Object Detection and Tracking using Tensor Flow

R. Sujeetha, Vaibhav Mishra

***Abstract:** Object detection and tracking could be a immense, vivacious however inconclusive and trending area of computer vision. Due to its immense use in official surveillances, tracking modules applied in security and lots of others applications have made researchers to devise a lot of optimized and specialized methods. However, problems are faced in implementing object detection and tracking in real-time; like tracking in real time and giving appropriate optimized results, over dynamic computation to find the efficient performance with respect to time factor, or multiple objects tracking create this task more difficult. Though, several techniques are devised but still lies a lot of scope of improvement, however during this literature review we've seen some illustrious and multiple ways of object detection and tracking. In this method we will be using Tensor Flow and Open CV library and CNN algorithm will be used and we will be labelling the detected layers with accuracy being checked at the same time .For validation purpose live input video will be taken for the same where objects will be getting detected and it can be simulated same for real-time through external hardware added .In the end we see the proper optimized and efficient algorithm for object tracking and detection.*

***Index Terms:** CNN, ML, OPENCV, TENSORFLOW*

I. INTRODUCTION

The object detection and tracking as a complete can be seen as an advanced mechanism to understand the objects present near to see. Years back when we see such method could have been literally compared to a virtually advanced artificial eye but with development of technology, we can figure out that the algorithm, machinery computation power and advanced datasets have made easier to devise a optimized method for object detection and tracking. The most important part of making or training an object detection architecture is using a proper dataset. The dataset should contain proper set of labelled images that should help in devising a proper technique getting used by the system to train various objects. We are here using Coco Dataset that contain thousands of images with hundreds of classifications that helps in detection and Classification of the objects that will be seen. Then using a proper classifier to train those images is very next thing to be kept under consideration. The faster rate defines the efficiency of object detection and tracking.

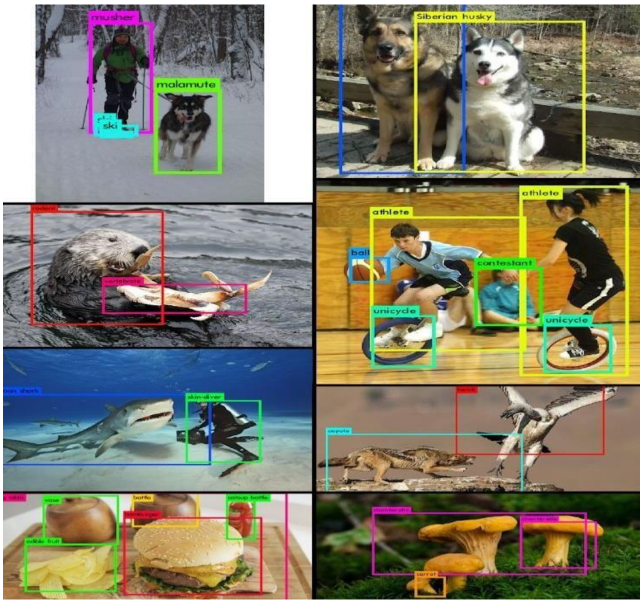

Fig 1 Detecting objects at multiple scenarios

Even though we lag behind the advanced algorithm detection systems' accuracy, we implemented the algorithm in a way that it is quick and efficient in detecting objects and people in images. Implementing this ideology in a video, we create a rough working of our base model. We still have difficulty in localizing some objects that are smaller in nature with respect to the the complete frame of the video. Although images that we require specificity comes with a steeper price tag than otherwise, which will most likely reduce the implementation of potential classification of datasets in foreseeable future. By the method of improvising the algorithm and usage of the data that we already possess, we intend to broaden our scope and implementation of the current detection system, using a Hierarchical method of classification, we will use coco datasets to our requiredsource.

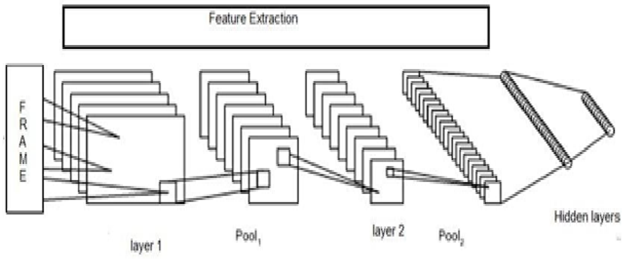

Fig 2 The Architecture The detection system 5 modules. The convolution layers are pretrained on the imagenet classifications with half the resolution and then trained with full resolution and coco datasets is being used

Revised Manuscript Received on 30 May 2019.

* Correspondence Author

R. Sujeetha*, Assistant Professor in Department of CSE, SRM Institute of Science and Technology, Chennai, Tamil Nadu

Vaibhav Mishra, student graduating from SRM Institute of Science and Technology, Chennai, Tamil Nadu

II. SYSTEMARCHITECTURE

Our goal is to improvise the algorithm to implement accurately even the farthest person. Datasets for our base project is limited, whereas datasets for other projects and tasks are abundant.

Coco dataset has over a thousands images with a huge range of category. The goal is to raise the detection scale to be at par with object classification.

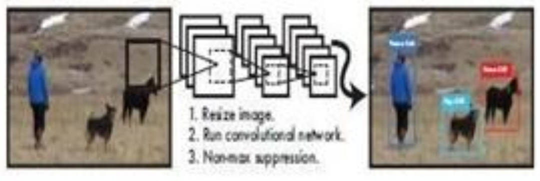

Fig 3 The Detection system: (1) resizes the input image (2) runs a single convolutional network (3) limits the detections by verifying with the dataset

A. LOCATIONDETECTION

When we are trying to locate the object in a scene the only problem that arises or we find is the instability of the model in predicting the object in image while creating the box around it using the coordinates that is being calculated using the given formula.Once the coordinates are known the box is set around the object by the neural layers predicts .

$$x = (tx *wa) - xa\ y = (ty *ha) - ya$$

This formula is not constrained so any bounding box will appear anywhere in the image. Instead of devising an architecture we have used an existing architecture to position the box of the cells .This makes the complete relative positioning of the bounding box ranged between 0 and 1.This logistic approach for constraints for the tracking and predicting the location of object falls between 0 AND 1p.

The architecture does not only predicts the boxes but also the ground values as coordinates that will be used to form box based on various values in terms of width and height (kx,ky),if box moves to either side the relative positioning changes as per the following formulas calculating it

$bx = \sigma(tx) + kx\, by = \sigma(ty) + ky\, bw = qwetw\ bh = qheth$

$(Pr(obj) *IOU(b, obj))= (\sigma(to))$

Then limit the location prediction which makes the architecture used easier to learn and the system becomes stable and optimized. It is improved by 5% when anchor boxes are used instead of dimension clusters

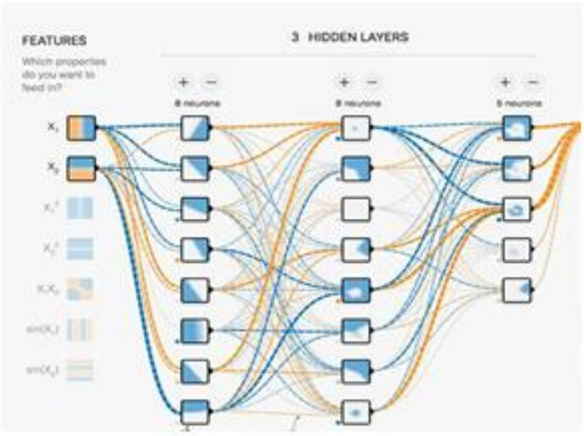

Fig 4 Bounding boxes with location predictions and dimension priors:

III. TENSORFLOW

Tensorflow is an open source library from Google that came few years back. This is a free library having huge application in industry as it meets all the organization standard expectations. It is a scientific library used for machine learning to handle analysis, prediction, stastical calculations on large scale and when working with neural layers.

TensorFlow was developed by the Google Brain team for their use in machine learning spectrum for theircompany.

Machine Learning has been today's pinnacle of technological advancement. Combining the latest tech with one of our most trusted and most basic application of a camera, Object detection.

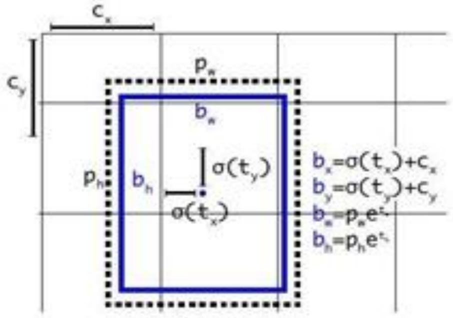

Fig 5 Tensorflow approach in designing neural layers

IV. TECHNIQUES FOR OBJECT DETECTION

There are various techniques that are used for acquiring the specifics of a object that make it unique. It is necessary to use these techniques to distinctly remember each object. There are three factors that define thisacquisition.

1. Better
2. Faster
3. Stronger

A. Better

This experiences an assortment of weaknesses with respect to cutting edge identification frameworks. Mistake investigation of contrasted with Fast R-CNN demonstrates that it makes a noteworthy number of confinement blunders. Moreover, it has generally low review contrasted with district proposition based strategies. In this way we center principally around improving review and confinement while keeping up order exactness. Geometric looks at distinguishing and differentiating features whereas photometric uses a calculated approach that transforms an image into values and then compares those with modular layers to eliminate the occurring variances.

In this, we lean towards using the geometric approach as it is easier to compute and store in adataset.

Some popular tracking algorithms are support vector machine, CNN, R-CNN, fast-CNN, (LDA)linear discriminant analysis, RNN, faster R-CNN matching using the SVM algorithm, the hidden Markov model, the multilinear subspace model using tensor representation, and the neural dynamic layer link matching.

Batch Normalization. Regularization being an important factor is removed by limiting everything to convergence by improving it in this method.

Convolutional With Anchor Boxes. Features get extracted by the neural layers and that helps to get coordinates that gives the important location on frame to form the boundary boxes predicting the object.

Fine-Grained Features:. Using a 13 cross squared image as feature map this models helps in predicting for larger objects giving advantages in image of finer grained features for detecting smaller objects. Faster R-CNN and SSD both run their models at varied frame in the network to get a range of predictions on varied resolutions in the image.

B. Faster

We wish detection to be correct however we have a tendency to additionally want it to be fast. As we see most of the applications lies in real time prediction no matter where it is used in surveillance or automatic cars or artificial intelligence depending on latency predictions levelling up the ground up for efficiency better in terms of time.

Training for detection We restructure the complete model of detection by removing an last layer replacing it with other layers of a 3*3 layers with filters ranging of 8 powers of 2 and then one last final layer for output at the time of detection. For VOC we predict five boxes that helps to provide the prediction of objects with 25 categories on 5 power 5 filters added to that model and then send all the classified object trained to the last model for fine grain for optimizing the complete spectrum of classification of objects.

As mentioned higher than, when our initial training on pictures at 224 × 224 we have a tendency to optimize our model to perform at a bigger size, 448. For this standardization we have a tendency to devise with the higher than expected frequency however for only ten epochs and initiate at a learning rate of model at 10−3. And when on this good value our model holds initial accuracy of 77% and at the best five accuracy of 93.3% starting from firstone.

Training for classification. Training the model on*the default given Coco class object detection dataset using gradient descent with initial learning rate of 1/10 using the neural network model at epochs value reaching to 160 .

This structure becomes better and faster in terms of accuracy as we deal with neural networks combined with tensorflow.

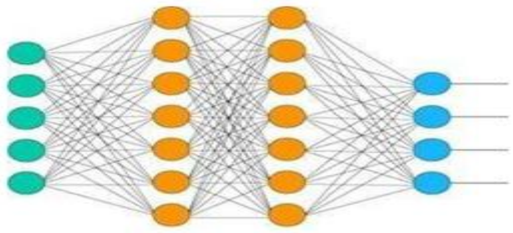

Figure6 Convolutional NeuralNetwork

C. Stronger

The model that we have structured for the detection and classification on the dataset it uses specific images labelled for the purpose of detection by forming the bounding box after the coordinate prediction and classification of that object in various classes.It brings down to images only that have specific class and that can be categorized based on the dataset after detection. So when our model sees an object in a frame for detection it back propagates to the classification of the object and loss function and render it back to the class of objects in the specific part of the architecture design to predict the output of theclass.The model that we are using presents us with some difficult as we are using multiple datasets to find the perfect object training model and label them while having Joint Classification and then labelling after the detection. While training such an complex large model we need to keep a check on the evaluation based on the combined dataset and our methodologies being used for the same so tha the challenges are not enclosed and do not tamper the model when brought down to applications so we need to add more categories for better application wise results. The corresponding coco for dataset has hundreds of classes of objects. Using RCNN model for this dataset with Imagenet so we can keep a balance the dataset by oversampling the COCO so that ImageNet is only greater by a 75 percent.The dataset being used for the same is Coco datasets trained on thousands of images with thousands of classes for training.

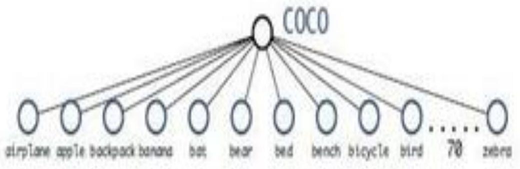

Figure 7 Combining datasets of Coco

Whenever it sees a image it initiates the classification and then back propagation module for limiting the amount of loss by finding the bounding box that is used for prediction and for that spectrum it calculates the loss that it is expected to throw. We expect .3IOU atleast label on the box and hen back propagate loss based on the classification and assumptions. This method reduces loss

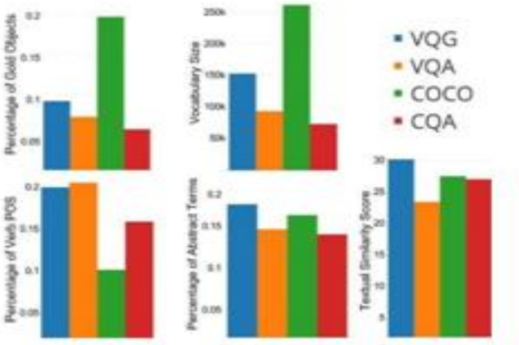

Figure 8 Comparison of datasets

D.COMBINATION OFTECHNIQUES

This combination of above stated algorithm in a manner that it holds the constraints very well tuned with optimization and efficiency and gives proper output as per detection of object.One thing to be taken under consideration that it requires or pushes high graphic on the system to utilise the training of the batch in the training process leading to delay which is being then maintained by optimizing the neural layers stacked on it.This then utilized on Coco dataset is trained using tensorflow library .The above listen properties holds the specific entities that makes this to execute properly on the local machine pushed by the training models .

E. R-CNN

R-CNN accelerates the arrangement phase of CNN however despite everything it depends upon the region and box that it frames up after training and classifying the objects after a certain period of time. As when it maps an image it forms the box after the completion of training and that may take upto 2 seconds at 0.5 fps in real time.The recent advancement in neural layers allowed the modules for a faster RCNN model that can give upto 10fps forming the bounding box but with less accurate model but at higher rate of output and it can go upto 18fps but with less score more.The faster RCNN model is mAP hi-gher than other models but slower in terms of many things when it comes to prediction of classified objects.

we compare this solemnly depends to the GPU and use of high computation power which makes it run at 30 hertz .which optimizes this giving it a great breakthrough and comparatively better mAP accessible in protest recognition frameworks. Quick YOLO is the speediest protest location technique on PASCAL , till the limits we know it has the quickest surviving and training object model with good number of prediction too and classification of objects in real time pushing upto 64% accuracy boost with53%maP

V. APPLICATIONS

A. APPLICATIONS IN SURVEILLANCE

With terror being one of the most concerned artifacts of the modern-day world, we plan on facing it with one of ours. Using the technicalities of it, we plan on starting on a small scale, our dataset will be unique. We, instead of using an already created dataset and running out of funds too soon, will be creating datasets with help of our program to work and detect humans of a particular campus. This will not only give us the edge over modern security, it will also alert us if any unusual activity occurs, by the implementation of Machine Learning mechanism, we learn the pattern of each individual's behavior, all in real-time.

B. APPLICATIONS INARTIFICIAL INTELLIGENCE

We are also going to implement the ideology of it to further improve the working and the functionality of pre-existing monitoring softwares/technologies. Monitoring systems of the present day heavily depend on human resource and the working of certain out-of-date technologies related to CCTV security[3]. On implementation of Real-Time Object detection with help of Machine Learning, we reduce the risk of human error and increase the efficiency of the level of security and feasibility. Eg. A child lost in the mall, here, our system helps to recognize the behavior pattern of a distressed child, a child without supervision, and similar constraints, data of which will be alerted to the security who can proceed with reasonablehelp.

C.APPLICATIONS INAUTOMOBILES

With the rapid growth in automobiles and automated cars and vehicles being designed we need an optimized system that can handle this task very efficiently.So this technology can be used very well in automated cars where humain fails or give errors machine may perform better at operations and may be beneficial in designing system for automated vehicles Tesla being the most advanced one in creating and manufacturing the same it needs to advance it approach to make the system more clear and this can help in for thesame.

VI. CONCLUSION

This is an ongoing structure for discovery more than thousand item classes by mutually upgrading identification also, arrangement. We utilize tensorflow to join information from different sources and our joint improvement strategy to prepare all the while on COCO. This is a solid advance towards shutting the dataset measure hole between recognition also,characterization.

A large number of our systems sum up outside of item location. Our portrayal of preparing offers a more extravagant, increasingly nitty gritty yield space for picture grouping. Dataset blend utilizing various leveled grouping would be valuable in the order and division areas.

Preparing procedures like multi-scale preparing could give advantage over an assortment of visual undertakings.

VII. FUTUREWORK

For future work we see how advancement of computation power and leading development in machine learning will enhance and help in optimizing the efficiency of object detection and tracking. Computer vision is in trending development phase and can totally change the scenario of visual artificial advancement. We will always be looking for ways that can get be brought together to create better models for this training spectrum. The perfect amalgamation of high computation power and huge classified datasets will help to improvise the current model enhancing it in all aspects. of the work or suggest applications and extensions.

REFERENCES

1. Joseph Redmon, AliFarhadi, YOLO9000:Better, Stronger, Faster, 2017 IEEE, pp. 6517-6526 Chengtao Cai, Boyu Wang, Xin Liang, A New Family Monitoring Alarm System Based on Improved YOLO Network, pp.4269-4274 Alexaender M., Micheal M. and Ron Kimel, 3-DFace Recognition, unpublished, First version: May 18, 2004; Second version: December 10,2004.
2. R. Bruneelli and T. Poggio, "Face Recognition: Features versus Templates", IEEE , 1993,(15)10:1042-1052[6] Albiol, A Oliver, J., Messi, J.M. bronsky using depth cameras. Computer Vision, IET. Vol 6(5),378-387.
3. http://www.aeroclubsocal.org/2015/02/15
/ -elon-musk- /
4. ICCV 2009 - International Conference on Computer Vision, Sep 2009, Kyoto, Japan. IEEE, pp.498-505, 2009, ⟨10.1109/ICCV.2009.5459197
5. H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In IEEE PAMI, volume 20,1998.
6. Better face Recongnition software computers at recognizing faces in recent tests. By Mark Willaims
Pontin May 30,2007MIT

7. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: 2016 IEEE Conference, pp.780-788

AUTHORS PROFILE

R. Sujeetha Working as an Assistant Professor in Department of CSE, SRMIST and Ramapuram from 21th June 2018 to till date

Vaibhav Mishra a pre final year student graduating from SRMIST ,machine learning enthusiast and keen interest in development and innovation in field of artificial intelligence