ISSN: 2366-1313

AI BASED OBJECT RECOGNIZATION AND TRACKING SYSTEM

[1]Mrs. C. ARCHANA, [2]Gone Vishnuvardhan, [3]Buyyani Harisha, [4]Anneboi Vishlesha, [5]Chinikal Preethi

[1]Assistant Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

archanachinna5491@gmail.com

[2]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

vishnuvardhangone25@gmail.com

[3]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

buyyaniharisha@gmail.com

[4]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

vishleshaanneboi@gmail.com

[5]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

preethichinikai@gmail.com

Abstract: *Object detection is a well-known computer technology connected with computer vision and image processing that focuses on detecting objects or its instances of a certain class (such as humans, flowers, animals) in digital images and videos. There are various applications of object detection that have been well researched including face detection, character recognition, and vehicle calculator. Object detection can be used for various purposes including retrieval and surveillance. In this study, various basic concepts used in object detection while making use of OpenCV library of python3, improving the efficiency and accuracy of object detection are presented. Deep learning has gained a tremendous influence on how the world is adapting to Artificial Intelligence since past few years. Some of the popular object detection algorithms are Region-based Convolutional Neural Networks (RCNN), FasterRCNN, Single Shot Detector (SSD) and You Only Look Once (YOLO). Amongst these, Faster-RCNN and SSD have better accuracy, while YOLO performs better when speed is given preference over accuracy. Deep learning combines SSD and Mobile Nets to perform efficient implementation of detection and tracking. This algorithm performs efficient object detection while not compromising on the performance.*

Keywords: *Object tracking, OpenCV, computer vision, Webcam, NumPy*

I. INTRODUCTION

Since AlexNet has stormed the research world in 2012 ImageNet on a large-scale

ISSN: 2366-1313

visual recognition challenge, for detection in-depth learning, far exceeding the most traditional methods of artificial vision used in literature. In artificial vision, the neural convolution networks are distinguished in the classification of images. Fig. 1. Basic block diagram of detection and Tracking Fig. 1 shows the basic block diagram of detection and tracking. In this paper, an SSD and MobileNets based algorithms are implemented for detection and tracking in python environment. Object detection involves detecting region of interest of object from given class of image. Different methods are –Frame differencing, Optical flow, Background subtraction. This is a method of detecting and locating an object which is in motion with the help of a camera. Detection and tracking algorithms are described by extracting the features of image and video for security applications. Features are extracted using CNN and deep learning. Classifiers are used for image classification and counting. YOLO based algorithm with GMM model by using the concepts of deep learning will give good accuracy for feature extraction and classification[1].

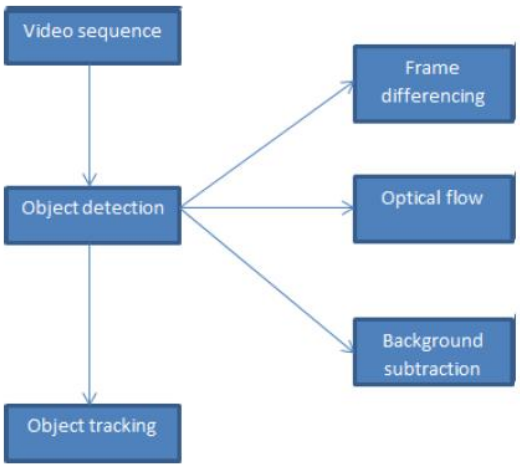

Fig.1 Object recognition

Over the past years domains like image analysis and video analysis has gained a wide scope of applications. CV and AI are two main technologies dominating technical society. Technologies try to depict the biology of human. Human vision is the sense through which a perception of outer 3D world is perceived. Human Intelligence is trained over years to distinguish and process scene captured by eyes. These intuitions act as a crux to budding new technologies. Rich resource is now accelerating researchers to excavate more details form the images. These developments are due to state of the-art methods like CNN. Applications from Google, Facebook, Microsoft, and Snapchat are all results of tremendous improvement in Computer vision and Deep learning. During time, the vision-based technology has transformed from just a sensing modality to intelligent computing systems which can understand

ISSN: 2366-1313

the real world. Computer vision applications like vehicle navigation, surveillance and autonomous robot navigation find Object detection and tracking as important challenges. For tracking vehicles and other real word objects, video surveillance is a dynamic environment. In this paper, efficient algorithm is designed for object detection and tracking for video Surveillance in complex environment. Object detection and tracking goes hand in hand for computer vision applications. Object detection is identifying object or locating the instance of interest in-group of suspected frames. Object tracking is identifying trajectory or path; object takes in the concurrent frames. Image obtained from dataset is, collection of frames[2].

Basic block diagram of object detection and tracking is shown in Fig. 1. Data set is divided into two parts. 80 % of images in dataset are used for training and 20 % for testing. Image is considered to find objects in it by using algorithms CNN and YOLOv3. A bounding box is formed across object with Intersection over union (IoU) > 0.5. Detected bounding box is sent as references for neural networks aiding them to perform Tracking. Bounded box is tracked in concurrent frames using Multi Object Tracking (MOT). Importance of this research work is used to estimate traffic density in traffic junctions, in autonomous vehicles to detect various kinds of objects with varying illumination, smart city development and intelligent transport systems [3].

Object detection is perhaps the main exploration research in computer vision. Object detection is a technique that distinguishes the semantic objects of a specific class in digital images and videos. One of its real time applications is self driving vehicles or even an application for outwardly hindered that identifies and advise the debilitated individual that some object is before them. Object detection algorithms can be isolated into the conventional strategies which utilized the method of sliding window where the window of explicit size travels through the whole image and the deep learning techniques that incorporates YOLO algorithm. In this, our point is to distinguish numerous objects from an image. The most well-known object to identify in this application are the animals, bottle, and people. For finding the objects in the image, we use ideas of object localization to find more than one object in real time. There are different techniques for object identification, they can be separated into two classifications, initial one is the algorithms dependent on Classifications.

ISSN: 2366-1313

CNN and RNN go under this classification. In this classification, we need to choose the interested areas from the image and afterward need to arrange them utilizing Convolution Neural Network. This strategy is slow as we need to run an expectation for each selected area. The subsequent class is the algorithms dependent on Regressions. YOLO strategy goes under this classification. In this, we won't need to choose the interested regions from the image. Rather here, we predict the classes and bounding boxes of the entire image at a single run of the algorithm and afterward distinguish different objects utilizing a single neural network. YOLO algorithm is quicker when contrasted with other grouping algorithms. YOLO algorithm makes localization errors but it predicts less false positives in the background. This document is template. We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace (copy-paste) the content with your own material. Number the reference items consecutively in square brackets (e.g. [1]). However, the authors name can be used along with the reference number in the running text. The order of reference in the running text should match with the list of references at the end of the paper.

MOTIVATION

Blind people do lead a normal life with their own style of doing things. But they definitely face troubles due to inaccessible infrastructure and social challenges. The biggest challenge for a blind person, especially the one with the complete loss of vision, is to navigate around places. Obviously, blind people roam easily around their house without any help because they know the position of everything in the house. Blind people have a tough time finding objects around them. So, we decided to make a REAL TIME OBJECT DETECTION System. We are interested in this project after we went through few papers in this area. As a result, we are highly motivated to develop a system that recognizes objects in the real time environment.

II. LITERATURE SURVEY

In the year 2017 Tsung-Yi Lin, Piotr Dollar, Ross Girshick, KaimingHe, BharathHariharan, and SergeBelongie proposed Feature Pyramid Networks for Object Detection. With the launch of Faster-RCNN, YOLO, and SSD in 2015, it seems like the overall structure an object identifier is resolved. Analysts begin to take a gander at improving every individual pieces of these networks. Highlight Pyramid Networks is an

ISSN: 2366-1313

endeavour to improve the identification head by utilizing highlights from various layers to frame a feature pyramid. This feature pyramid thought isn't novel in computer vision research. In those days when highlights are still physically planned, feature pyramid is now a powerful method to recognize patterns at various levels. Utilizing the Feature Pyramid in deep learning is likewise not a ground breaking thought: SSPNet, FCN, and SSD all showed the advantage of aggregating multiple layer highlights before classification. Nonetheless, how to share the feature pyramid among RPN and the region-based detector is still yet to be resolved.

In the year 2017 Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick proposed Mask R-CNN. In this paper Mask RCNN is certainly not a commonplace object detection network. It was intended to settle a difficult example division task, i.e, making a mask for each object in the scene Nonetheless, Mask R-CNN indicated an incredible augmentation to the Faster R-CNN framework, and furthermore thusly motivated object location research. The fundamental thought is to add a binary mask prediction branch after ROI pooling alongside the current bounding box and characterization branches. Obviously, both perform multiple tasks preparing (division + detection) and the new ROI Align layer add to some improvement over the bounding box benchmark.

In the year 2017 Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis proposed Soft-NMS – Improving Object Detection with One Line of Code. In this paper Non-maximum suppression (NMS) is broadly utilized in anchor based object detection networks to diminish copy positive proposition that are close-by. All the more explicitly, NMS iteratively wipes out applicant boxes on the off chance that they have a high IOU with a surer applicant box. This could prompt some sudden conduct when two objects with a similar class are to be sure near one another. Soft NMS rolled out a little improvement to just downsizing the certainty score of the overlapped applicant boxes with a boundary. This scaling boundary gives us more control when tuning the localization execution, and furthermore prompts a superior exactness when a high review is likewise required.

In the year 2017 ZhaoweiCai UC San Diego, Nuno Vasconcelos UC San proposed Cascade R-CNN: Delving into High Quality Object Detection. While FPN investigating how to plan a superior R-CNN neck to utilize backbone highlights Cascade R-CNN examined an

ISSN: 2366-1313

upgrade of R-CNN grouping and regression head. The basic assumption that is straightforward yet sagacious: the higher IOU rules we utilize while planning positive focuses on, the less false positive predictions the network will figure out how to make. In any case, we can't just increment such IOU threshold from regularly utilized 0.5 to more forceful 0.7, in light of the fact that it could likewise prompt all the more overpowering negative models during training. Cascade R-CNN's answer is to chain various recognition head together, each will depend on the bounding box recommendations from the past detection head.

In the year 2017 Tsung-Yi Lin PriyaGoyal Ross GirshickKaiming He Piotr Dollar proposed Focal Loss for Dense Object Detection. To comprehend why one-stage locators are typically not comparable to two-stage detectors, Retina Net explored the frontal area foundation class unevenness issue from one-stage detectors dense predictions. Take YOLO for instance, it attempted to predict classes and bounding boxes for all potential areas meanwhile, so the majority of the yields are coordinated to negative class during training. SSD tended to this issue by online hard model mining. YOLO utilized an objectiveness score to certainly prepare a closer view classifier in the beginning phase of training. Retina Net thinks the two of them didn't get the way in to the issue, so it developed another loss function work called Focal Loss to assist the network with realizing what's significant. Focal Loss added a power γ to Cross-Entropy loss. The α boundary is utilized to adjust such a focusing effect.

In the year 2018 Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, JiayaJia proposed Path Aggregation Network for Instance Segmentation. In this paper Occurrence division has a close relationship with object detection, so regularly another case segmentation network could likewise profit object recognition research in a roundabout way. PANet targets boosting data stream in the FPN neck of Mask R-CNN by adding an extra base up path after the first top-down path. To picture this change, we have a ↑↓ structure in the first FPN neck, and PANet makes it more like a ↑↓↑ structure prior to pooling highlights from various layers. Likewise, rather than having separate pooling for each element layer, PANet added an "adaptive feature pooling" layer after Mask RCNN's ROI Align to merge multi-scale features.

In the year 2018 ChengjiLiu, Yufan Tao, JiaweiLiang, Kai Li, Yihang Chen proposed Object Detection Based on YOLO Network.In this paper YOLO v3 is

ISSN: 2366-1313

the latestform ofthe YOLO versions. Following YOLOv2's convention, YOLOv3 acquired more thoughts from past exploration and got a powerful incredible one-stage finder like a beast. YOLO v3 adjusted the speed, exactness, and execution unpredictability really well. Also, it got truly main stream in the business as a result of its quick speed and basic parts. Basically, YOLO v3's success comes from its all the more impressive backbone include extractor and a RetinaNet-like identification head with a FPN neck. The new spine networkDarknet-53 utilized ResNet's skip connections with accomplish a precision that is comparable to ResNet-50 yet a lot quicker.

III. PROPOSED WORK

Object Detection is the process of finding and recognizing real-world object instances such as car, bike, TV, flowers, and humans out of an images or videos. An object detection technique lets you understand the details of an image or a video as it allows for the recognition, localization, and detection of multiple objects within an image. It is usually utilized in applications like image retrieval, security, surveillance, and advanced driver assistance systems (ADAS)

Object Detection is done through many ways:

➢ Feature Based Object Detection

➢ Viola Jones Object Detection

➢ SVM Classifications with HOG Features

➢ Deep Learning Object Detection

Object detection from a video in video surveillance applications is the major task these days. Object detection technique is used to identify required objects in video sequences and to cluster pixels of these objects. The detection of an object in video sequence plays a major role in several applications specifically as video surveillance applications.

Object detection in a video stream can be done by processes like pre-processing, segmentation, foreground and background extraction, feature extraction. Humans can easily detect and identify objects present in an image. The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects with little conscious thought. With the availability of large amounts of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy.

ISSN: 2366-1313

ALGORITHMS

R-CNN

R-CNN is a progressive visual object detection system that combines bottom-up region proposals with rich options computed by a convolution neural network. R-CNN uses region proposal ways to initial generate potential bounding boxes in a picture and then run a classifier on these proposed boxes.

SINGLE SIZE MULTI BOX DETECTOR

SSD discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At the time of prediction, the network generates scores for the presence of each object category in each default box and generates adjustments to the box to better match the object shape.

Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

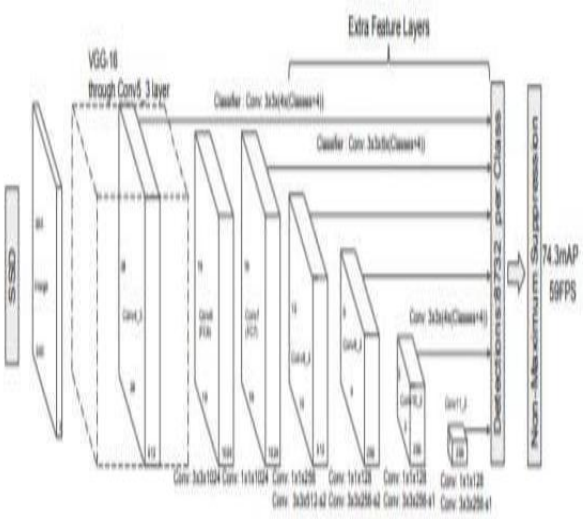

Fig.2 VGG-16 SSD Model

ALEXNET

AlexNet is a convolutional neural Network used for classification which has 5 Convolutional layers, 3 fully connected layers and 1 soft max layer with 1000 outputs for classification as his architecture.

YOLO

YOLO is real-time object detection. It applies one neural network to the complete image dividing the image into regions and predicts bounding boxes and possibilities for every region.

Predicted probabilities are the basis on which these bounding boxes are weighted. A single neural network predicts bounding boxes and class possibilities directly from full pictures in one evaluation. Since the full detection pipeline is a single network,

ISSN: 2366-1313

it can be optimized end-to-end directly on detection performance.

VGG

VGG network is another convolution neural network architecture used for image classification.

TENSOR FLOW

Tensor flow is an open-source software library for high performance numerical computation. It allows simple deployment of computation across a range of platforms (CPUs, GPUs, TPUs) due to its versatile design also from desktops to clusters of servers to mobile and edge devices. Tensor flow was designed and developed by researchers and engineers from the Google Brain team at intervals Google's AI organization, it comes with robust support for machine learning and deep learning and the versatile numerical computation core is used across several alternative scientific domains.

To construct, train and deploy Object Detection Models TensorFlow is used that makes it easy and also it provides a collection of Detection Models pre-trained on the COCO dataset, the Kitti dataset, and the Open Images dataset. One among the numerous Detection Models is that the combination of Single Shot Detector (SSDs) and Mobile Nets architecture that is quick, efficient and doesn't need huge computational capability to accomplish the object Detection.

CONVOLUTIONAL NEURAL NETWORKS (CNN)

The convolutional neural network, or CNN for brief, could also be a specialized kind of neural network model designed for working with two-dimensional image data, although they're going to be used with one-dimensional and three-dimensional data.

Central the convolutional neural network is the convolutional layer that gives the network its name. This layer per forms an operation known as "convolution".

In the context of a convolutional neural network, a convolution may be a linear operation that involves the multiplication of a group of weights with the input, very similar to a standard neural network. as long as the technique was designed for two-dimensional input, the multiplication is performed between an array of input file and a two -dimensional array of weights, called a filter or a kernel.

The filter is smaller than the input file and therefore the before the sort of multiplication applied between a filter-sized patch of the input and the filter may be a scalar product. A scalar product is that the element-wise multiplication between the filter-sized patch of the input and filter, which is then summed, always leading to

ISSN: 2366-1313

one value. Because it leads to 1 value, the operation is conventionally represented and mentioned because the "scalar product".Using a filter smaller than the input is intentional because it allows an equivalent filter (set of weights) to be multiplied by the input array multiple times at distinct points on the input. Specifically, the filter is applied systematically to every overlapping part or filter-sized patch of the input file, left to right, top to bottom.

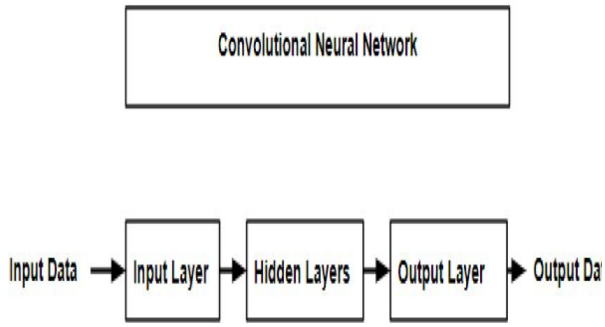

Fig.3 Sample block diagram indicating the flow of image processing using CNN.

This systematic application of an equivalent filter across a picture may be a powerful idea. If the filter is meant to detect a selected sort of feature within the input, then the appliance of that filter systematically across the whole input image allows the filter a chance to get that feature anywhere within the image. This capability is usually represented and mentioned as translation invariance, e.g. the total altogether concern in whether the feature is present instead of where it should had been present.

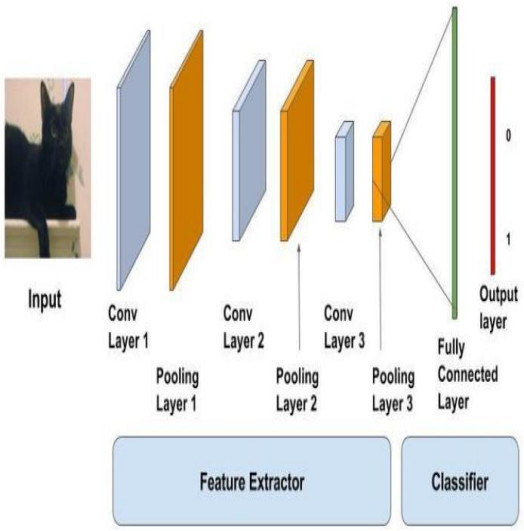

Fig.4 Image classification using CNN

IV. RESULTS

ISSN: 2366-1313

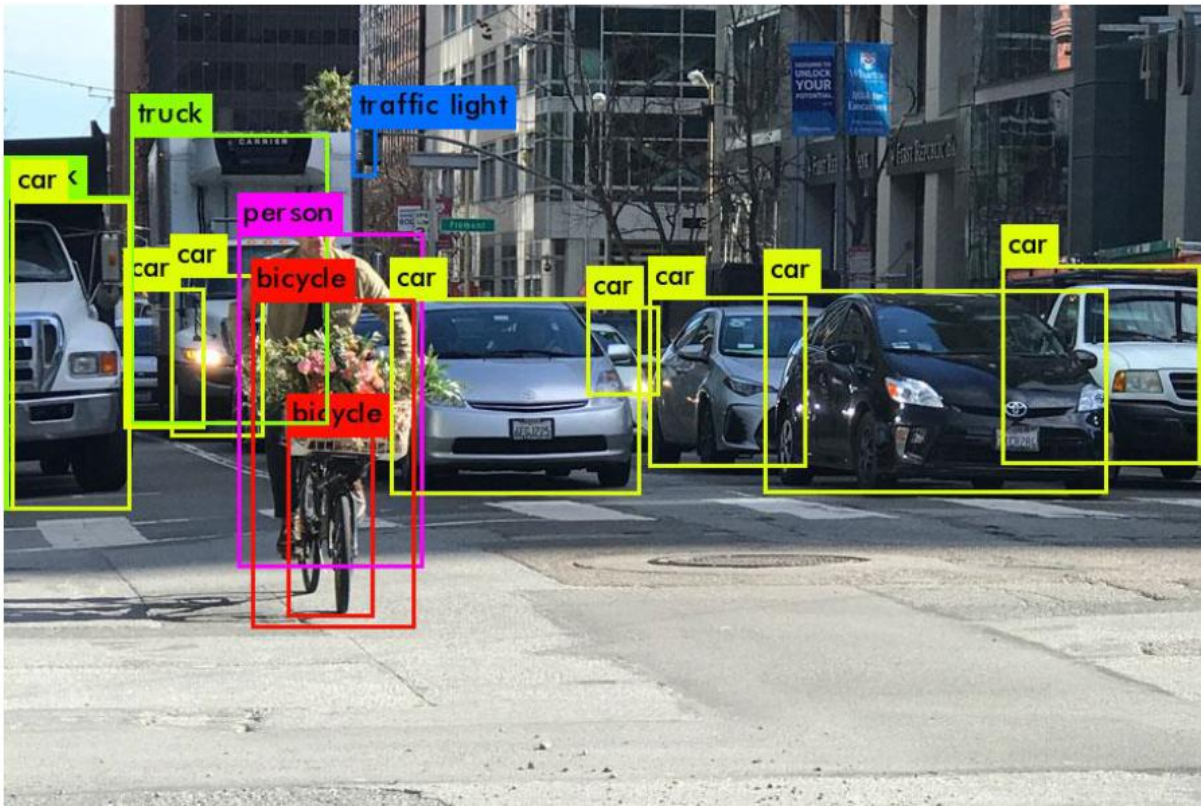

Fig.5 Final result with object detection

V. CONCLUSION

Deep learning-based object detection has been a research hotspot in recent years. This project starts on generic object detection pipelines which provide base architectures for other related tasks. With the help of this the three other common tasks, namely object detection, face detection and pedestrian detection, can be accomplished. Authors accomplished this by combing two things: Object detection with deep learning and OpenCV and Efficient, threaded video streams with OpenCV. The camera sensor noise and lightening condition can change the result as it can create problem in recognizing the object. The end result is a deep learning-based object detector that can process around 6-8 FPS.

Objects are detected using SSD algorithm in real time scenarios. Additionally, SSD have shown results with considerable confidence level. Main Objective of SSD algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained and can further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video

ISSN: 2366-1313

surveillance. This real time analysis of the ecosystem can yield great results by enabling security, order and utility for any enterprise. Further extending the work to detect ammunition and guns in order to trigger alarm in case of terrorist attacks. The model can be deployed in CCTVs, drones and other surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted.

REFERENCES

1. Bruckner, Daniel. Ml-o-scope: a diagnostic visualization system for deep machine learning pipelines. No. UCB/EECS-2014-99.CALIFORNIA UNIV BERKELEY DEPT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES, 2014.
2. K Saleh, Imad, Mehdi Ammi, and Samuel Szoniecky, eds. Challenges of the Internet of Things: Technique, Use, Ethics. John Wiley & Sons, 2018.
3. Petrov, Yordan. Improving object detection by exploiting semantic relations between objects.MS thesis.UniversitatPolitècnica de Catalunya, 2017.
4. Nikouei, SeyedYahya, et al. "Intelligent Surveillance as an Edge Network Service: from Harr-Cascade, SVM to a Lightweight CNN." arXiv preprint arXiv:1805.00331 (2018).
5. Thakar, Kartikey, et al. "Implementation and analysis of template matching for image registration on DevKit- 8500D." Optik-International Journal for Light and Electron Optics 130 (2017): 935-944.
6. Bradski, Gary, and Adrian Kaehler.Learning OpenCV: Computer vision with the OpenCV library." O'Reilly Media, Inc.", 2008.
7. Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).
8. Kong, Tao, et al. "Ron: Reverse connection with objectness prior networks for object detection." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).IEEE, 2017.
9. Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision.Springer, Cham, 2016.
10. Veiga, Francisco José Lopes. "Image Processing for Detection of Vehicles In Motion." (2018).
11. Huaizheng Zhang, Han Hu, GuanyuGao, Yonggang Wen, Kyle Guan, "Deepqoe: A Unified Framework for Learning to Predict Video QoE", Multimedia and Expo

ISSN: 2366-1313

(ICME) 2018 IEEE International Conference on, pp. 1- 6, 2018.