Object Detection Model

By

Smith Bhojak
18BCE353
&
Sharan Chhugani
18BCE354

NIRMA
UNIVERSITY
INSTITUTE OF TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY
Ahmedabad 382481

OBJECT DETECTION MODEL

Mini Project - I

Submitted in fulfillment of the requirements

For the degree of

Bachelor of Technology in Information Technology

By

SMITH BHOJAK
18BCE353
SHARAN CHHUGANI
18BCE353

Guided By
DVIJESH BHATT
DEPARTMENT OF INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY
Ahmedabad 382481

CERTIFICATE

This is to certify that the project entitled “OBJECT DETECTION MODEL” submitted by SMITH BHOJAK (18BCE353.) & SHARAN CHHUGANI (18BCE354.)], towards the partial fulfillment of the requirements for the degree of Bachelor of Technology in Information Technology of Nirma University is the record of work carried out by him/her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination.

PROF. Dvijesh Bhatt
Assistant Professor
Department of Information Technology,
Institute of Technology,
Nirma University,
Ahmedabad

Dr. Madhuri Bhavsar
HOD, Dept. of Information Technology
Institute of Technology,
Nirma University,
Ahmedabad

ACKNOWLEDGEMENT

We would like to acknowledge with sincere gratitude and appreciation towards all those who helped/guided us for this mini project-I report, such as Prof. Dvijesh Bhatt’s support and criticism towards our work.Also we would like show appreciation for the faculties who helped us with areas that we were facing difficulties in.

ABSTRACT/ Outline

Object Detection Model as the name suggests is a process of classification,segmentation,captioning etc.This report touches the basic introduction to an Object Detection Model and later goes into explaining of CNN,YOLO,Darknet etc and then working of the project.

CONTENTS

Introduction

1.1 General

Image Classification is all about creating a network of Convoluted and Hidden layers that breaks down the image and maps it to classes' with various possible architectures and datasets.
The goal is to be able to detect the vehicles inside the images captured by the cctv and predict the no of vehicles than be accomodated.
We want to calculate the total amount of area available after vehicles each vehicle comes and goes and be able to give how many more of them can be accomodated.

1.2 Scope of Study

Our project is meant to cover the topics of how an image is taken as an input and how objects are detected using different methods,networks and systems.

Literature Survey

2.1 General

The Literature Survey covered documents such as papers, authentic articles that introduced Sentiment analysis, Deep learning techniques, and sentiment analysis using deep learning. The papers and articles are chosen in such a way that topics that were described in scope are covered.

About System

3.1 Convolutional Neural Networks

-In brief, an image going through various no of layers in form of matrices of pixelated values and then coming out as a single no representing a class of particular object.

-Cnn makes the machine learn to recognise objects in images and videos which has a wide range of applications.

-Making artificial neurons available to machine which fire when exposed to various components of images and forming a network of layers out of them, simulating the human neural system.

3.2 Layers and Network

3.2.1 Convolutional Layer

Convolutional layer is the first layer in the CNN. A 32*32*3 array of pixel is given as input to this convolutional layer. There is a filter which slides over the entire matrix (filter can be a 5*5*3 or 3*3*3 or 2*2*2) making the area on the input image touched as receptive field. This filter consists of weights which are multiplied with the pixel values of the input image in the receptive field and this process goes on until the filter glides over the entire image or array. The values after each multiplication are added and the result will be a 28*28*1 array of pixels known as an activation or feature map.This is the maths going behind the convolutional layer.

3.2.2 Fully Connected Layers

The fully connected layer is the last layer in the architecture.It takes as input, the output of the previous layer and gives as output a N dimensional vector where N is the total number of classes. Basically it takes into consideration, the high level features and tries to match it with all the classes and the probability for each class is given.

3.2.3 ReLU (Rectified Linear Units) Layers

This is said to be a non linear layer and should be present after each convolutional layer. Earlier sigmoid and tanh were used but ReLu was found more effective.It simply applied a function and converts all the negative values to 0.

3.2.4 Pooling Layer

Pooling layer or a downsampling layer is applied after the ReLu layer. Max Pooling is most common one of its variants. it takes a filter and convolves it on the input array with the same stride and takes the maximum number from the activations present. The output layer is much less in dimension as compared to the input layer.

3.2.5 Dropout Layers

Dropout layer are used to avoid overfitting while training the model with the train dataset. As its name suggests it simply takes a few activations and drops them or sets them to 0.

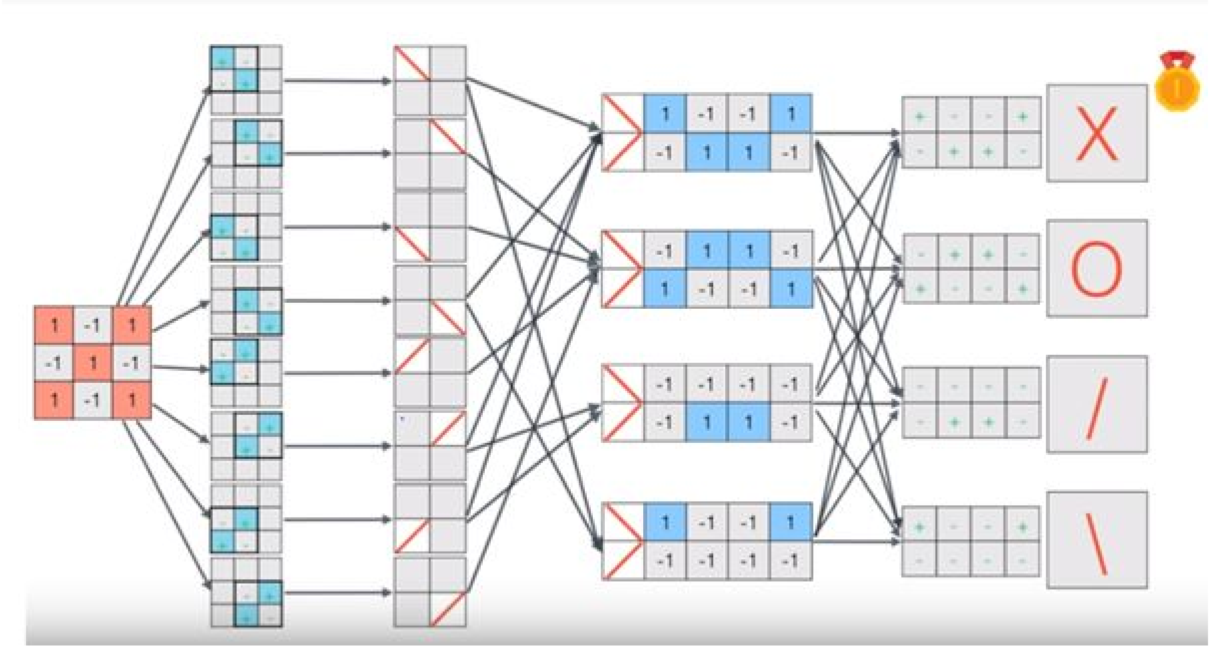
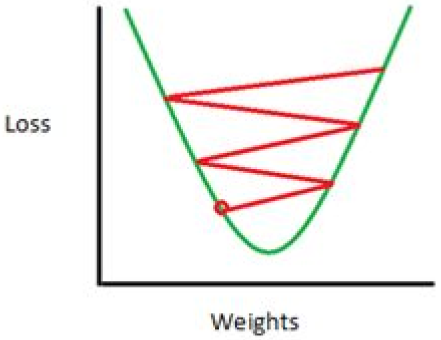


Fig3.1

3.3 Important Terms

-Error and Loss Function



Consequence of a high learning rate where the jumps are too large and we are not able to minimize the loss.

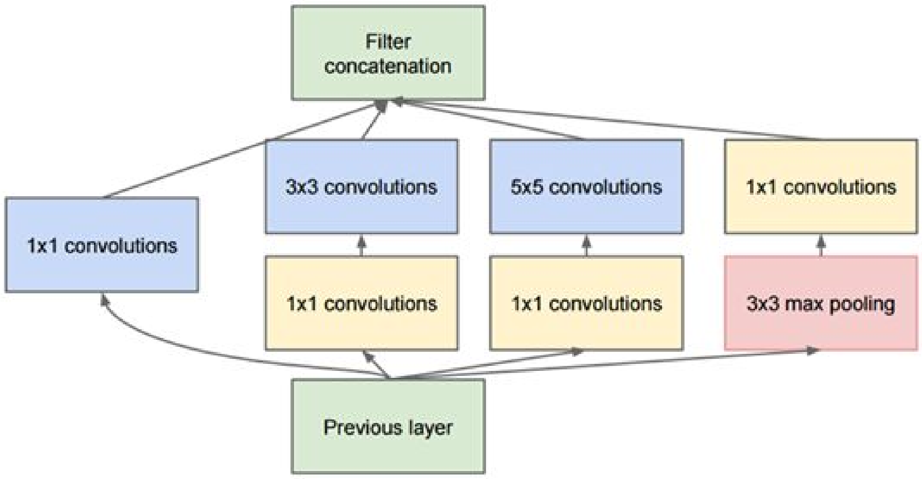
fig 3.2

-Stride And Padding

-Object Classification and Segmentation

-Transfer Learning

-Data Augmentation Techniques



Full Inception module

fig 3.3

3.5 Applications of the Network

-Image Recognition

-Object detections

-Facebook Tagging Algorithms

-Google SEO

-Products Recommendations ,etc

-Image recognition is just what it does at the core , its applications are way wider than that!

Implementation

4.1 Technology to be used

YOLO (real-time object detection system)

Darknet (Neural Network framework)

ImageNet (Image Database - for classification)

COCO (Common Object in Context - object detection, and captioning dataset)

4.2 You Only Look Once (YOLO)

- Provides real-time object detection system.
- Uses a single neural network to the full image.
- Network divides the image into regions and predicts bounding boxes and probabilities for each region.
- The Bounding boxes are weighted by the predicted probabilities.
- Predictions are informed by global context in the image.
- Makes predictions with a single network evaluation
- More than 1000x faster than R-CNN
- 100x faster than Fast R-CNN
- Multi-scale predictions

4.3 Darknet

- It is an open-source neural network framework that has been written in C and CUDA

- Supports both CPU and GPU computation so has much faster processing
- It also has Pre-trained weights so is user doesn't want to train the network they can just use the pre trained weights network

4.4 COCO

- It is a large-scale object detection, segmentation, and captioning dataset. which has wide array of features and flexibility
- Support for Object Segmentation
- Recognizing with context unlike usual recognition which only uses features COCO has context capability
- Has at-least 5 Captions per image so identifying any image can be simple and quick
- Around 80+ Object Categories
- 330K+ images of dataset makes it widely chosen among all other datasets

Pre-processing

5.1 Walkthrough of Pre-processing

Our primary goal here is to detect object from images in order to which we will need to pre-process the image.

- We will perform the following operation such as
  - Converting image to grayscale
  - Applying Bilateral filter (medianblur())on the image to blur unwanted details and preserve the edges.
  - Edge detection is performed using canny edge detection algorithm.
  - Dilation and erosion is performed.

5.2 Grayscale

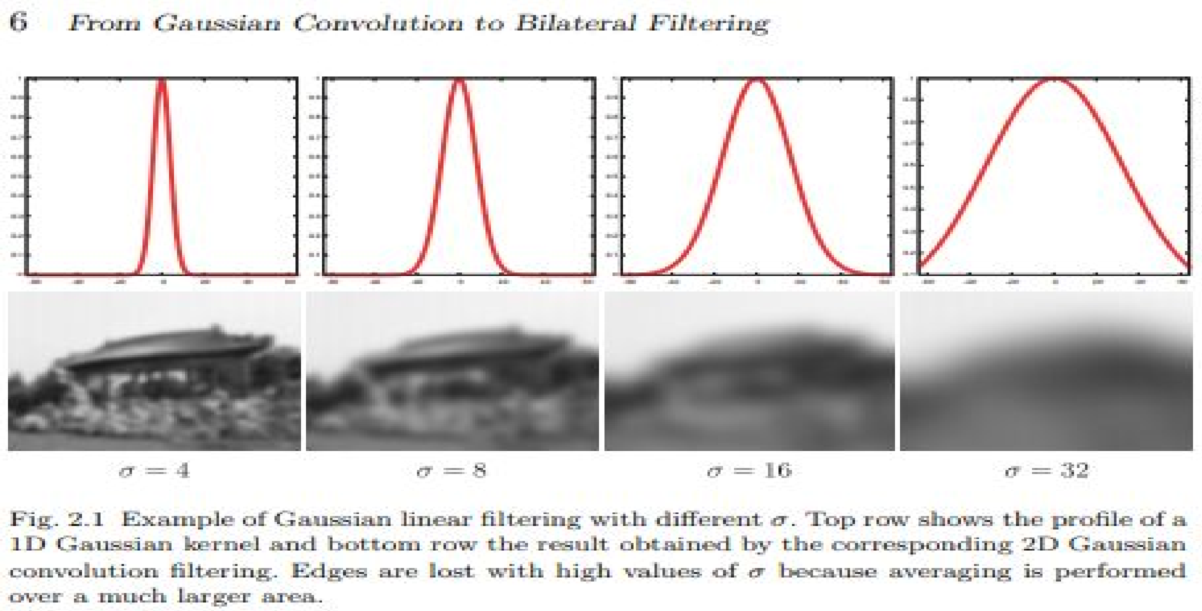
The reason for we differentiate images from any other sort of color image is because less information needs to be provided for each pixel. In fact a `gray' color is one in which the red, green and blue components al will be having equal intensity, and so it is only necessary to specify a single common intensity value unlike 3 values (RGB) for each pixel, as opposed to the three intensities needed to specify each pixel in a full color image.

Usually, the grayscale color intensity is to be stored as an 8-bit integer giving 256 possible different shades of gray from black to white. If all the levels of gray are evenly spread and spaced then differentiating between other gray-levels is significantly better than the gray-level perception power of the human eye.

Grayscale images are sufficient for most of the tasks that we need to carry out such as edge detection and classifying features and so there is no need to use any other harder-to-process color images.

5.3 Bilateral Filter

- The bilateral filter converts any input image to a smoothed version. It removes most of the unwanted texture, noise, and fine details, but preserves what we want, large sharp edges.
- The Median blur is a similar operation to the other averaging methods. Here, the central element of the image is replaced by the median of all the pixels in the kernel area. Hence this filter processes the main edges while discarding and removing the unwanted noise.
- Other image blurring and filtering functions are:-
    - Averaging
    - Gaussian Blur (each output image pixel value is a weighted sum of its neighbours in the input image)
    - Median blur
- This filter takes into consideration the differences in the values with the neighbours to preserve edges while performing smoothing of image.



6 *From Gaussian Convolution to Bilateral Filtering*

$\sigma = 4$ $\sigma = 8$ $\sigma = 16$ $\sigma = 32$

Fig. 2.1 Example of Gaussian linear filtering with different σ. Top row shows the profile of a 1D Gaussian kernel and bottom row the result obtained by the corresponding 2D Gaussian convolution filtering. Edges are lost with high values of σ because averaging is performed over a much larger area.

Fig 5.1

5.4 Canny Edge Detection

- Canny is a famous method to find a edge detection method, the 1st and most obvious criteria of edge detection is low error rate. The 2nd criteria is that the edge points should be well localized. The 3rd criteria is to have only one single response to a single edge.
- Non-maximum suppression . if value is less then threshold it will turn to zero otherwise an edge.
- Advantages:-Using probability for finding error rate, Localization and response. Improving signal to noise ratio, It has much better detection even in noisy images.
- Other possible edge detection methods
  - Sobel operator (3*3 convolution kernel detects horizontal and vertical edges)
  - Robert’s Cross operator (2*2 convolution kernel)

- Prewitt’s Operator
- Laplacian of Gaussian(It highlights all the regions of rapid intensity change , and has three kernels)(Malfunctioning at the corners, curves and where the gray level intensity function varies. Not finding what is the orientation of a particular edge because it uses the Laplacian filter)

5.5 Dilation

- Dilation adds pixels to the boundaries of the object of an image.
  It grows or thickens around a structuring element.
  An erosion after Dilation is Opening function.

5.6 Erosion

- Erosion removes boundaries on the object in the image
  It eliminates unwanted data.
  Dilation followed by Erosion is a closing function.

Working of System

6.1 Walkthrough

The following steps will be performed for detecting objects after detecting edges

Finding contours and sorting them from left to right. (Contours that form area less than the threshold area of any average car is discarded).
Bounding boxes are formed around the detected contours.
The midpoint of the bounding box is calculated
Distance between the midpoints is calculated using euclidean function
Then we draw the lines between the midpoint to show the final bounding box and then calculate the Pixel Per Metric with reference to width entered by the user.

Then we calculate 2 points A and B, divided by the Pixel Per Metric value A being the length and B being the width of the bounding box around the contour of edges detected.

Now we have our measurements of the objects we just need to identify the total space available to us in the parking structure.

First we simple calculate the total space occupied by objects detected and we subtract it from total estimated space of the available parking structure.

We also extract all the detected contours(objects) from the image and classify them whether they are vehicles or not.

Summary & Conclusion

7.1 Summary

In this report we have discussed the basics of object detection techniques and also covered what are concepts behind all of the functionality that is happening in the system/network.With which any beginner would find helpful for understanding and implementing object detection in images.

7.2 Conclusion

We can understand and realize the gravity and wide array of applications for object detection and image classification as there are no limits to the application,it can be used for something as simple as detecting cars inside parking lot and can be used for detecting cancer cells to act before harm is done.

References

1. ***Bilateral Filtering: Theory and Applications***

   [Sylvain Paris][Pierre Kornprobst][Jack Tumblin][Fr´edo Durand]

2. ***Study and Comparison of Various Image Edge Detection Techniques***

   [Raman Maini][Dr. Himanshu Aggarwal]

3. ***DEPTH AND GEOMETRY FROM A SINGLE 2D IMAGE USING TRIANGULATION***

   [Yasir Salih],[Aamir S. Malik]

Appendix A - List of Useful Websites

1. A Beginner's Guide To Understanding Convolutional Neural Networks (https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/)
2. How to measure the size of an object using a camera (https://web.archive.org/web/20131123073926/http://forestjohnson.blogspot.com/2010/01/how-to-measure-size-of-object-using.html)
3. Measuring size of objects in an image with OpenCV (https://www.pyimagesearch.com/2016/03/28/measuring-size-of-objects-in-an-image-with-opencv/)
4. Smoothing Images (https://docs.opencv.org/3.1.0/d4/d13/tutorial_py_filtering.html)
5. COMPARISON OF DIFFERENT EDGE DETECTION METHODS FOR VARIOUS COOL WEATHER CONDITIONS (https://www.researchgate.net/publication/310674105_COMPARISON_OF_DIFFERENT_EDGE_DETECTION_METHODS_FOR_VARIOUS_COOL_WEATHER_CONDITIONS)