Object Recognition and localization using Deep Convolutional Neural Network(CNN)

Author : Assefa Fikadu Medhin

Course: Intermediate Project

Instructor:Witold Paluszyński Ph.D.

Embedded Robotics,
Chair of Cybernetics and Robotics, Faculty of Electronics,

Wrocław University of Technology

January 2021

1 Abstract

Vision systems are essential in building a mobile robot that will complete a certain task like navigation, object grasping ,surveillance and other application.The main objective of this project is object detection which help vision system of mobile Robotics.In this project I will use, pre-trained Convolutional Neural Networks (CNN) MobileNet SSD(Single Shot Multi-Box Detector) model is used to detect object and localize the object within the image boundary by adding a box above the object including the name of the object class and accuracy. The image may contain different objects and if there multiple objects the system will detect multiple objects from single image.I have used my laptops web camera for real-time object detection experiment.

2 Introduction

Object detection is important concept in Robotics vision system.Object detection algorithms seek to detect the location of where an object resides in an image.Image classification involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. Object detection is more challenging and combines these two tasks and draws a bounding box around each object of interest in the image and assigns them a class label.object detection can be used for many applications like object grasping , video surveillance, medical imaging and robot navigation.

2.1 Convolutional Neural Networks (CNN)

The convolutional neural network is a specialized type of neural network model designed for working with image data.[3] Convolutional neural network (CNN) is a class of deep,feed-forward artificial neural network that has been utilized to produce an accurate performance in computer vision tasks, such as image classification and detection.IN deep learning each input image will pass it through a series of layers such us convolution layer, fully connected layers (FC), max pooling layer apply Softmax function to classify an object with probabilistic values between 0 and 1.

2.2 Mobilenet-ssd model

The mobilenet-ssd model is a Single-Shot multibox Detection (SSD) network intended to perform object detection. By using SSD, we only need to take one single shot to detect multiple objects within the image.This model is implemented using the Caffe* framework. Caffe is a deep learning framework made with expression, speed, and modularity in mind.

2.3 Single Shot MultiBox Detector(SSD)

The core of SSD is predicting category scores and box offsets for a fixed set of default bounding boxes using small convolutional filters applied to feature maps.This discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location.

- Single Shot: this is to mean that the tasks of object localization and classification are done in a single forward pass of the network.
- MultiBox: this is the name of a technique for bounding box regression developed by Szegedy et al[5]
- Detector: The network is an object detector that also classifies those detected objects

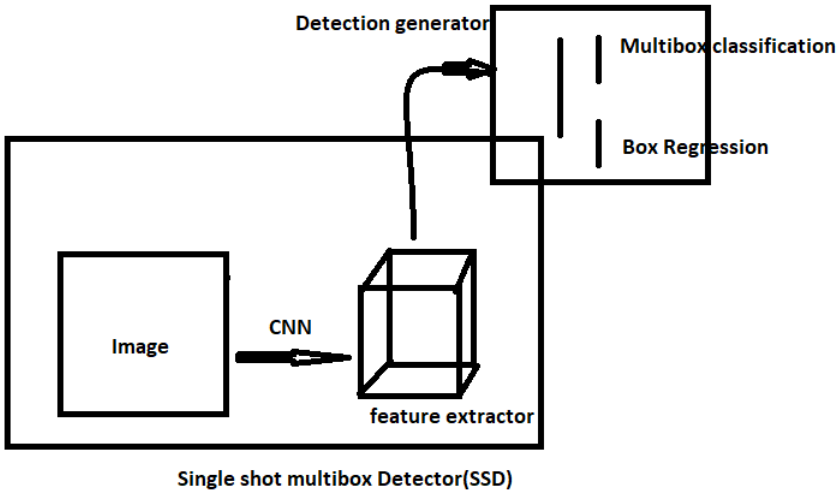

Figure 1: Architecture of SSD

2.4 Transfer learning

It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems.In this case I will used pre-trained model based approach where pre-trained model can then be used for object detection.The advantage of using this technique is it saves time to train the network from the start and less data is needed to get good results.[1]

2.5 Scope of the project

object detection model work for sample images and for real time using webcamera.Goals in this project is object detection which includes two points.

- object Recognition
  classify the object to a class from a set of predefined categories.

- object localization
  This is to locate where the object is and draw a bounding box around it.I will also add confidence level of the detected image[2].

3 Methodology

To implement object detection using CNN i will use opencv library using Python.the over all steps of the detection process can be see in the figure below.

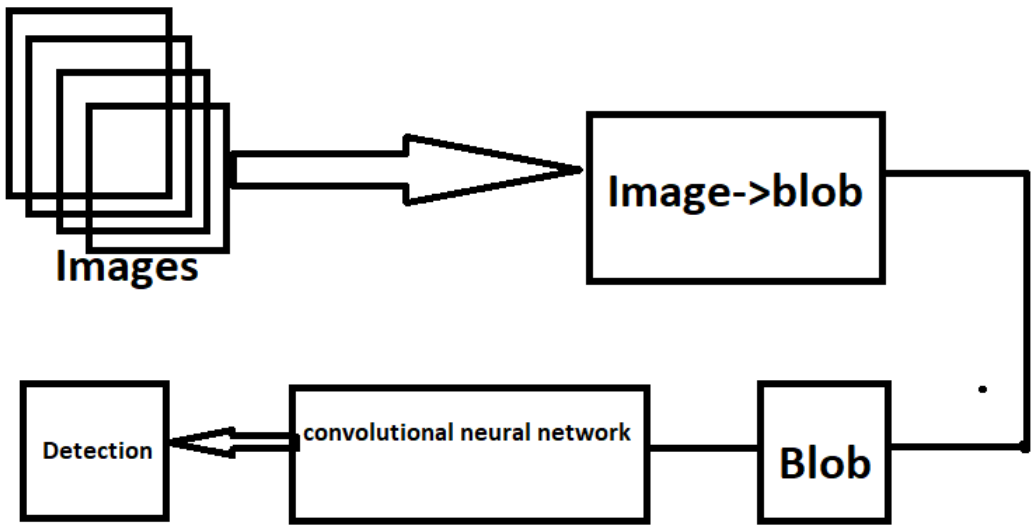

Figure 2: object detection process

3.1 Steps

- First Import the pre-trained object detection model using cv2.dnn.readNetFromCaffe() opencv library[4] which contains two files.
  - Binary file MobileNetSSD deploy caffemodel
  - prototxt file which contains Network schema for the above model.
- Then import the image
- Before passing to the Network we have to do image Pre-progressing.This is done using cv2.dnn.blobFromImage () opencv library.
  - The model input is a blob that consists of a single image of 1x3x300x300 in BGR order, also like the densenet-121 model. The BGR mean values need to be subtracted as follows: [127.5, 127.5, 127.5] before passing the image blob into the network.

  - After we perform mean subtraction we have to scale our images by some scaling factor.The scale factor should be

$$\frac{1}{\sigma} = \frac{1}{mean}$$

  - Generate blob from image as The model input is a blob that consists of a single image of 1x3x300x300 in BGR order.

- after pre processing we will pass the blob through the network and we will get Detections.[6]

  - The detection has the format: [image-id, label, conf, x-min, y-min, x-max, y-max]
  - image-id - ID of the image in the batch
  - label - predicted class ID
  - conf - confidence for the predicted class
  - (x-min, y-min) - coordinates of the top left bounding box corner (coordinates are in normalized format, in range [0, 1])
  - (x-max, y-max) - coordinates of the bottom right bounding box corner (coordinates are in normalized format, in range [0, 1])

- Finally, loop through the Detection and so that to extract features and detect object from the detection.

4 Result

object detection for sample images and for real time object detection using laptop web camera are shown below.

Figure 3: sample image 1

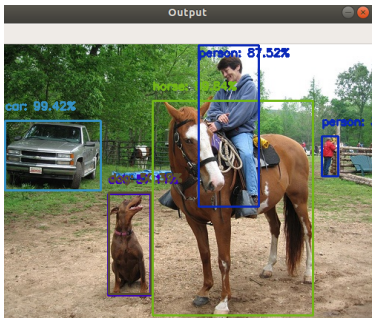

Figure 4: detection result 1

Figure 5: object detection for sample images

Figure 6: sample image 2

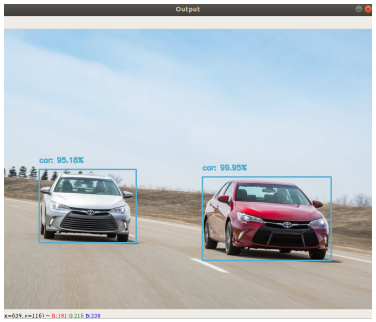

Figure 7: detection result 2

Figure 8: object detection for sample images

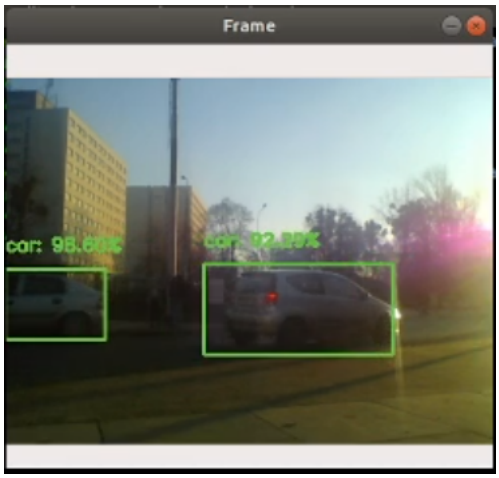

Figure 9: object detection 1

Figure 10: object detection 2

Figure 11: detection result from web camera

5 Conclusion

As we can see from the Result object detection procedure was successful and can detect the items such as a car, bicycle, bus, person etc using the MobileNet SSD detector.The performance of the model when detecting using web camera was quite good because the model was trained with a lot of data.such systems can be used for robots in different application areas for example we can use object grasping robot and object tracking and navigation.Generally object detection is applicable in robotic vision system.

6 References

[1] Younis Ayesha ,Shixin Li, Jn Shelembi , Hai, Zhang. (2020). Real-Time Object Detection Using Pre-Trained Deep Learning Models MobileNet-SSD 978-1-4503-7673-0. 44-48.

[2] Galvez Reagan , Bandala Argel , Dadios Elmer , Vicerra, Ryan Maningo, Jose Martin. (2018). Object Detection Using Convolutional Neural Networks.2023-2027.10.1109/TENCON.2018.8650517.

[3] A Brief History of CNNs in Image Segmentation

[4] Start Here with Computer Vision, Deep Learning, and OpenCV

[5] understanding ssd multibox real time object detection in deep-learning

[6] Details of Mobilenet-ssd Models public Mobilenet ssd Mobilenet ssd