Machine Learning based Object Detection and Classification using Drone

Sugam Dembla[1], Niyati Dolas[2], Ashish Karigar[3], Dr. Santosh Sonavane[4]

[1]Student, Symbiosis Skills and Professional University, Pune, Maharashtra, India
[2]Student, Symbiosis Skills and Professional University, Pune, Maharashtra, India
[3]Student, Symbiosis Skills and Professional University, Pune, Maharashtra, India
[4]Director, School of Mechatronics, Symbiosis Skills and Professional University, Pune, Maharashtra, India

---***---

Abstract - *The issue of safety and surveillance has been in debate for ages as it comes under our day-to-day life struggle. The reasons for these issues are violation and negligence of mandatory rules and regulations. Violation of traffic rules, safety norms have resulted in many unfortunate situations such as life loss. Apart from this, there are instances of Breaking of Traffic Signals, Road Accidents Emergencies, Burglaries, Shootings, and Explosions. There has been a rise in the criminal record due to a lack of evidence. Updating ourselves with technology can help in demolishing the above issues. Our project works on the theme suggested above, using a collaboration of mechanics and computer vision to solve pressing issues. Our review works on dealing with these issues and how we could solve them using machine learning models, by reviewing different object detection and classification models and various mechanical models. We will be trying to design a video surveillance system that combines three phases of data processing: moving object extraction, moving object recognition and tracking, and decisions about actions which will direct on automatic identification of events of interest which comes under Event Recognition or Anomaly Detection, Object Detection, Classification and Tracking of moving vehicles for vehicle number detection and Safety Gear Detection like Helmet while driving and Workers for Safety vest and boot on a construction site and various such aspects.*

Key Words: Classification, Computer Vision, Key word3, Key word3, Key word3

1.INTRODUCTION

In today's globalized world, there have been still usage of old practices regarding safety and surveillance. With the modernizing world, there is a need to update our surveillance methods. Using the latest methods of machine learning-based object detection and classification using a drone has been listed down. We aimed to assess the effectiveness of how this concept can change the age-old methods. Learning is the main hallmark of human intelligence and, therefore, the basics suggest that to get information. Machine learning is a fundamental way to make a computer intelligent. Machine learning is a study where we can train our computers to carry out human simulation activities. We can acquire an accurate prediction of outcomes without being explicitly programmed. Joining hands with computer vision techniques for interest can be helpful. As we know, computers can gain a high level of understanding from digital images or videos, surveillance devices like surveillance cameras which have been used in day-to-day life. These devices have not been very efficient in evidence collection or similar means. As there has been an increase in various accidents due to violation of safety norms and regulations. As our topic works on identifying these violations and trying to raise the alarm for them, violations of traffic rules are one of the pressing issues, this issue is the reason for accidents [1]. A survey conducted by Fact Checker showed that violation of traffic rules such as over-speeding, driving on the wrong side, drink driving, use of mobile phones, and jumping a red light caused 80% (117,914) of all road accident deaths (147,913) in a single or approximately 323 deaths every day. As our review works on safety and surveillance issues, due to lack of evidence there have been many pending cases in India [2]. Among the pending criminal cases, nearly 36% of the cases are pending at the trial stages due to Lack of Evidence, Arguments & Judgement. Approximately 61% at this stage are pending due to a lack of substantial evidence [32]. Another department that takes the lead in violation of safety norms in the construction industry. About 38 Fatal accidents take place every day and around 48,000 workers die in India due to occupational accidents in the construction sector.

Our idea of reviewing this topic of how Machine learning using object detection and classification can be useful to limit the above issues.

Machine learning consists of various algorithms, identifying the model with higher accuracy and training the model with a selected model which will help in identifying the objects of the face/human or anomaly events and classifying these objects, people, or events according to the created dataset and raising an alarm if there is a violation of regulations.

As previously noted, it will consist of mechanical aspects also, as will be read ahead about the selection of drone and why not any other means.

2. EXISTING OBJECT DETECTION & CLASSIFICATION MODELS

2.1 OverFeat (2013)

Inspired by the first success of AlexNet within the 2012 ImageNet competition, wherever CNN-based feature extraction defeated all handmade feature extractors, OverFeat quickly introduced CNN back to the thing detection space still. The thought is incredibly straight forward: if we can classify one image victimization CNN, what regarding covetously scrolling through the total image with completely different sizes of windows, and check out to regress and classify them one-by-one employing a CNN? This leverages the facility of CNN for feature extraction and classification, and conjointly bypasses the arduous region proposal drawback by pre-defined slippery windows. Also, since a close-by convolution kernel will share a part of the computation result, it's not necessary to reckon convolutions for the overlapping space, thus reducing value heaps. OverFeat could be a pioneer within the one-stage object detector. It tried to mix feature extraction, location regression, and region classification within the same CNN.

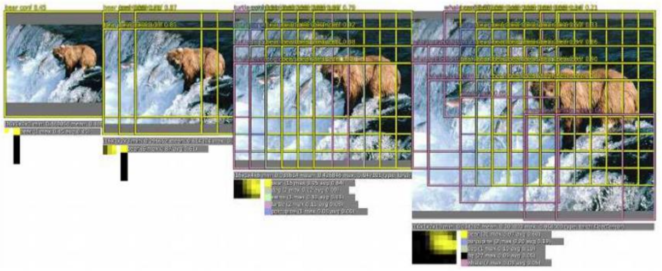

Fig -1: "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks" [3]

2.2 R-CNN (2014)

R-CNN also came out and proposed in 2014, R-CNN came out late compared with OverFeat. Nevertheless, region-based CNN approach in the long run led to a big impact on the field of object detection and classification research due to its two-stage framework, which comprises - first region proposal stage, and second region classification & refinement stage. Selective Search is the technique R-CNN uses by first extracting potential regions of interest from an input image, video, or any live feed. Selective search does not try to process the foreground object, but instead, it collects and makes groups of similar pixels i.e., images which have similar pixels will most probably belong to the same object in the image. Thus, the output of a selective search in R-CNN has a very high probability to contain something of interest and meaning. Next, what R-CNN tried to do is it surrounds these images and video region proposals into fixed-size images and videos with some paddings and then tries to feed these images and videos into the second stage of the network for more fine-grained recognition.

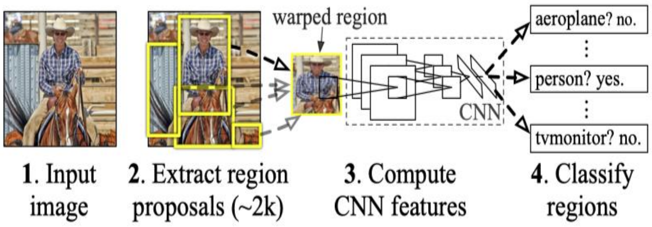

Fig -2: "Region-based Convolutional Networks for Accurate Object Detection and Segmentation" [4]

2.3 Resnet (2015)

Residual neural network (ResNet) was proposed in 2015 which is a type of Artificial Neural Network (ANN) that builds on constructs known from pyramidal cells within the neural structure within the brain. Residual neural networks (ResNet) does this by using shortcuts or skip connections, to skip and jump over some layers of the neural network. A Typical Residual Neural Network (ResNet) model contains non-linearities which is known as (ReLU) which is implemented with double or triple layer skips and batch normalization in between. By using this technique an additional matrix is generated containing weights that may be used to learn the skip weights, and these models are known as HighwayNets. DenseNets are the models with several parallel skips. In the context of Residual Neural Networks (ResNet), a non-residual network could also be described as a clear network.

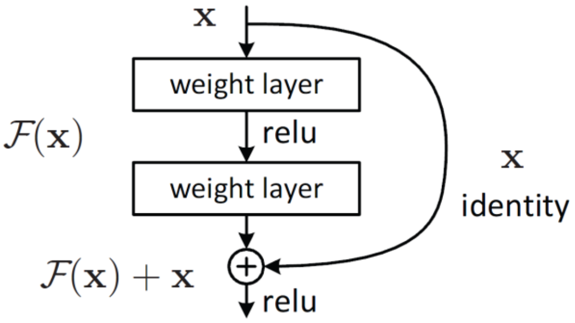
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Fig -3: "Deep Residual Learning for Image Recognition" [5]

2.4 Fast-RCNN (2015)

A quick follow-up for R-CNN is to scale back the duplicate convolution over multiple region proposals. Since these region proposals all come from one image, it's naturally to enhance R-CNN by running CNN over the whole image once and share the computation among many region proposals. However, different region proposals have different sizes, which also result in different output feature map sizes if we are using the same CNN feature extractor. These feature maps with various sizes will prevent us from using fully connected layers for further classification and regression because the FC layer only works with a fixed size input. With a shared feature extractor and thus the scale-invariant ROI pooling layer, Fast R-CNN can reach a uniform localization accuracy but having 10~20x faster training and 100~200x faster inference. The near real-time inference and a neater E2E training protocol for the detection part make Fast R-CNN a well-liked choice within the industry also.

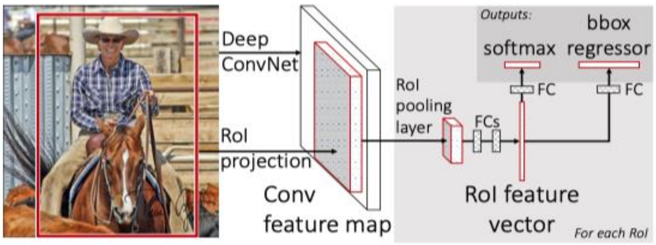

Fig -4: "Fast R-CNN" [6]

2.5 Faster-RCNN (2015)

2.5 Faster-RCNN (2015) - In early 2015, Ross Girshick proposed an improved version of R-CNN called Fast R-CNN. Just a few months later, Ross and his team came back with another improvement again. This new network Faster R-CNN is not only faster than previous versions but also marks a milestone for object detection with a deep learning method. Although the sliding window has a fixed size, our objects may appear on different scales. Therefore, Faster R-CNN introduced a technique called anchor box. Anchor boxes are pre-defined prior boxes with different aspect ratios and sizes but share the same central location. In Faster R-CNN there are k=9 anchors for each sliding window location, which covers 3 aspect ratios for 3 scales each. These repeated anchor boxes over different scales bring nice translation-invariance and scale-invariance features to the network while sharing outputs of the same feature map. Note that the bounding box regression will be computed from these anchor box instead of the whole image.

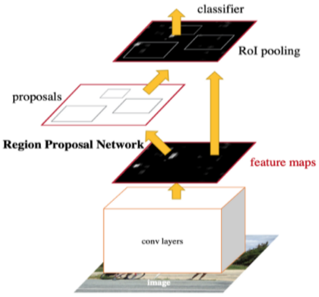

Fig -5: "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" [7]

2.6 YOLO v1 (2015)

The basic and simplest idea is to separate an input image or a video into an SxS grid and have each cell directly regress the bounding box location and thus probability of correct score if the item center falls into that cell. Because objects may have different sizes, there'll be over one bounding box regressor per cell. During training, the regressor with the absolute best IOU are getting to be assigned to match with the ground-truth label, so regressors at the same location will learn to handle different scales over time. Meanwhile, each cell of the model will try to predict C class probabilities, conditioning on the grid cell of the model containing an object with high confidence score. This way of approach is later described as dense predictions because YOLO tried to predict classes and bounding boxes for all possible locations during a picture.

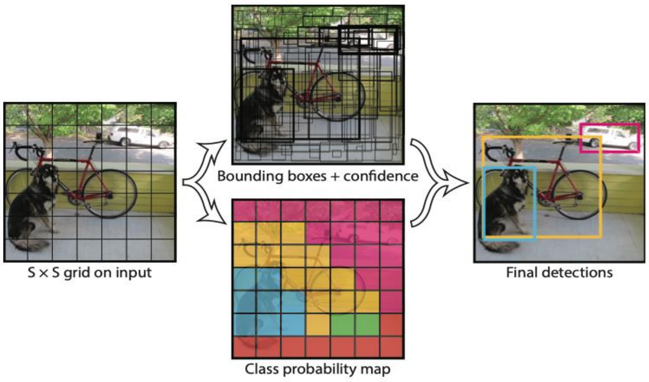

Fig -6: "You Only Look Once: Unified, Real-Time Object Detection" [8]

2.7 YOLO v2 (2016)

Joseph Redmon, on the opposite hand, was also busy improving his one-stage YOLO detector. The initial version of YOLO suffers from many shortcomings: predictions supported a rough grid brought lower localization accuracy, two scale-agnostic regressors per grid cell also made it difficult to acknowledge small packed objects. In 2015, we came across many state-of-the-art and great innovations in object detection & classification and in computer vision areas. YOLO v2 just has to search out one and only because of integrating all of them to become better, faster, and stronger. To stabilize early training YOLO v2 tries to constraint the center of the object regression tx and ty in that particular grid cell rather than predicting offsets to anchor boxes, to boost the detection of small objects, YOLO v2 added a passthrough layer to merge features from an early layer.

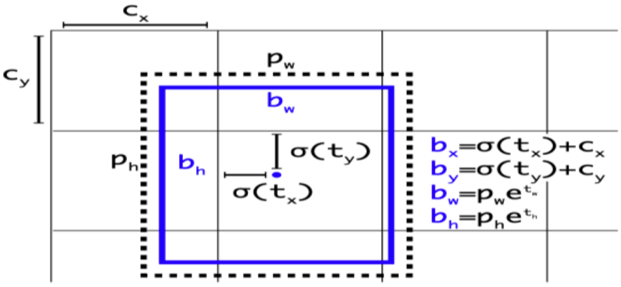

Fig -7: "YOLO9000: Better, Faster, Stronger" [9]

2.8 YOLO v3 (2018)

YOLO v3 is the last version of the official YOLO series. YOLO v3 balanced the speed, accuracy, and implementation complexity almost. And it got really popular within the industry due to its fast speed and straightforward components. The success and higher accuracy of the model comes from its feature extractor which works as its powerful backbone and another comes from RetinaNet-like detection head with an FPN neck. The new backbone network Darknet-53 leveraged ResNet's skip connections to understand an accuracy that's on par with ResNet-50 but much faster. Also, YOLO v3 ditched v2's under layers and fully embraced FPN's multi-scale predictions design. Since then, YOLO v3 finally changed each person's perspective of its poor performance while handling small objects.

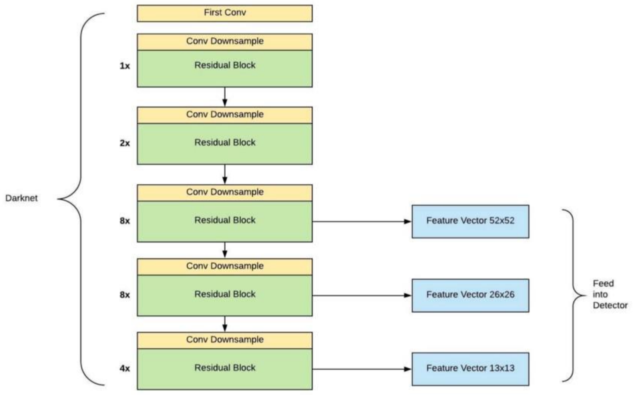

Fig -8: "YOLOv3: An Incremental Improvement" [10]

3. LIMITATIONS OR SHORTCOMINGS OF EXISTING MODELS [11]

3.1 OverFeat (2013)

Unfortunately, OverFeat is a one-stage approach and being a one-stage approach OverFeat suffers from very poorer accuracy because less prior knowledge is used. Thus, OverFeat failed to lead a hype for one-stage detector research, until a much more elegant solution came out 2 years later.

3.2 R-CNN (2014)

Since it is a neural network R-CNN takes a lot of time to train the network as we have to classify and separate over 2000 region proposals per image [12]. It cannot be implemented in real-time because it takes around 47 seconds for each and every test image. R-CNN was introduced in the year of 2014 combining region proposals with CNN. Major drawbacks are that it is slow, hard to train, and consumes large memory. This method uses a selective search to generate regions and to detect the object and It's not fully convolutional because selective search is not E2E trainable.

3.3 ResNet (2015)

It is the most powerful deep neural networks which have achieved state-of-the-art performance on the ILSVRC 2015 classification challenge [13]. The first implemented was the VGG network. Suppose the input size of an image or a video is given as 300 and 320 and even more than that, even though the ResNet-101 layers are deeper than the VGG-16 layers, the main disadvantage and con of ResNet-101 is that it decreases the accuracy. The main thing is that it has the capacity to undergo a deeper layer.

3.4 Fast R-CNN (2015)

Fast R-CNN uses a single method that extracts features from the regions, then divides them into different classes, and returns boundary boxes for the identified classes. It takes time to concentrate on increasing accuracy and decreasing time. This method was introduced in the year of 2015 [14]. It has high accuracy compared to the previous method and detects the object in a faster way.

3.5 Faster R-CNN (2015)

In the above methods, a selective search is used to detect the region proposals, which is time consuming and slow in the process. To overcome these problems, a new method was introduced, i.e., RoI Pool layer which is used to classify the image within the proposed region and can find the values for the boundary boxes. This method also consumes time and detects smaller or hidden objects that were introduced in the year of 2015. [15]

3.6 YOLO (2015-2018)

YOLO is a single-stage detector. The first breakthrough was in 2015 by Redmon et al. [16], it detects the real-time object and is very fast, better, and stronger compared to other methods. Its accuracy is very high. But the initial version of YOLO suffers from many shortcomings: predictions based on a coarse grid brought lower localization accuracy, two scale-agnostic regressors per grid cell also made it difficult to recognize small packed objects. One of the drawbacks of YOLO V1 is the bad performance in localization of boxes, because bounding boxes are learning totally from data.

Fortunately, in 2016 in YOLO v2, the authors added prior (anchor boxes) to help the localization. In order to introducing the anchors, some modifications are done on the architecture of the network

Later in 2018 in YOLO v3, the author makes some more modifications in YOLO v2 and hence improves

1. multi-scale prediction (class FPN)
2. It has better basic classification network like ResNet- 101 and classifier like darknet-53
3. Classifier-category prediction [17]

Moreover, it has a COCO dataset to store the data of images and videos and has excellent results on the COCO dataset. YOLO helps to detect moving objects, recognizes and helps to display in a rectangular bounding box with a provided caption. The major advantage is that the fast-moving objects are captured very quickly compared to the rest of the methods. This method is mainly used for speed. It is faster compared to any other method.

4. COMPARISON OF EXISTING MODELS

4.1 Accuracy

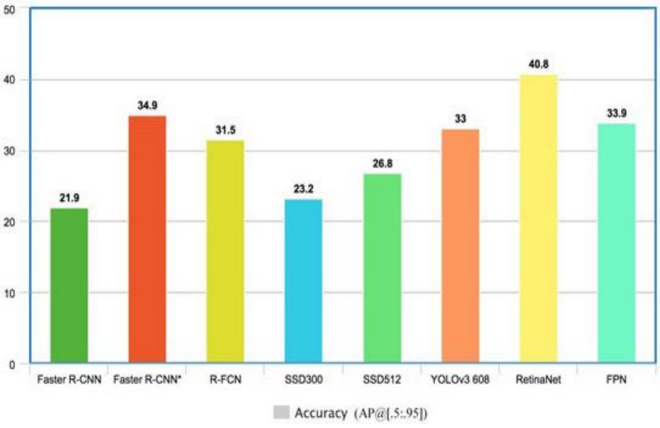

Fig -9: "Simple Surveillance System with the Tensorflow Object Detection API" [18]

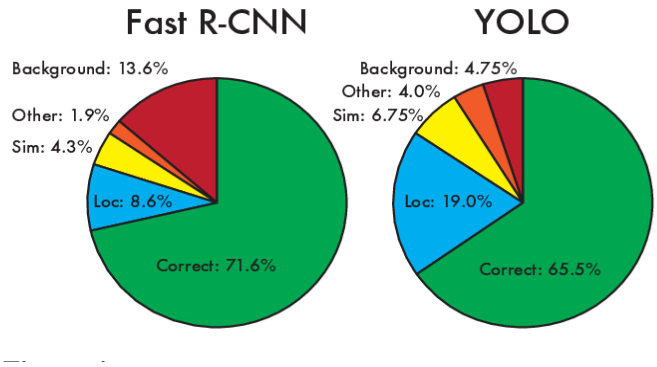

Fig -10: "A 2019 Guide to Object Detection" [30]

4.2 Frames per second

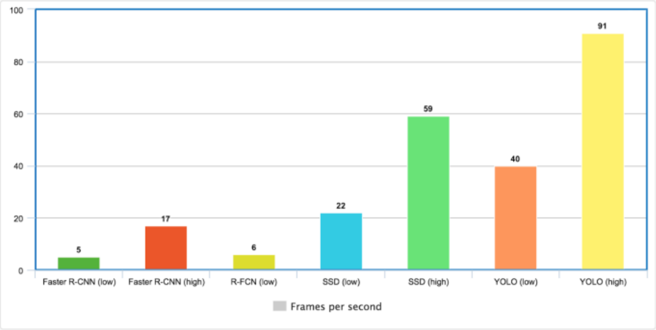

Fig -11: "A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework" [18]

4.3 Time Taken to Process the Images

Table -1: Time Taken by different Algorithms

Methods	Fast-RCNN	Faster-RCNN	YOLO
Time (Sec)	83.81045	73.88907	19.389161

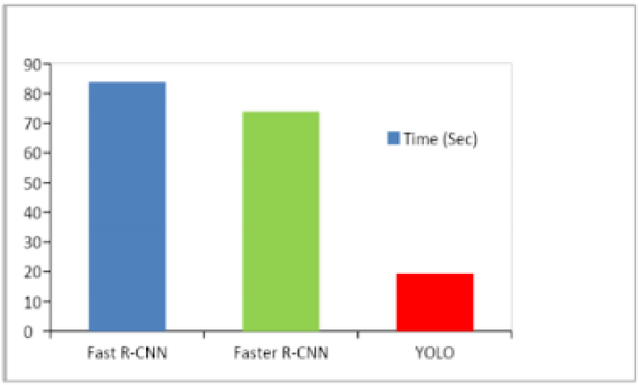

Fig -12: "Comparative Analysis on YOLO Object Detection with OpenCV" [11]

5. EXPLORING COMPUTER VISION LIBRARIES

5.1 Python Libraries for Computer Vision

5.1.1 Fastai - It is a deep learning library that provides high-level components that provides state-of-the-art results in standard deep learning domains. It also provides researchers with low-level features that can be mixed and matched to build new approaches.[19]

5.1.2 Keras - Written in Python, Keras is a high-level neural networks library which is capable of running either of the frameworks TensorFlow or Theano. This library was developed with a focus of fast experimentation. This deep learning library provides many features, including support for convolutional networks and recurrent networks, allowing easy and rapid prototyping.

5.1.3 Imutils - Imutils is a computer vision package that includes a series of OpenCV + convenience functions to make essential image processing functions such as translation, rotation, resizing, skeletonization, displaying python library Matplotlib images, sorting contours, detecting edges, among others relatively easy.

5.1.4 PyTorchCV - PyTorchCV is a PyTorch-based framework specialized for computer vision tasks. This framework is a collection of image classification, segmentation, detection, and pose estimation models. There are several implemented models in this framework, including AlexNet, ResNet, ResNeXt, PyramidNet, SparseNet, DRN-C/DRN-D, and more.

5.1.5 OpenCV - OpenCV is a popular and open-source computer vision library essential for real-time applications. OpenCV library has a modular structure and includes several hundreds of computer vision algorithms. OpenCV consists of a number of modules, including image processing, video analysis, 2D feature framework, object detection, camera calibration, 3D reconstruction and more.

5.1.6 Caffe - CAFFE (Convolutional Architecture for Fast Feature Embedding) is a deep learning framework, Caffe supports many different types of deep learning architectures geared towards image classification and image segmentation. It helps CNN, RCNN, LSTM and fully connected neural network designs. Caffe supports GPU- and CPU-based acceleration computational kernel libraries such as NVIDIA cuDNN and Intel MKL. [20] Caffe provides a complete toolkit for training, testing, finetuning, and deploying models, with well-documented examples for all of these tasks.[21]

5.2 Performance of different Computer Vision Libraries

5.2.1 Image Classification

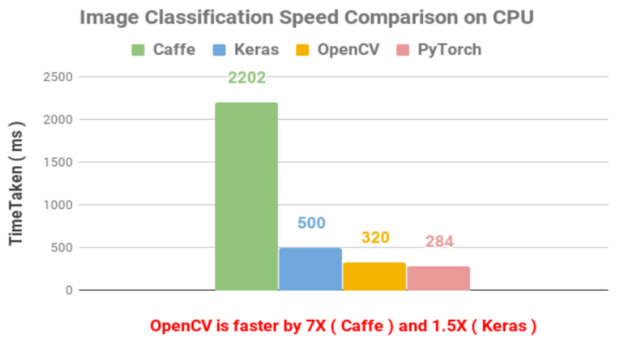

Fig -13: "Image Classification Speed Comparison on CPU"

5.2.2 Object Tracking

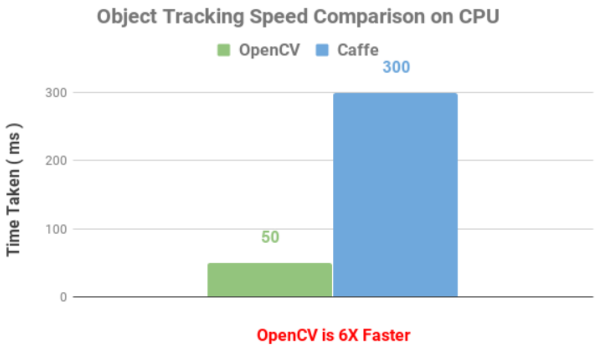

Fig -14: "Object Tracking Speed Comparison on CPU"

5.2.3 Pose Estimation

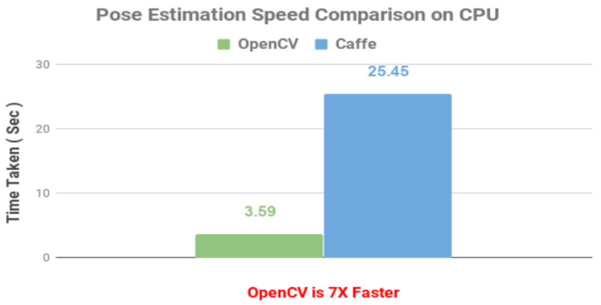

Fig -15: "Pose Estimation Speed Comparison on CPU"

Results show OpenCV is faster and more reliable being open source

5.3 OpenCV

It is a library of programming functions mainly used for image processing. It provides a de-facto standard API for computer vision applications. We can solve many real-time problems using image processing applications. Image processing is a form of signal processing in which the input can be an image or a video frame, the output is an image or set of characteristics related to the image. OpenCV is a library mainly used for image processing. It is freely available on the open-source Berkeley Software Distribution license. It was started as a research project by Intel. OpenCV contains various tools to solve computer vision problems. It contains low level image processing functions and high-level algorithms for face detection, feature matching and tracking.

5.3.1 Using OpenCV and YOLO for detection

YOLO and OpenCV methods are used for object detection, which detects each and every object clearly. The last step is to have the boundary boxes and labeled images. It is easy to understand and consumes less time to detect the object. In the table below it says about how the objects are detected and checks the process that goes through to detect an object.

There are four steps to follow to get a proper object detected image. The steps involved in this method are as follows. [31]

Steps	Description
Step I	Consider an image and we need to create a grid that will give us the features of an object.
Step II	In this step, we make using of OpenCV which will read the input image and data points and specify the file path to an image in a Numpy array.
Step III	Detecting an image in a grid view after the process of reading the image by OpenCV and Numpy and converting the grid to rectangular boxes.
Step IV	The final step consists of displaying the image with the rectangular box along with the caption on the window. This is done using YOLO and COCO dataset.

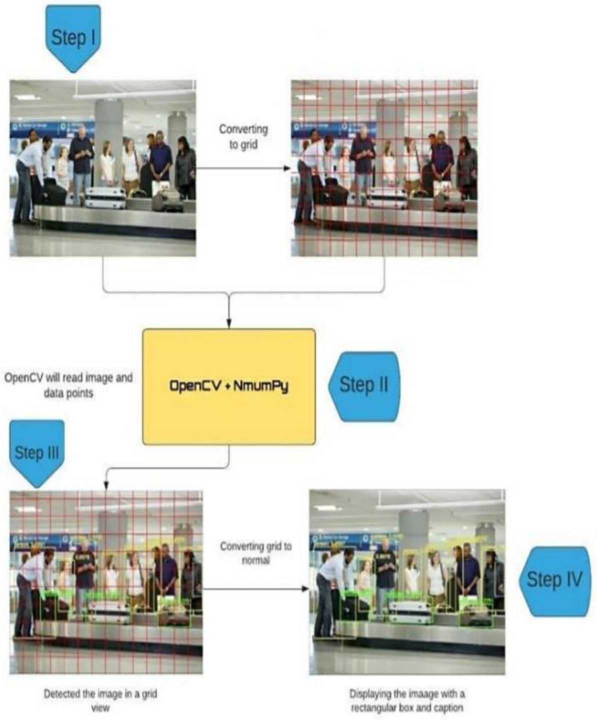
Step I
Converting to grid
OpenCV will read image and data points
OpenCV + NmumPy
Step II
Step III
Converting grid to normal
Step IV
Detected the image in a grid view
Displaying the imaage with a rectangular box and caption

Fig -16: Flowchart of the proposed algorithm.

6. HARDWARE SELECTION

Working on the topic related to Surveillance systems, the devices which come under surveillance systems are Surveillance Camera, Rover Board and Drone.

6.1 Surveillance Camera

CCTV video surveillance systems with the help of machine learning algorithms will be useful to classify and recognize objects such as humans, vehicles, etc. Once indexed, the video metadata can be used for configuring intelligent alerts, triggering real-time calls to action when certain objects or behaviors are detected – or when anomalous activity occurs [23]. Count-based alerts, for example, send notifications when the number of people in a certain area or building exceeds the limit defined by the organization. We can generate alerts which can be activated using face recognition and object identification and classification methods.

Surveillance recorded videos through CCTV cameras have given beneficiation in unstructured big data. The contribution of either real-time live video feeds or recorded video streams received from CCTV cameras have turned out to be as important as other sources like social media data, Bio data, agriculture data, sensor data, medical data, or data collected and evolved from space research. CCTV cameras are implemented in all places where security has much importance. Manual surveillance seems tedious and time consuming. We can define security in various terms and in various contexts like theft identification, violence detection, chances of explosion etc. In crowded public places the term security covers almost all types of abnormal events. Out of all the detection, violence detection is a tough nut to crack among all the activities which involves group activity. Anomalous or abnormal activity analysis in a video with a crowd scene is very difficult due to several real-world constraints [24].

Fig -17: Surveillance Camera

6.2 Rover

Evolution of rovers for four generations used on the red planet have been gathering scientific data, for sending evocative photographs and surviving in harsh conditions [26]. Similarly implementing the same technique on earth in our day-to-day life. As we know Rover boards on Mars have well efficient cameras which would be of greater help in our life, as it comes with the advantage of mobility over a surveillance camera. The mobility provides us access to compact areas which can prove to be advantageous and can be helpful in many sectors.

Fig -17: Rover

6.3 Drones

Mini-drones are increasingly used in video surveillance. Their aerial mobility and ability to carry video cameras provide new perspectives in visual surveillance. Drones provide the ideal solution to the problems and limitations faced by other surveillance methods. Surveillance using drones have a number of advantages, such as they are easier, faster and cheaper methods of Data collection and many more key advantages such as drones produce less noise, it can enter narrow and confined spaces where humans can't and also with various sensors it can detect what a human eye cannot.

Drones are also used to track and apprehend criminals. It is not uncommon to find them being used in war zones to track fugitives and terrorists. They can be also used in the dark for escorting people. This facility of providing security in secluded places can help reduce criminal incidents like having a real-time surveillance drone in concealed spaces like under the bridged, parking spaces can help to reduce criminal activities. They can even help monitor educational institutions and residential areas, detect abnormal activity and potential threats, and immediately notify concerned departments and law enforcement agencies.[28]

Drones can be useful as it will help in improving surveys, faster reporting, visibility of sites and easier inspection processes. Their advanced features can carry out difficult and dangerous operations efficiently and inspections over expensive and large sites can reduce human risk and costs. It acquires high accuracy and acceleration in data collection, allowing construction workers to track a site's progress with certain precision which is not possible using manual labor in the industry before.[28]. Drones can help to detect people working without safety gear and can help to raise alarm and avoid accidents.
Drones can help us with the update of what's going on around the site in real-time.

Fig -18: Drone

7. LIMITATIONS

7.1 Surveillance Camera

The background subtraction method detects objects with noise and output is not accurate. Object behind the object is not detected. Problems occur during identification of objects when any obstacles come before the object. If the position of the camera is not proper and the object in image is not captured properly then it cannot be identified. As the Camera is only placed at specific areas it cannot capture overall information of the day[25]. To get overall angle there might be a need for more cameras or Motion cameras which is not cost efficient.

7.2 Rover

Every robotic mobility comes with its advantages and disadvantages, similarly Rover comes with its own, Relatively low slope climb capacity which can cause wheel slippage, Obstacle traverse capability relatively low compared to other concepts, Friction in tracks can cause its inefficiency, One of the major disadvantage of the Rover is its operation speed which is Low and also its mobility is slow, on other hand jamming of parts which would make it prone to failure, it one complex structure to build mechanically (Control of walking)[27]. Additionally, getting it up to a higher level is a task itself and the camera won't be able to get image or video data due to its lower level.

7.3 Drone

Major drawback is that drones currently have less flight time, as batteries lose full capacity after a certain amount of time, reducing time more drastically. Lithium polymer batteries used in drones have a sensitivity to moisture. Therefore, flying in rain is avoided. Weather being a deciding factor, dense fog can be a limitation too, which might affect the image. In addition, the fog may hit the drone and cause water droplets, which may cause a malfunction. The drone has a wireless connection between the main unit and the controller. In wireless communication frequency plays an important and problem in it will create problems. Drones are controlled manually which requires good training with aviation, which might be difficult for trained personnel too. For data with higher accuracy the drones need to be with stable flight capabilities. And another issue is Privacy violation, the issue which is in discussion.

8. PROPOSED WORK

We have arrived on these points and we are proposing that we will be using YOLO v3 Algorithm with OpenCV for Object Detection and Classification due to

Its implementation in

1. Real Time
2. Higher accuracy
3. Higher Frames per second (fps)
4. Less time taken to process the input images and videos

Studying pros and cons of mechanical aspects in surveillance, Drone comes with limitations which are resolvable. Drones can fit out with various surveillance equipment which can collect high-definition video/image day and night. These equipment with technology will allow drones to intercept cell phone calls, GPS locations and gather license plate information. Nowadays, Drone comes with in-built libraries which can make collection of datasets and implementing algorithms easier.

9. CONCLUSIONS

In this paper, we have revisited important computer vision and Object detection and Classification based survey papers. Due to its powerful learning ability and advantages in dealing with occlusion, scale transformation and background switches, deep learning-based object detection has been a research hotspot in recent years. This review paper provides a not only a detailed review on deep learning-based object detection frameworks but also a detailed review on different types of Surveillance Systems which can be used to perform this Project. Then, we explored various Computer Vision Libraries and provided information on all the libraries and researched on Person Detection, Vehicle Detection and Anomaly detection or Event Detection techniques that can be applied for road network entities involving vehicles, people, and their interaction with the environment. Finally, we propose several promising future directions to gain a thorough understanding of the Computer Vision techniques or landscapes which include Object detection, its Packages (OpenCV) and its Libraries and Surveillance Systems. This review is also meaningful in the development of our Final Year Project.

REFERENCES

[1] Fact Checker Team, “Traffic Violations Caused 323 Deaths Every Day In 2017, Yet Fines Alone Not A Solution,” FactChecker.in is India’s first dedicated Fact Check initiative., Sep. 24, 2019. https://www.factchecker.in/traffic-violations-caused-323-deaths-every-day-in-2017-yet-fines-alone-not-a-solution/#:~:text=Mumbai%3A%20Violation%20of%20traffic%20rules,In%20India%2D2017%20data%2C%20the (accessed Jan. 23, 2021).

[2] “At which stage are cases pending in Indian courts?,” FACTLY, Aug. 21, 2019. https://factly.in/at-which-stage-are-cases-pending-in-indian-courts/ (accessed Jan. 23, 2021).

[3] P. Sermanet, D. Eigen, X. Zhang, and M. Mathieu, “OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks,” International Conference on Learning Representations, Feb. 2014, doi: 1312.6229.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 1, pp. 142–158, Jan. 2016, doi: 10.1109/tpami.2015.2437384.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[9] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690.

[10] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” IEEE Conference on Computer Vision and Pattern Recognition, Apr. 2018, doi: 1804.02767.

[11] H. Deshpande, A. Singh, and H. Herunde, "Comparative analysis on YOLO object detection with OpenCV," International Journal of Research in Industrial Engineering, no. 1, Mar. 2020, doi: 10.22105/riej.2020.226863.1130.

[12] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich features hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587). http://openaccess.thecvf.com/menu.py

[13] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). ImageNet: a large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.

[14] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE international conference on computer vision (pp. 1440-1448). http://openaccess.thecvf.com/menu.py

[15] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems (pp. 91-99). Neural Information Processing Systems Foundation, Inc.

[16] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). http://openaccess.thecvf.com/menu.py

[17] "Target detection--YOLO v1, v2, v3 - Programmer Sought,"ProgrammerSought. https://www.programmersought.com/article/40594604346/ (accessed Jan. 23, 2021).

[18] "Simple Surveillance System with the Tensorflow Object Detection API_Gilbert Tanner - MdEditor," Markdown-MdEditor. https://www.mdeditor.tw/pl/2SvU/zh-hk (accessed Jan. 23, 2021).

[19] S. A. Sanchez, H. J. Romero, and A. D. Morales, "A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework," IOP Conference Series: Materials Science and Engineering, p. 012024, Jun. 2020, doi: 10.1088/1757-899x/844/1/012024.

[20] A. Choudhury, "10 Best Python Libraries For Computer Vision Tasks," Analytics India Magazine, Sep. 28, 2020. https://analyticsindiamag.com/10-best-python-libraries-for-computer-vision/ (accessed Jan. 23, 2021).

[21] Contributors to Wikimedia projects and Y. Jia, "Caffe (software) - Wikipedia," Wikipedia, the free encyclopedia, Mar. 29, 2017. https://en.wikipedia.org/wiki/Caffe_(software) (accessed Jan. 25, 2021).

[22] Y. Jia, E. Shelhamer, and J. Donahue, "Caffe: Convolutional Architecture for Fast Feature Embedding," Nov. 2014, doi: 10.1145/2647868.2654889.

[23] "How Can Machine Learning be Used for CCTV Video Surveillance?|BriefCam,"BriefCam. https://www.briefcam.com/resources/blog/how-can-machine-learning-be-used-for-cctv-video-surveillance/ (accessed Jan. 25, 2021).

[24] Sreenu, G., Saleem Durai, M.A. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. J Big Data 6, 48 (2019). Intelligent video surveillance: a review through deep learning techniques for crowdanalysis|https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0212-5#Abs1

[25] Payal Panchal, Gaurav Prajapati, Savan Patel, Hinal Shahand, Jitendra Nasriwala| INTERNATIONAL JOURNAL FOR RESEARCH IN EMERGING SCIENCE AND TECHNOLOGY, VOLUME-2, ISSUE-1, JANUARY-2015|A Review on Object Detection and Tracking Methods|https://ijrest.net/downloads/volume-2/issue-1/pid-21201506.pdf

[26] "Deep learning will help future Mars rovers go farther, faster, and do more science -- ScienceDaily," ScienceDaily. https://www.sciencedaily.com/releases/2020/08/200819120700.htm (accessed Jan. 25, 2021).

[27] Grimm, Christian. (2011). Concept Development and Design of a Flexible Metallic Wheel with an Adaptive Mechanism for Soft Planetary Soils.|https://www.researchgate.net/publication/270648864_Concept_Development_and_Design_of_a_Flexible_Metallic_Wheel_with_an_Adaptive_Mechanism_for_Soft_Planetary_Soils/citation/download

[28] "Use Machine Learning APIs with Drones," Intelligent document processing with AI | Nanonets. https://nanonets.com/drone/ (accessed Jan. 25, 2021).

[29] Mario, "What Are The Disadvantages of Drones? – Drone Tech Planet," Drone Tech Planet, Dec. 11, 2019. https://www.dronetechplanet.com/what-are-the-disadvantages-of-drones/ (accessed Jan. 25, 2021).

[30] D. Mwiti, "A 2019 Guide to Object Detection. Common model architectures and a few… | by Derrick Mwiti | Heartbeat," Medium, Jul.18,2019. https://heartbeat.fritz.ai/a-2019-guide-to-object-detection-9509987954c3 (accessed Jan. 25, 2021).

[31] H. Deshpande , A. Singh, H. Herunde | Department of Computer Application, Jain (Deemed to-be) University, Bengaluru, Karnataka, India. | Comparative Analysis on YOLO Object Detection with OpenCV

[32] E. News Service, "Accidents at workplaces in India 'under reported'; 38 per day in construction sector: Study | India News,The Indian Express," The Indian Express,Nov.21,2017.https://indianexpress.com/article/india/accidents-at-workplaces-in-india-under-reported-38-per-day-in-construction-sector-study-4947079/ (accessed Jan. 23, 2021).

BIOGRAPHIES

Sugam Dembla, Final year B. Tech Mechatronics Student currently studying in Symbiosis Skills & Professional University (SSPU), Pune.

Niyati Dolas, Final year B. Tech Mechatronics Student currently studying in Symbiosis Skills & Professional University (SSPU), Pune.

Ashish Karigar, Final year B. Tech Mechatronics Student currently studying in Symbiosis Skills & Professional University (SSPU), Pune.

Dr. S. S. Sonavane currently working as the Director at School of Mechatronics Engineering of Symbiosis Skills & Professional University (SSPU), Pune.