 Step 4

Once a model is trained, it can be used to recognize (or predict) an unknown image. The new image will also go through feature extraction process.