IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

VEHICLE DETECTION SYSTEM & COUNTING OF VEHICLES IN STILL IMAGES USING DEEP LEARNING

[1] Prof. Pallavi Hiwarkar, [2] Damini Bambal, [3] Rishabh Roy
[1] Assistant professor, [2] Student, [3] Student
[1] Department of Computer Science & Engineering,
[1] G H Raisoni University, Saikheda, Madhya Pradesh, India.

Abstract: Vehicle detection and counting are turning out to be progressively significant in the field of Highway management. In any case, because of the various sizes of vehicles, their identification stays a test that straight forwardly influences the exactness of vehicle counts. To resolve this issue, we propose a vehicle detection system and counting framework. In the proposed vehicle detection and counting system we uses Yolov3 for vehicle detection and counting of vehicles from still images that can detect, classify and count numerous vehicles from CCTV footage. This system is implemented in Google Colab and any recorded video of Vehicles is used to test it. The execution time is 5 minutes, with 99% accuracy.

INTRODUCTION

Vehicle identification and measurements in roadway observing video scenes are of impressive importance to astute traffic the board and control of the expressway. With the famous establishment of traffic reconnaissance cameras, a tremendous information base of traffic video film has been acquired for examination. For the most part, at a high survey point, a more-far off street surface can be thought of. The item size of the vehicle changes significantly at this survey point, and the location exactness of a little article far away from the street is low. Not with standing perplexing camera scenes, it is fundamental to actually take care of the above issues and further apply them. In this article, we center on the above issues to propose a reasonable arrangement, and we apply the vehicle identification results to multi-object following and vehicle counting.

Related work on vehicle detection

As of now, vision-based vehicle object discovery is partitioned into conventional machine vision techniques and complex profound learning strategies. Customary machine vision techniques utilize the movement of a vehicle to isolate it from a decent foundation picture. This technique can be separated into three classifications: the strategy for utilizing foundation deduction , the strategy for utilizing persistent video outline contrast , and the technique for utilizing optical stream . Utilizing the video outline contrast technique, the fluctuation is determined by the pixel upsides of a few successive video outlines. In addition, the moving forefront area is isolated by the edge . By utilizing this strategy and stifling commotion, the halting of the vehicle can likewise be identified. At the point when the foundation picture in the video is fixed, the foundation data is utilized to lay out the foundation model. Then, at that point, each edge picture is contrasted and the foundation model and the moving item can likewise be divided. The strategy for involving optical stream can distinguish the movement district in the video. The produced optical stream field addresses every pixel's bearing of movement and pixel speed. Vehicle location strategies utilizing vehicle highlights, like the Scale Invariant Feature Transform (SIFT) and Speeded up Robust Features (SURF) techniques, have been generally utilized. For instance, 3D models have been utilized to finish vehicle location and grouping errands. Utilizing the relationship bends of 3D edges on the external surface of the vehicle, the vehicles are partitioned into three classifications: vehicles, SUVs, and minibuses.
The utilization of profound convolutional networks (CNNs) has made astounding progress in the field of vehicle object identification. CNNs have a solid capacity to learn picture includes and can play out various related assignments, for example, grouping and jumping box relapse. The discovery strategy can be for the most part isolated into two classifications. The two-stage technique creates a competitor box of the article by means of different calculations and afterward characterizes the item by a convolutional neural organization. The one-stage strategy doesn't create an applicant box however straightforwardly changes over the situating issue of the article jumping enclose to a relapse issue for handling. In the two-stage strategy, Region-CNN (R-CNN) utilizes specific locale search in the picture. The picture contribution to the convolutional network should be fixed-size, and the further design of the organization requires a long preparation time and consumes a lot of capacity memory. Drawing on the possibility of spatial pyramid coordinating, SPP NET permits the organization to enter pictures of different sizes and to have fixed results. R-FCN, FPN, and Mask RCNN have further developed the element extraction techniques; include determination, and arrangement capacities of convolutional networks in various ways. Among the one-stage strategies, the most significant are the Single Shot Multibox Detector (SSD) and You Only Look Once (YOLO) structures. The Multibox , Region Proposal Network

(RPN) and multi-scale portrayal techniques are utilized in SSD, which utilizes a default set of anchor boxes with various angle proportions to all the more precisely position the article. Dissimilar to SSD, the YOLO network isolates the picture into a decent number of lattices. Every framework is answerable for anticipating objects whose middle focuses are inside the network. YOLOv2 added the BN (Batch Normalization) layer, which causes the organization to standardize the contribution of each layer and speed up the organization assembly speed. YOLOv2 utilizes a multi-scale preparing technique to haphazardly choose another picture size for each ten clusters. Our vehicle object identification utilizes the YOLOv3 network. In light of YOLOv2, YOLOv3 involves strategic relapse for the item classification. The classification misfortune strategy is two-class cross-entropy misfortune, which can deal with different name issues for a similar article. Also, calculated relapse is utilized to relapse the container certainty to decide whether the IOU of the deduced box and the real box is more prominent than 0.5. Assuming that more than one need box fulfills the condition, hands down the biggest earlier box of the IOU is taken. In the last item expectation, YOLOv3 utilizes three unique scales to anticipate the article in the picture. The conventional machine vision technique has a quicker speed while recognizing the vehicle however doesn't deliver a decent outcome when the picture changes in splendor, there is occasional movement behind the scenes, and where there are sluggish vehicles or complex scenes. Progressed CNN has accomplished great outcomes in object recognition; be that as it may, CNN is delicate to scale changes in object identification .The one phase strategy utilizes lattices to anticipate objects, and the matrix's spatial requirements make it difficult to have higher accuracy with the two-stage approach, particularly for little items. The two phase strategy utilizes area of interest pooling to section up-and-comer locales into blocks as per given boundaries, and in the event that the up-and-comer district is more modest than the size of the given boundaries, the up-and-comer district is cushioned to the size of the given boundaries. Thusly, the trademark construction of a little item is obliterated and its identification precision is low. The current strategies don't recognize if enormous and little articles have a place with a similar classification. A similar technique is utilized to manage a similar sort of article, which will likewise prompt mistaken recognition. The utilization of picture pyramids or multi-scale input pictures can tackle the above issues, albeit the estimation necessities are huge.

Vehicle location research in Europe

Vision-based vehicle location techniques in Europe have accomplished plentiful outcomes. In, between the "Hofolding" and "Weyern" areas of the A8 motorway in Munich, Germany, the Multivariate Alteration Detection (MAD) technique was utilized to distinguish the difference in two pictures with a brief time frame slack. The moving vehicles are featured in a change picture, which is utilized to assess the vehicle thickness of the street. In, utilizing the motorways A95 and A96 close to Munich, the A4 close to Dresden, and the "Mittlere Ring" in Munich as the test conditions, the Canny edge calculation is applied to the street picture, and the histogram of the edge steepness is determined. Then, at that point, utilizing the k-implies calculation, the edge steepness measurements are isolated into three sections, and a shut vehicle model is distinguished in view of the steepness. A differentiation based methodology was utilized to make a shading model to distinguish and eliminate vehicle shadow regions, which kills obstruction brought about by development in the scene. Subsequent to dispensing with the shadow region, the vehicle recognition execution can be fundamentally gotten to the next level. The analysis in was led on Italian and French thruways. The HOG and Haar-like elements were looked at in, and the two highlights were converged to develop a locator for vehicle discovery that was tried on French vehicle pictures. Notwithstanding, when the above strategy is utilized for vehicle recognition, the sort of vehicle can't be identified. Moreover, when the brightening is deficient, it is hard to separate the edge of the vehicle or identify the moving vehicle, which brings on some issues in low vehicle recognition precision and influences the location results for additional utilization. Pictures of elevated view points were utilized by yet can't obviously catch the qualities of every vehicle and produce bogus vehicle discoveries. In any case, with the improvement of profound learning innovation, vehicle recognition in view of CNN has been effectively applied in Europe. In, Fast R-CNN was utilized for vehicle identification in rush hour gridlock scenes in the city of Karlsruhe, Germany. Quick R-CNN utilizes a particular pursuit methodology to observe all up-and-comer outlines, which is remarkably tedious, and the vehicle identification speed is slow.
So, research on vision-based vehicle location is as yet advancing, and significant difficulties are progressively being survived, which will make a critical commitment to the improvement of European traffic development.

LITERATURE SURVEY

1. Vision-based vehicle detection and counting system using deep learning in highway scenes- They proposes a vision-based vehicle detection and counting system in which the highway road surface in the image is first extracted and divided into a remote area and a proximal area by a newly proposed segmentation method
2. Vehicle detection and recognition- The surveillance system includes detection of moving vehicles and recognizing them, counting number of vehicles and verification of their permit with the organization.
3. Video-Based Vehicle Counting for Expressway- vehicle counting method is designed based on the tracking results, in which the driving direction information of the vehicle is added in the counting process.

THE FRAMEWORK STRUCTURE

This part depicts the fundamental construction of the vehicle recognition and counting framework. To start with, the video information of the traffic scene are placed. Then, at that point, the street surface region is extricated and isolated. The YOLOv3 profound learning object identification technique is utilized to identify the vehicle object in the interstate rush hour gridlock scene. At long last, ORB include extraction is performed on the distinguished vehicle box to finish multi-object following and get vehicle traffic data.
As per figure, the street surface division technique is utilized to separate the street region of the thruway. The street region is partitioned into two sections in view of the position where the camera is raised, a distant region and a proximal region. Then, at that point, the vehicles in the two street regions are identified utilizing the YOLOv3 object identification calculation. This calculation can further develop the little item recognition impact and tackle the issue that the article is hard to recognize because of the sharp difference in the article scale. The ORB calculation is then utilized for multi-object following. The ORB calculation removes the identified box's highlights and matches them to accomplish connection between's a similar article and different video

outlines. At long last, traffic insights are determined. The direction produced by the article following is created, the vehicle driving not set in stone, and traffic data, for example, the quantity of vehicles in every classification is gathered. This framework works on the precision of article identification from the roadway observation video viewpoint and develops a discovery following and traffic data obtaining plan inside the full field of the camera view.

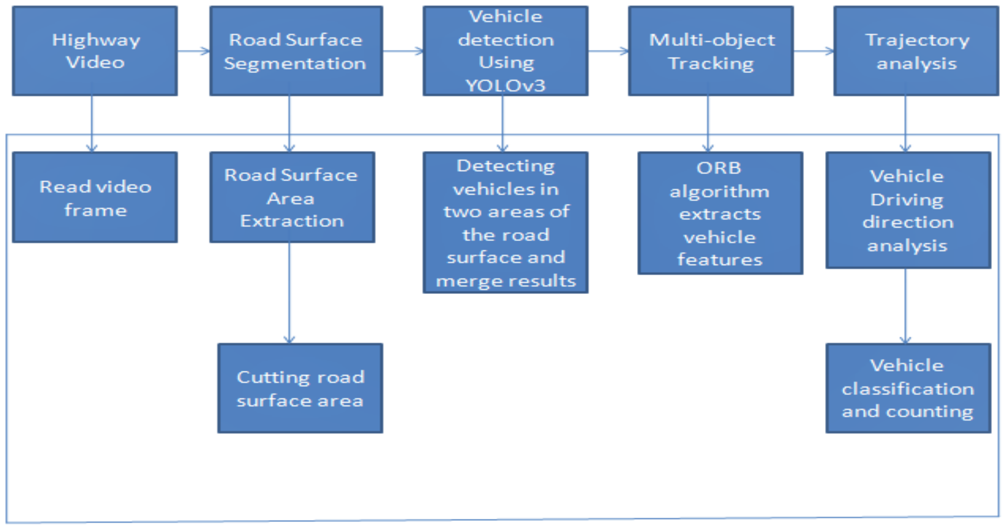


Fig. The Framework Structure

METHODOLOGY

1. Street surface division -This part portrays the technique for roadway street surface extraction and division. We executed surface extraction and division utilizing picture handling strategies, for example, Gaussian combination demonstrating, which empowers better vehicle recognition results while utilizing the profound learning object identification technique. The expressway video picture observation enormous field of view. The vehicle is the focal point of consideration in this review, and the district of interest in the picture is accordingly the roadway street surface region.Simultaneously, as per the camera's shooting point, the street surface region is amassed in a particular scope of the picture. With this element, we could separate the thruway street surface regions in the video. The course of street surface extraction is shown

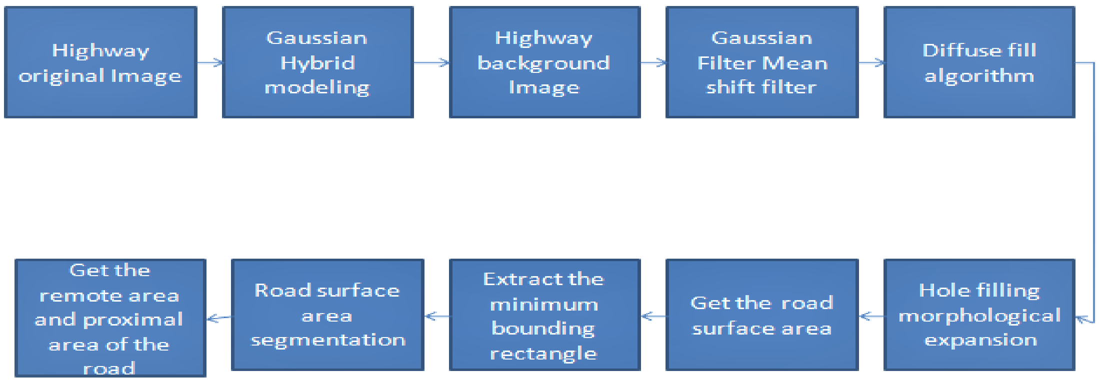


Fig. Street surface division

As displayed in Fig.1 to kill the impact of vehicles out and about region division, we utilized the Gaussian blend displaying strategy to extricate the foundation in the initial 500 casings of the video. The worth of the pixel in the picture is Gaussian around a specific focal worth in a specific time range, and every pixel in each casing of the picture is counted. Assuming the pixel is a long way from the middle, the pixel has a place with the closer view. In the event that the worth of the pixel point strays from the middle worth inside a specific change, the pixel point is considered to have a place with the foundation. The blended Gaussian model is particularly helpful in pictures where foundation pixels have multi-top attributes, for example, the roadway reconnaissance pictures utilized in this review.

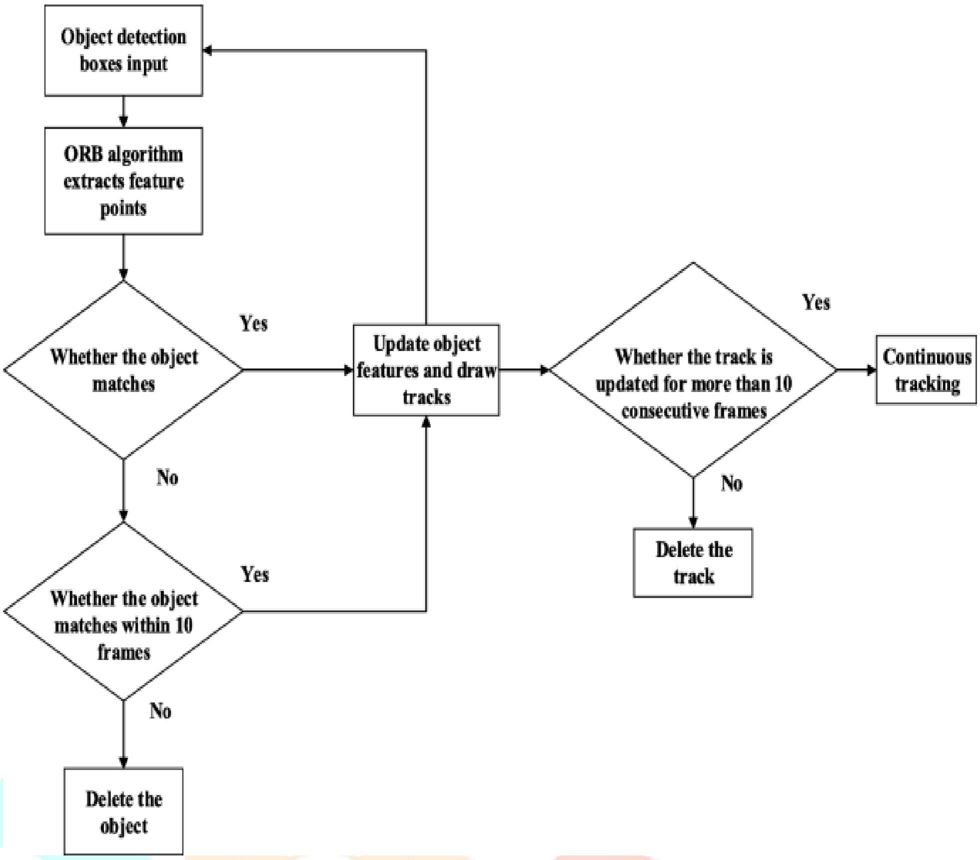
After extraction, the foundation picture is smoothed by a Gaussian channel with a 3*3 piece. The Mean Shift calculation is utilized to smooth the shade of the info picture, kill the shading with a comparative shading circulation, and disintegrate the shading region with a more modest region. On this premise, the flooding filling calculation is utilized to isolate the street surface region. The flooding filling calculation chooses a point in the street surface region as a seed point and fills the neighboring nonstop street surface regions with the pixel worth of the seed point. The pixel worth of the neighboring nonstop street surface regions is near the seed point pixel esteem. At last, the opening filling & morphological development activities are performed to all the more totally separate the street surface. We extricated the street surfaces of various roadway scenes.

After that we portioned the street surface region to give precise contribution to resulting vehicle discovery. For the extricated street surface picture, a base encompassed square shape is created for the picture without revolution. The handled picture is separated into five equivalent parts, the 1/5 region neighboring the beginning of the direction hub is characterized as the close to far off region of the street surface, and the excess 4/5 region is characterized as the close to proximal region of the street surface. The close to proximal region and the close to distant region cross-over by 100 pixels address the issue that the vehicle in the picture might be partitioned into two sections by the above methodology. The pixel upsides of the close to proximal region and the close to far off region are looked through segment by section. Assuming that the pixel values in the segment are every one of the zero, the picture of the segment is all dark and isn't the street surface region; it is then erased. After the not-street surface regions are rejected, the saved regions are called distant regions and proximal region of the street surface.

2. Vehicle identification utilizing YOLOv3- This segment portrays the item location strategies utilized in this review. The execution of the roadway vehicle identification system utilized the YOLOv3 organization. The YOLOv3 calculation proceeds with the fundamental thought of the initial two ages of YOLO calculations. The convolutional brain network is utilized to extricate the highlights of the information picture. As per the size of the component map, for example, 13*13, the info picture is partitioned into13*13 networks. The focal point of the article mark confine is a framework unit, and the network unit is answerable for anticipating the item. The organization structure embraced by YOLOv3 is called Darknet-53. This design embraces the full convolution technique and replaces the past rendition of the direct-associated convolutional brain network with the lingering structure. The branch is utilized to straightforwardly associate the contribution to the profound layer of the organization direct learning of residuals guarantees the uprightness of picture include data, improves on the intricacy of preparing, and further develops the general discovery precision of the net-work. In YOLOv3, every framework unit will have three bouncing boxes of various scales for one item. The competitor box that has the biggest covering region with the commented on box will be the last expectation result. Also, the YOLOv3 network has three result scales, and the three scale branches are in the end consolidated. Shallow elements are utilized to identify little items, and profound highlights are utilized to distinguish enormous articles; the organization can accordingly recognize objects with scale changes. The identification speed is quick, and the recognition precision is high. Traffic scenes taken by parkway observation video have great versatility to the YOLOv3 organization .The organization will at long last result the directions, certainty, and classification of the article. while utilizing YOLO location, pictures are resized to a similar size, for example, 416*416, when they are shipped off the organization. Since the picture is fragmented, the size of the distant street surface becomes disfigured and bigger. Subsequently, more element points of a little vehicle object can be gained to stay away from the deficiency of an item includes because of the vehicle object being excessively little. The dataset introduced in "Vehicle dataset" segment is set into theYOLOv3 network for preparing, and the vehicle object identification model is acquired. The vehicle object recognition model can recognize three sorts of vehicles: vehicles, transports, and trucks . Since there are not many cruisers on the thruway, they were excluded from our identification. The distant region and proximal region of the street surface are shipped off the organization for identification. The distinguished vehicle box places of the two regions are planned back to the first picture, and the right item position is gotten in the first picture. Utilizing the vehicle object recognition strategy for getting the classification and area of the vehicle can give important information to protest following. The above data is adequate for vehicle counting, and the vehicle location strategy hence doesn't recognize the particular attributes of the vehicle or the state of the vehicle.

3. Multi-object following-This segment depicts how different articles are followed in light of the item enclose recognized "Vehicle recognition utilizing YOLOv3" segment. In this review, the ORB calculation was utilized to remove the highlights of the recognized vehicles, and great outcomes were acquired. The ORB calculation shows unrivaled execution as far as computational execution and matching expenses. This calculation is a great option in contrast to the SIFT and SURF picture portrayal calculations. The ORB calculation utilizes the Features From Accelerated Segment Test (FAST) to identify include focuses and afterward utilizes the Harris administrator to perform corner discovery. In the wake of acquiring the element focuses, the descriptor is determined utilizing the BRIEF calculation. The direction framework is laid out by taking the element point as the focal point of the circle and involving the centroid of the point locale as the x-pivot of the direction framework. Along these lines, when the picture is turned, the direction framework can be pivoted by the revolution of the picture, and the element point descriptor hence has turn consistency. Whenever the image point is changed, a reliable point can likewise be proposed. Subsequent to getting the paired element point descriptor, the XOR activity is utilized to match the component focuses, which works on the matching proficiency.

The following system is displayed in Fig. Whenever the quantity of matching focuses got is more prominent than the set limit, the point is viewed as effectively coordinated and the matching box of the item is drawn. The wellspring of the expectation box is as per the following: include point cleaning is performed utilizing the RANSAC calculation, which can avoid the wrong commotion points of the matching blunders, and the homography grid is assessed. As indicated by the assessed homography grid and the place of the first article discovery box, a point of view change is performed to get a relating expectation box.



We utilized the ORB calculation to extricate highlight focuses in the item recognition box got by the vehicle discovery calculation. The article include extraction isn't performed from the whole street surface region, which drastically lessens how much computation. In object following, the expectation box of the article in the following edge is drawn since the difference in the vehicle object in the persistent casing of the video is unobtrusive as indicated by the ORB include separated in the item box. Assuming the forecast box and the identification box of the following casing meet the most brief distance necessity of the middle point, a similar article effectively matches between the two casings. We characterize an edge T that alludes to the most extreme pixel distance of the recognized focus point of the vehicle object box, which moves between two neighboring video outlines. The positional development of similar vehicle in the adjoining two casings is not exactly the edge T. Along these lines, when the middle place of the vehicle object confine moves over T the two nearby edges, the vehicles in the two edges are not something very similar, and the information affiliation comes up short. Considering the scale change during the development of the vehicle, the worth of the edge T is connected with the size of the vehicle object box. Different vehicle object boxes have various edges. This definition can address the issues of vehicle development and different info video sizes.
We erase the direction that isn't refreshed for ten back to back outlines, which is reasonable for the camera scene with a wide-point of picture assortment on the expressway under study. In this kind of scene, the street surface caught by the camera is far off. In ten back to back video outlines, the vehicle will move farther away. Hence, when the direction isn't refreshed for ten casings, the direction is erased. Simultaneously, the vehicle direction and the location line will just cross once, and the edge setting in this manner doesn't influence the last counting result. Assuming the Forecast enclose neglects to match successive casings, the item is viewed as missing from the video scene, and the expectation box is erased. From the above interaction, the worldwide article location results and following directions from the total expressway observing video point of view are gotten.

4. Direction examination
This segment portrays the investigation of the directions of moving items and the counting of different article traffic data.

EXPERIMENTATION

We use opencv with pre-prepared YOLOv3, there are a couple reasons. We might need to involve opencv for YOLO:
a] Data sets used for proposed method - we have taken CCTV footage of any highway. It consists of annotated images with classes of objects.
b]Implementation details – this work is carried out in Google Colab. Keras is used to train the image processing network & opencv.
c]Framework used - YOLOv3 (You Only Look Once, Version 3) is a real time object detection algorithm identifies specific objects in videos, live feeds, or images. YOLO uses features learned by a deep convolutional neural network to detect an object. Versions 1-3 of YOLO were created by Joseph Redmon and Ali Farhadi. The first version of YOLO was created in 2016, and version 3, which is discussed extensively in this article, was made two years later in 2018. YOLOv3 is an improved version of YOLO and YOLOv2 .YOLO is implemented using the Keras or OpenCV deep learning libraries. For example, in a live feed of traffic, YOLO can be used to detect different kinds of vehicles depending on which regions of the video score highly in comparison to predefined classes of vehicles.

RESULT & ANALYSIS

The algorithm, Yolo Version 3,The model can identify the cars, buses, bicycle, motorcycle, trucks & pedestrian. In the figure shows that the model can identify the cars, trucks, buses, bicycle, motorcycle & pedestrians.(Fig.1)

Fig.1

Total detection time: - The time is taken to detect the testing data with the manually created YOLO weights for both the algorithms. The total time obtained when we run the code is 7 seconds

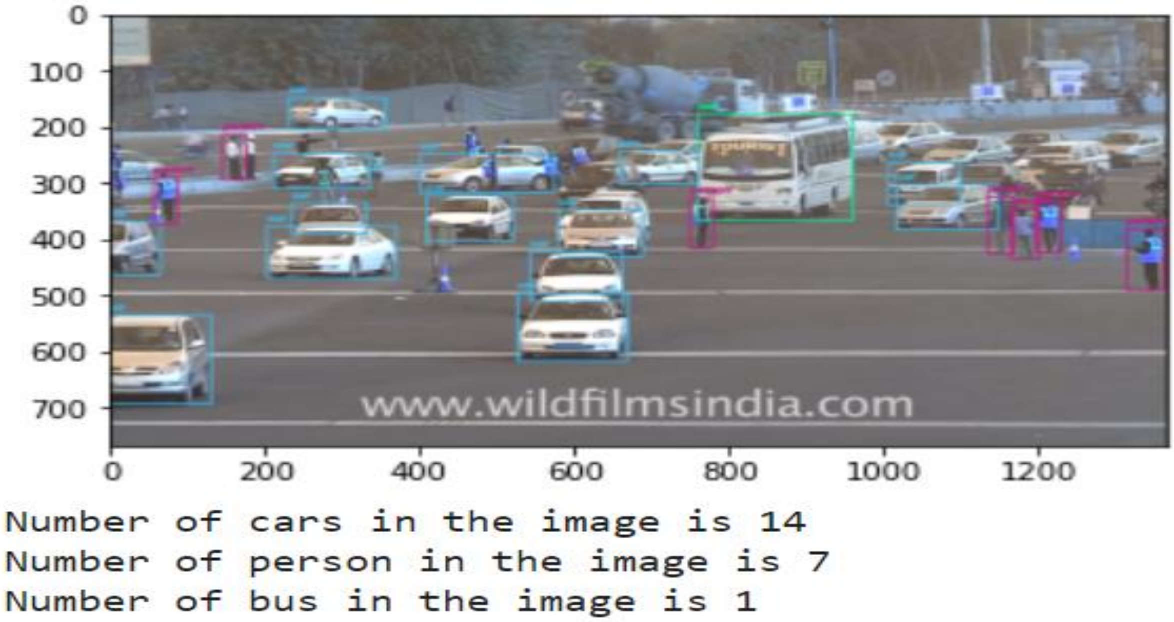


Fig.2

In this figure shows that the model can count number of cars, person & buses from a given input video image.(Fig.2)

The accuracy which we get after detecting the vehicle from CCTV Footage is 98% & the counting of vehicle from still image is 99%.

CONCLUSION:

In this task, we have seen that in the video of vehicle detection framework, our model is making boundary boxes outside the vehicle and gives the name and exactness of vehicle as well as showing us counting of vehicle and people on foot from an information picture without the requirement for costly preprocessing or costly profound evalutions; YOLOv3 as higher precision. In view of exploratory outcomes we can distinguish vehicle and walkers all the more definitively and recognize the article independently with precise area of an item.

REFERENCES:

[1]Vision-based vehicle detection and counting system using deep learning in highway scenes Huanshen Song, Haoxiang Liang, Huaiyu Li, Zhe Dai & Xu Yun 51 (2019)

[2] Dennis Hein. “Traffic Light Detection with Convolutional Neural Networks and 2D Camera Data”. PhD thesis. fu-berlin, 2020.

[3] Ablajan Sulaiman. “Image, Video and Real-Time Webcam Object Detection & Instance Segmentation using Mask R-CNN”. In: (2020). url: https : / / medium . com / @toarches / image - video - and - real - time webcam-object-detection-and-instance-segmentation-with-mask-rcnn-37a4675dcb49.

[4] Li, Y.; Guo, J.; Guo, X.; Liu, K.; Zhao, W.; Luo, Y.; Wang, Z. A Novel Target Detection Method of the Unmanned Surface Vehicle under All-Weather Conditions with an Improved YOLOV3. Sensors 2020, 20, 4885. [CrossRef]

[5] Chen, X.Z.; Chang, C.M.; Yu, C.W.; Chen, Y.L. A Real-Time Vehicle Detection System under Various Bad Weather Conditions Based on a Deep Learning Model without Retraining. Sensors 2020, 20, 5731. [CrossRef] [PubMed]

[6] Shakil, S., Rajjak, A., Kureshi, A.K. (2020). Object detection and tracking using YOLO v3 framework for increased resolution video. International Journal of Innovative Technology and Exploring Engineering (IJITEE)

[7] Liu, L., Ouyang, W.L., Wang, X.G., Fieguth, P., Chen, J., Liu, X.W., Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. International Journal of Computer Vision, 128(2): 261-318. https://doi.org/10.1007/s11263-019-01247-4

[8] Cui, S., Zhou, Y., Wang, Y., Zhai, L. (2020). Fish detection using deep learning. Applied Computational Intelligence and Soft Computing, 2020: 3738108. https://doi.org/10.1155/2020/3738108

[9] Chethan Kumar B, Punitha R, and Mohana, “YOLOv3 and YOLOv4: Multiple Object Detection for Surveillance Applications” Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020) IEEE Xplore Part Number: CFP20P17- ART; ISBN: 978-1-7281-5821-1

[10] Hassan, N. I., Tahir, N. M., Zaman, F. H. K., & Hashim, H, “People Detection System Using YOLOv3 Algorithm” 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE). doi:10.1109/iccsce50387.2020.9204925

[11] Viraf, “Master the COCO Dataset for Semantic Image Segmentation”, May 2020.

[12] Joseph Redmon, Ali Farhadi, “YOLOv3: An Incremental Improvement”, University of Washington.

[13] Karlijn Alderliesten, “YOLOv3 — Real-time object detection”, May 28 2020.

[14] Arka Prava Jana, Abhiraj Biswas, Mohana, “YOLO based Detection and Classification of Objects in video records” 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India.

[15] Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, “Object Tracking Algorithms for video surveillance applications” International conference on communication and signal processing (ICCSP), India, 2018, pp. 0676-0680.