



## Study and Implementation of SSD and MobileNetV1-V2 for Object detection and tracing

Smrati Srivastava, Tanish Agarwal, Raksha Chaudhary, Nitish Raj

<sup>1</sup>UG student, Computer Science Engg., IIMT College of Engineering, Greater Noida, India  
smrati12.iimt@gmail.com

<sup>2</sup>UG student, Computer Science Engg., IIMT College of Engineering, Greater Noida, India  
tanish.iimt@gmail.com

<sup>3</sup>UG student, Computer Science Engg., IIMT College of Engineering, Greater Noida, India  
raksha1999.iimt@gmail.com

<sup>4</sup>UG student, Computer Science Engg., IIMT College of Engineering, Greater Noida, India  
nitish.rraj@outlook.com

### Abstract-

Artificial intelligence is turning into thought day by day, from industrial work to unit chores. AI has begun to take over the globe in terms of Technology and use. currently once we state AI, deep learning and machine learning are the 2 basic ideas that strike our mind. For the past decade the advancement in AI and deep learning has gained tremendous traction over the traditional use of Technology. For a man-made machine to figure properly it desires vision over the skin world or with reference to its surroundings, so as for the machine to know the task and therefore the operating surroundings its managing. There are several algorithms out there for such functions, such as (R CNN) region primarily based convolutional neural network, faster-R CNN, you merely Look Once (YOLO) and Single Shot Detector (SSD). Among the algorithms mentioned on top of Yolo is that the quickest however others have higher accuracy compared to others. we are going to be gazing however object detection and object trailing is completed expeditiously with the mix of SSD and mobile nets as a victim of deep learning. This algorithmic rule not solely find object expeditiously it conjointly does not compromise on performance of the hardware that we have a tendency to use.

**Keywords**—MobileNetsV2, Single Shot Detector, Single Shot Multibox Detector, COCO.

### Introduction:

Fig.1 shows the fundamental diagram of detection and pursuit. during this paper, associate degree SSD and Mobile Nets primarily based algorithms are enforced for detection and pursuit in python setting. Object detection involves police investigation region of interest of object from given category of image. completely different strategies are –Frame differencing, Optical flow, Background subtraction. this is often a technique of police investigation associate degreeed locating an object that is in motion with the assistance of a camera. Detection and

pursuit algorithms are delineate by extracting the options of image and video for applications. options are extracted mistreatment CNN and deep learning. Classifiers are used for image classification and count. Ease of Use

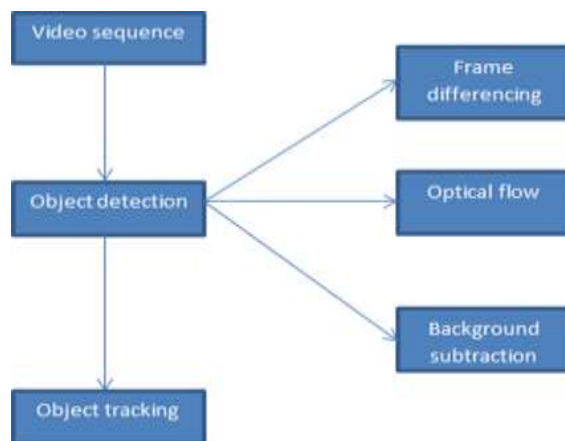


Fig. 1. Basic block diagram of detection and Tracking

## 1.1 SSD (single shot multibox detector)

SSD (single shot multibox detector), being quicker than faster- RCNN. it's become a move to alternative in object detection applications. SSD divide the image employing a grid and have a grid cell be chargeable for detection objects therein region of the image. it had been developed by Google around 2016. It takes just one shot at detection objects.

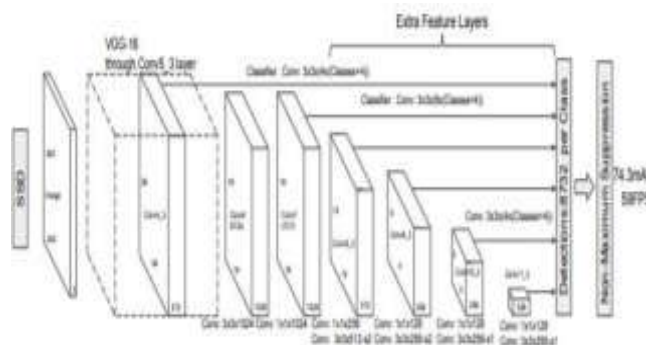


Fig.2 SSD (Single Shot Detector)

MobileNetsV2 square measure supported contour design that uses step wise dissociable convolution to create light-weight deep neural networks. it had been style for mobiles and embedded vision applications.

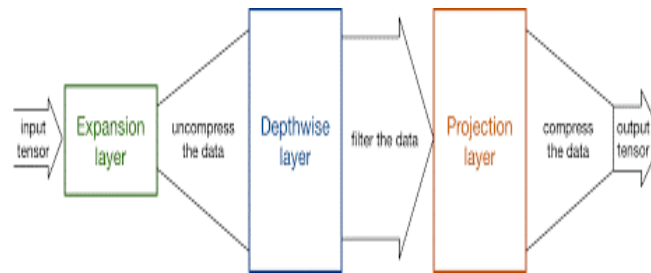


Fig.3 Mobile nets V2

Mobile-net was integrated into the SSD framework to tackle the sensible limitations of running high resource and power intense neural networks on low finish devices. It helps in real time applications to run sleek and need less hardware. therefore creating it price effective and additional correct. along it becomes MobileNetSSD.

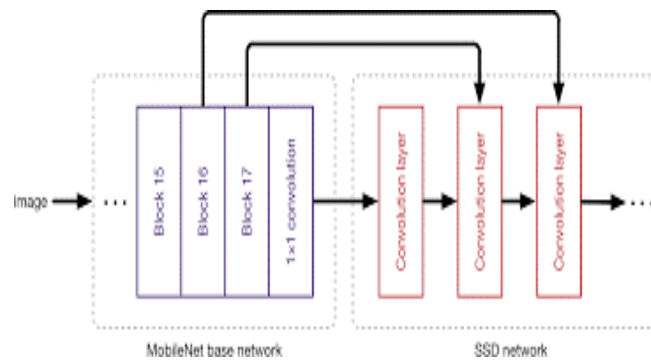


Fig.4 Mobile net SSD

The COCO bounding box format is [top left x position, prime leftposition, width, height]. The class id corresponds to one class per the classes section. every associate degree notation additionally has an id (unique to all or any different annotations within the dataset).

**1.2(a)How It is done:**

- The COCO file format.
- Changing associate degree existing dataset to COCO format.
- Loading a COCO dataset
- Visualizing and exploring your dataset.
- Generating predictions from associate degree object detection model.
- Evaluating map of a model on your COCO dataset.

Section II any elaborates however SSD and Mobile Nets rule works along, Section III explains methodology of implementation, and Section IV describes simulation results and analysis.

## 1.2 Object Detection and Tracking Algorithm

### 1.2.1 Single shot Detector (SSD)

SSD uses an identical part whereas coaching, to match the acceptable anchor box with the bounding boxes of every ground truth object at intervals a picture. primarily, the associate degree chore box with the very best degree of overlap with an object is chargeable for predicting that object's category and its location.

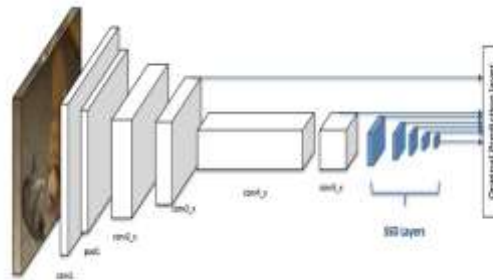


Fig.5 Inner working

### 1.2.2 Grid cell

Instead of victimization window, SSD divides the image employing a grid and have every grid cell be chargeable for detection objects therein region of the image. Detection objects merely suggests that predicting the category and site of associate degree object at intervals that region. If no object is gift, we tend to contemplate it because the background category and also the location is unnoticed.

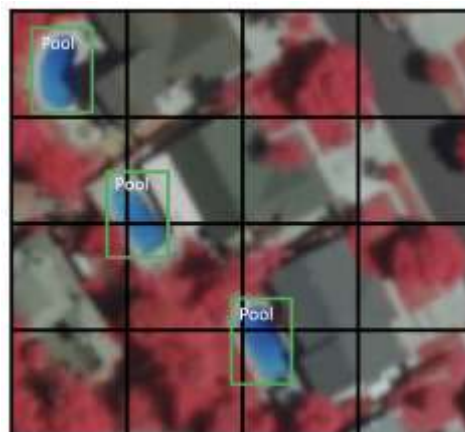


Fig.6

For instance, we tend to might use a 4x4 grid within the example below. every grid cell is ready to output the position and form of the item it contains.

### 1.2.3 Anchor box

Each grid cell in SSD are often allotted with multiple anchor/prior boxes. These anchor boxes square measure pre-defined and everyone is chargeable for a size and form at intervals a grid cell. for instance, the pool within the image below corresponds to the taller anchor box whereas the building corresponds to the broader box.

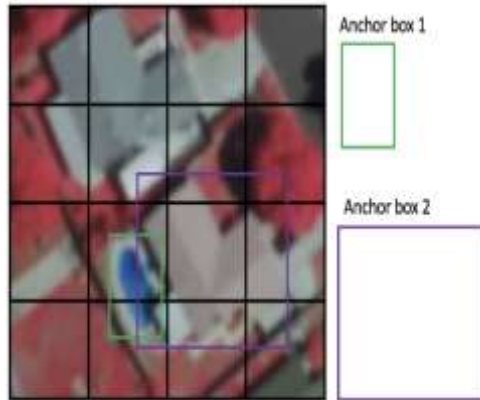


Fig.7

For matching the acceptable anchor box with the bounding boxes of every ground truth object at intervals a picture. SSD uses matching part whereas coaching. Anchor balls having the very best degree of overlap at intervals associate degree object is chargeable for predicting that object category and its location. every anchor box is specified by a facet quantitative relation and a zoom level.

#### 1.2.4 Aspect Ratio and Zoom Level:

Not all the objects within the world square measure of sq. form. Shapes of the item vary tall, length and breath. For attempt this downside, SSD permits predefined facet ratios of anchor boxes. The parameters of those ratios are often wont to specify the various facet ratios of the anchor box is related to every grid cell at every zoom / scale level.

#### Zoom Level

All the anchor boxes having identical size at the grid cell isn't necessary and that we additionally may well be fascinated by finding smaller or larger objects at intervals a grid cell. To specify the quantity of up and down scaling of the anchor boxes with relevance the grid cell, zoom parameter is employed.

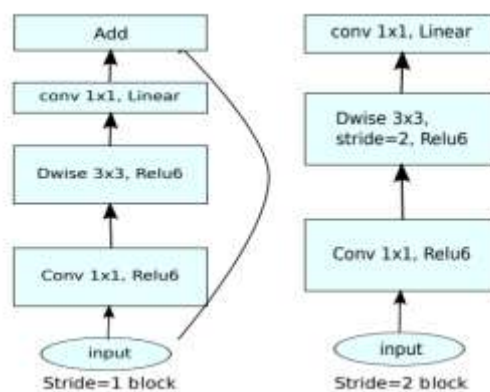


Fig.8 Mobile net V2

Based upon convolutional neural spec that seeks to perform well on mobile devices or in straightforward words it needs less hardware and power to work expeditiously. Latency low power models parameterized to fulfill the resource on sense of a range of use cases. It includes the progressive performance of mobile models on multiple task and Benchmarks further as across spectrum of various model sizes.

There square measure a complete of 3 layers:

- The 1st layer may be a one X one convolution with ReLU6, known as some extent wise convolution that is chargeable for building new options through computing linear mixtures of the input channels.
- Second layer is named depth wise convolution, perform light-weight filtering by applying one convolution filter for unit channel.
- Third there's another one x one convolution however with none non-linearity.

### 1.2.5 RELU6:

Its associate degree activation operates ordinarily employed in deep convolutional neural networks it's additionally you in mobile machine learning cases as a result of its optimize for Google's MobileNets design.

### 1.2.6 Strategies of Implementation:

#### (A) Object detection Frame differencing:

Frames square measure captured from camera at regular intervals of your time. distinction is calculable from the consecutive frames. Optical Flow

This technique estimates and calculates the optical flow field with rule used for optical flow. a neighbourhood mean rule is employed then to reinforce it. To filter noise a self-adaptive rule takes place. It contains a good adaptation to the amount and size of the objects and useful in avoiding time intense and complex preprocessing strategies.

#### (B) Background Subtraction:

Background subtraction (BS) methodology may be a fast methodology of localizing objects in motion from a video captured by a stationary camera. This forms the first step of a multi-stage vision system. this sort of method separates out background from the foreground object in sequence in pictures.



Fig.9 Detection of human from background subtraction



Fig. half-dozen depicts Detection of human from background subtraction. Foreground or person is detected and separated from the background of the image for any preprocessing. The separation result is shown step wise, once that localization of region of interest takes place.

## (C)Object tracking

It is drained video sequences like security cameras and CCTV police investigation feed; the target is to trace the trail followed, speed of associate degree object. the speed of real time detection are often accrued by using object trailing and running classification in few frames captured in an exceedingly fastened interval of your time. Object detection will run on a slow frame rates searching for objects to lock onto and once those objects square measure detected and fastened, then object trailing, will run in quicker frame speed.

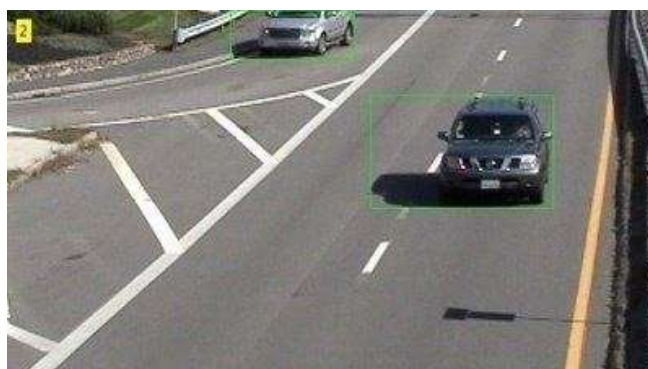


Fig 10 Tracking of car

Fig.10 shows the trailing of automotive. 2 ways in which during which the item are often half-track within the higher than example are: (1)-Tracking in an exceedingly sequence of detection. during this methodology a CCTV video sequence of a traffic that is in motion takes place. Suppose somebody needs to trace a automotive or person's movement here, he can take totally {different completely different} pictures or frames at different interval of your time. With the assistance of those pictures one will target the item sort of a automotive or person. Then, by checking however my object has touched in numerous frames of the video, I will track it. speed of the item are often calculated by corroboratory the object's displacement with the assistance totally different{of various} frames taken at different interval of your time. This methodology is really a flaw wherever one isn't trailing however detection the item at completely different intervals of your time. Improved methodology is "detection with dynamics". during this methodology estimation of car's mechanical phenomenon or movement takes place. By checking it's position at a selected time't' and estimating its position at over again interval let's say't+10'. From this particular image of automotive at 't+10' time are often planned with the assistance of estimation.

SIMULATION RESULTS AND ANALYSIS supported SSD rule, a python program was developed for the rule and enforced in OpenCv. OpenCV is run in Ubuntu IDE. Total twenty one objects were trained during this

model. the subsequent results square measure obtained once productive scanning, detection and trailing of video sequence provided by camera.



Fig.11(a) Detection of Bicycle with confidence level of 99.49%



Fig.11(b) Detection of Bus with confidence level of 98.68%



Fig.11(c) Detection of Train with confidence level of 99.99%





Fig.11(d) Detection of Dog with confidence level of 97.77%

Fig.11(a) to 11(b) shows the \$64000 time detection of bicycle, bus, train, and dog confidently levels ninety nine.49%, 98.68%, 99.99% and 97.77% respectively.

The model was trained to find twenty one objects category like dog, motorbike, person, potted plant, bird, car, cat, sofa, sheep, bottle, chair, aero plane, train, bicycle etc. with accurately of ninety nine.

### Conclusion:

We have with success studied the SSD rule operating with MobileNetV2 so as to perform object detection and tracing task we've pointed out that victimization SSD rule in real time state of affairs is incredibly effective and it shows results sizeable confidence level that's with smart accuracy. it's a reliable model because it needs less hardware creating it price effective and still it does not compromises on the performance.

Applications of search hardware and algorithms is actually big selection because it are often employed in drones, CCTV, security functions. additionally it leads towards the a part of automation with additional dependability and fewer maintenance. With low power consumption these algorithms are often employed in much nearly each piece of hardware and that they opens tons of potentialities within the field of computer science and advancement towards it.

Additionally they're GPU destined Algorithms, leading to quicker speed and fewer load on the hardware of the hardware.

### References:

- [1] <https://developers.arcgis.com/python/guide/how-ssd-works/#:-:text=SSD%20uses%20a%20matching%20phase,object's%20class%20and%20its%20location>
- [2] <https://heartbeat.fritz.ai/real-time-object-detection-using-ssd-mobilenet-v2-on-video-streams-3bfc1577399c>
- [3] <https://paperswithcode.com/method/ssd>



- [4] [https://medium.com/@smallfishbigsea/understand-ssd-and-implement-your-own-caa3232cd6ad#:~:text=SSD%20\(Single%20Shot%20MultiBox%20Detector,generally%20faster%20than%20Faster%20RCNN](https://medium.com/@smallfishbigsea/understand-ssd-and-implement-your-own-caa3232cd6ad#:~:text=SSD%20(Single%20Shot%20MultiBox%20Detector,generally%20faster%20than%20Faster%20RCNN)
- [5] <https://www.xailient.com/post/real-time-vehicle-detection-with-mobilenet-ssd-and-xailient#:~:text=Single%20Shot%20object%20detection%20or,class%20instances%20in%20those%20boxes>
- [6] <https://arxiv.org/pdf/1512.02325.pdf>
- [7] <https://paperswithcode.com/method/mobilenetv2#:~:text=MobileNetV2%20is%20a%20convolutional%20neural,perform%20well%20on%20mobile%20devices.&text=the%20bottleneck%20layers.-,The%20intermediate%20expansion%20layer%20uses%20lightweight%20depthwise%20convolutions%20to%20filter,a%20source%20of%20non%2Dlinearity>
- [8] <https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c>