INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

Review on Techniques Used to Detect Objects

VENKATA RAMANA MADAM
Research Scholar
Department of Computer Science
J.S UNIVERSITY, Shikohabad, UP

Dr. VIKRAM SINGH RATHORE
Professor
Department of Computer Science
J.S. UNIVERSITY, Shikohabad, UP

Dr. VITTAPU MANI SHARMA
Professor
Department of Computer Science
J.S. UNIVERSITY, Shikohabad, UP

Abstarct - This paper presents review of different traditional and deep learning algorithms techniques in detecting different objects. The traditional techniques solved problem of object detection using different features such as SIFT, Haar features, HOG features, and deep learning techniques used pretrained CNN models for feature extraction followed by different architectures for object detection. It covers the advantages and limitations of existing models and techniques used for detecting the objects. Further the work done in traffic light detection is also discussed. It also discusses the experimental work done for classification on CIFAR 10 dataset and a dataset created from benchmarked dataset ImageNet, using Visual Geometry Group 16 (VGG16) model, MobileNet model, InceptionV3 model and Capsule Networks. The performance of the different architectures is evaluated and compared.

Keywords- CNN, Algorithms, Deep Learning, Object Detection, Traffic Light.

1. INTRODUCTION

Object detection is a method for identifying and classifying objects in visual media. Object detection is a powerful tool for improving visual comprehension and has several practical uses, such as automated face tagging, video surveillance, pedestrian identification, image retrieval, and face detection.

Conventional object detection methods consisted of two stages: feature extraction and classification. Classifiers like Support Vector Machines (SVM) and Adaboost—both of which do a great job of identifying the objects' classes—used feature vectors generated using methods like SIFT, Histogram of Gradients (HOG), Haar features, and so on. The main problem with the old methods was that it was hard to build a general pipeline, as the performance of the classifier depended greatly on the characteristics that were gathered.

Variations in size and background, lighting, look, and stance further complicate item identification. A lot has changed in object identification since 2012, with challenges like ILSVRC and PASCALVOC bringing solutions to object detection problems. Pascal VOC was in office from 2005 until 2012. The ILSVRC has been conducted yearly since 2010 and replaced Pascal VOC as the principal standard for object identification after its 2012 termination. The ILSVRC competition makes use of the ImageNet dataset, which has over fourteen million images organized into one thousand object classification classes and two hundred detection classes. In computer vision applications, CNN plays a crucial role. In 2012, the efficiency of object recognition was greatly enhanced by AlexNet, a deep convolutional neural network (CNN) trained to divide a huge dataset of high-resolution images into a thousand distinct classes.

ISSN: 2057-5688

A 30% improvement in accuracy on the datasets used in the PASCAL VOC 2012 competition was achieved by the Region Based Convolutional Neural Network (R- CNN), the first major development. Convolutional neural networks (CNNs) use deeper designs to extract hierarchical features that help extract useful information. The emergence of massive datasets, high-speed parallel systems, GPU, and improvements in network topologies like as ResNet, AlexNet, GoogLeNet, and VGG are only a few of the factors that have contributed to the dramatic improvement in the performance of deep learning algorithms. In order to achieve high-level performance, deep learning requires large datasets.

Popular open datasets include ImageNet, which has millions of images across 20,000 object classes, PASCAL VOC 2012, which has 20,530 images across 11,530 classes, and Microsoft MS COCO, which has 300,000 images across 80+ object categories.

The following performance measures are used for evaluation purposes: Accuracy, Precision-Recall (PR) curve, Intersection over Union (IoU) threshold, Accuracy, mAP, and Receiver Operating Curve (ROC) curve.

With the use of machine learning (ML), computers may improve themselves over time by learning from data automatically, without any human input whatsoever. Watched and unsupervised are the two main types. Supervised methods use labeled datasets from the past to forecast the results of future events, in contrast to unsupervised methods that learn from the unlabeled data. The standard procedure for machine learning involves extracting features from training data before the classifier can make a prediction about the objects' classes. Due to the fact that the training data does not include all potential inputs, the learning algorithm has to be trained to predict the outcome for unseen events. Once the features have been automatically extracted, deep learning algorithms move on to the classification phase.

2. LITERATURE REVIREW

In visual content, object detection is one method for identifying and classifying things. Object detection is a

rful technique for improving visual perception; it several practical uses, such as automatic face ng, video surveillance, pedestrian recognition, e retrieval, and face detection.

past, object identification methods relied on two processes: feature extraction and classification. llent object classifiers like Support Vector ines (SVM) and Adaboost used feature vectors ated by methods like SIFT, Histogram of ents (HOG), Haar features, etc. Because the rmance of the classifier was highly dependent on athered characteristics, the main problem with the methods was that it was hard to build a generic ne.

Size, background, lighting, look, and posture differences significantly complicate item identification. Problems with object detection have been addressed via challenges such as ILSVRC and PASCALVOC from 2012, leading to considerable progress in object identification. Pascal VOC served as an official from 2005 to 2012. The ILSVRC, which has been conducted yearly since 2010, succeeded Pascal VOC as the principal item identification benchmark upon its termination in 2012. In the ILSVRC competition, participants utilize the ImageNet dataset, which contains over fourteen million pictures divided into two hundred detection classes and one thousand object classification classes. Computer vision applications rely on CNN. In 2012, the object identification efficiency was greatly enhanced using AlexNet, a deep convolutional neural network (CNN) trained to categorize a large dataset of high-resolution pictures into a thousand distinct groups.

R-Convolutional Neural Networks (R-CNNs) were the first major development; they increased accuracy on PASCAL VOC 2012 datasets by 30%. Convolutional neural networks (CNNs) use deeper architectures to retrieve hierarchical characteristics that facilitate the extraction of useful information. Many factors have contributed to deep learning algorithms' remarkable performance boost in recent years. These include the proliferation of massive

ISSN: 2057-5688

datasets, the advent of GPUs and high-speed parallel systems, and innovations in network topologies such as ResNet, AlexNet, GoogLeNet, and VGG. For deep learning to function at its best, large datasets are required.

A few well-known public datasets include: ImageNet (containing millions of pictures in 20,000 object classes), PASCAL VOC 2012 (containing 20,530 images in 11,530 classes), and Microsoft MS COCO (containing 300,000 photographs in more than 80 object categories). Performance measures used for evaluation include Accuracy, Precision-Recall (PR) curve, Intersection over Union (IoU) threshold, Accuracy, mAP, and Receiver Operating Curve (ROC) curve.

With the use of machine learning (ML), computers may learn from data on their own, without any human intervention, and hence improve over time. Two main types exist: those that are monitored and those that are not. Supervised methods use labeled datasets from the past to forecast the results of future events, as opposed to unsupervised methods that learn from the unlabeled data. The standard procedure for machine learning requires feature extraction from training data before the classifier can make object class predictions. Since not all inputs are included in the training data, the learning algorithm needs to be trained to anticipate the outcome of unseen events. Deep learning algorithms proceed to the classification step once the features have been automatically extracted.

3. REVIEW OF DEEP LEARNING TECHNIQUES FOR OBJECT DETECTION

Pascal VOC was in office from 2005 until 2012. The ILSVRC has been conducted yearly since 2010 and replaced Pascal VOC as the principal standard for object identification after its 2012 termination. The ILSVRC competition makes use of the ImageNet dataset, which has more than fourteen million images and a large number of object categorization and detection classes (thousands of classes). Many factors have contributed to deep learning algorithms' dramatically improved performance in recent years. These include improvements in network topologies like as ResNet, t, GoogleLeNet, and VGG, as well as the ation of large datasets and high-speed parallel ers.

ject recognition, Szegedy et al. (2013) proposed rNet, a CNN-based Deep Neural Network . It is possible to use DNN-based regression to get metric data. Following the use of the multi-scale nference method, they produced detections. ing to Krizhevsky et al. (2017), they used t to tweak the final classifier's regression layers. twork first clustered the pixels, then after making ast, it converted the masks into bounding boxes. f its operation on several crops, the network ns subparly. According to Sermanet et al. (2013), t deep learning object detector, Overfeat Network, ated by combining CNN with the sliding window ue. Predictions were generated by first ing each visual component as an object or a non-

Krizhevsky et al. (2017) constructed a deep convolutional neural network (DCNN) with 60 million parameters. The network consists of five convolutional layers, three FC layers, and SoftMax layers. It was trained to recognize 1.2 million pictures in 1000 separate item categories for the ILSVRC-2010 competition. The model has a Top-1 error rate of 37.5% and a Top-5 error rate of 17.0% on the test dataset, as per the results. Their efforts were fruitful in GPU-efficient 2D convolution process implementation. In order to decrease the model's overfitting, the authors used augmentation and dropout techniques.

In order to develop RCNN, a simple and extensible detection method, Girshick et al. (2014) combined CNN with region recommendations. The processes of Region Proposal and Object Detection are shown in Figure 2.1 of the detector. Using the SS method, they were able to produce 2000 area suggestions. We used convolutional neural networks (CNNs) to trim the generated area proposals to a certain size so that we could obtain 4096 dimensional information. Using linear SVMs, the regions were sorted into positive and negative groups according to their scores.

ISSN: 2057-5688

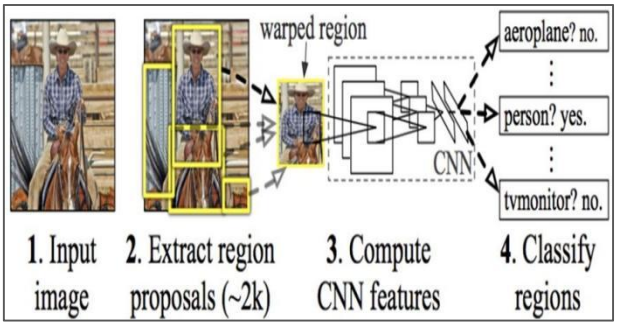

Figure 2.1: Region Convolutional Neural Network (RCNN) (Source: Girshick et al., 2014)

ISSN: 2057-5688

After then, the NMS method was used to determine the ultimate bounding boxes. The model achieved a mAP score of 53.3% on the VOC2012 dataset. Reducing the size of the features vector and sharing CNN parameters across all categories were the two main requirements for good detection, according to the authors. Pretraining the network under supervision was also shown to be very effective when working with small datasets. Algorithms that use saliency cues improve the accuracy of candidate bounding boxes of different sizes. Based on the approach discussed earlier, CNN's main drawback is that it requires input pictures of fixed size, and RCNN training requires Support Vector Machines (SVM) to match the recovered features. Second, there was a huge storage need since training the model took longer and there was redundancy in the indicated locations.

By using a pooling mechanism, the innovative model SPPNet (He et al., 2015) eliminated the need for a set size of the input image. SPP used a combination of regional features and levelled the image. The SPP layer's principal function was to aggregate different attributes into fixed-size outputs that the FC layers could use as inputs. An SPP layer was superimposed on top of the network's prior convolutional layer. The proposed approach could manage images with varying dimensions, aspect ratios, and scales. Findings from the study indicate that the proposed network excelled in both detection and classification.

In their 2014 paper, Simonyan and Zisserman proposed a CNN that took an input image with dimensions 224 x 224 x 3 and processed it using a stack of convolutional layers with 3x3 filters, a 1 pixel convolution stride, and 1 pixel padding. There were three FC levels in the spatial pooling network: two 1000-channel layers and one 4096-channel layer. A 2x2 pixel window with a two-pixel stride was employed for each MaxPooling layer. While every hidden layer made use of the ReLU activation function, the SoftMax function was used by the final layer. From 11 to 19 layers deep, they examined how the categorization error dropped. It was made feasible to train and evaluate on several GPUs by dividing batches of training photos and processing them on parallel GPUs. The testing were conducted using the ILSVRC-2012.

Szegedy et al. proposed Inception, a powerful design that leverages multiscale processing and the Hebbian principle (2015). The parameters of the network increase in proportion to its growth, which causes overfitting and causes the use of computing resources to skyrocket. To get around this problem, the writers used sparse connected architectures. Nevertheless, sparse designs were not well-supported by the existing infrastructures, and libraries relied on dense matrix multiplication. In order to create the sparse structure, the authors used their device-agnostic GoogLeNet inception architecture. In total, there were one hundred levels, and twenty-two of them had parameters. The network training was conducted using the asynchronous stochastic gradient technique. In the ILSVRC 2014 Classification Challenge, the network tested on both the validation and testing datasets, and it earned a Top-5 error rate of 6.67%. Findings from the study showed that dense building blocks could be employed to mimic sparse systems, which improved network performance.

Because previous systems produced several separate candidate area ideas, the authors devised the idea of an area Proposal Network (RPN) to create object suggestions from input photographs of any
ize, thereby addressing the main drawback of hose methods. igure 2.2 shows that Faster RCNN was ntroduced by Ren et al. (2016). It used RPN nstead of SS to generate region ecommendations. They classified using haracteristics from the backbone network's ast shared convolutional layer, and Fast RCN vas in charge of the process. Using anchors, RPN projected several aspect ratios and image egions. Results from testing using the VGG16 nodel as its foundation shown that the roposed model could provide a 5FPS frame ate on a GPU. Its mAP scores were 70.4% on he 2012 dataset and 73.2% on the 2007 lataset of PASCAL VOC.

ISSN: 2057-5688

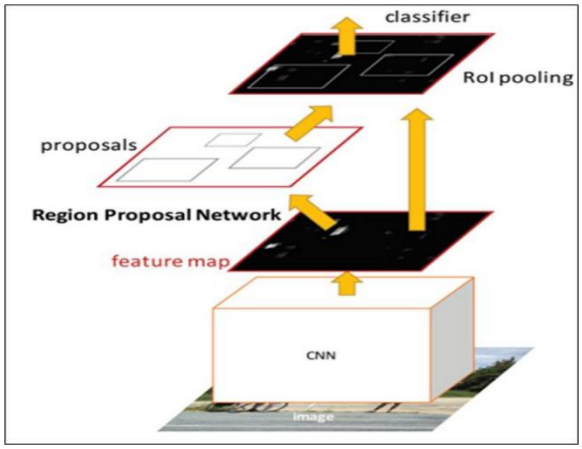

Figure 2.2: Faster RCNN (Source: Ren et al., 2016)

Lin et al. (2017) introduced a neural network codenamed Feature Pyramid Network (FPN) that attained scale invariance by generating various feature maps that represent image at different scales. After generation of feature maps, Faster RCNN network was run on each scale and multiscale testing was performed. The architecture of FPN as shown in Figure 2.3 has the capability to handle any sized input image and produced the output as proportionally sized feature maps at multiple levels.

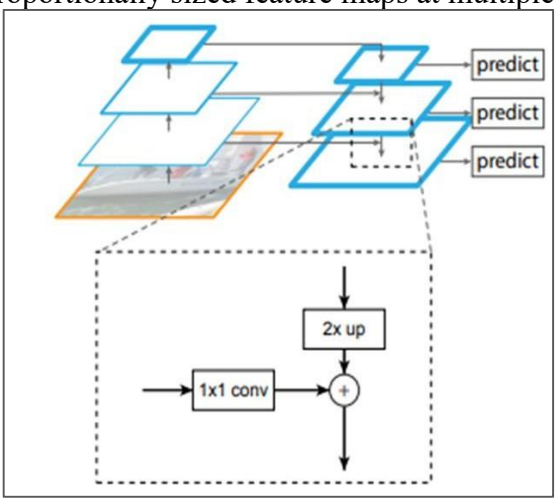

Figure 2.3: Feature Pyramid Network (FPN) (Source: Lin et al., 2017)

4. TRAFFIC LIGHT OBJECT DETECTION

Intelligent cars and advanced driver assistance systems both rely on traffic light detection. It determines where and how the traffic signal is functioning. The most difficult issues with traffic signal identification are computation time, background variation, lighting conditions, vehicle rear lights, low resolution, and occlusion. Fairfield and Urmson et al. (2011) proposed mapping the 3D positions of the traffic lights using earlier map information that captures the control semantics of different lights. Nighttime performance was satisfactory, but rain and light glare are the two biggest obstacles to accurate identification.

In their work, Siogkas et al. (2012) developed a system that utilizes a monocular onboard camera to detect traffic lights in challenging environmental conditions. They improved the system's capacity to discern between different colored lights and controlled the blooming effect by including a color pre-processing module. A rapid color space based on CIE-L*a*b* is used by the approach.

To reduce the number of false positives, radial symmetry transformations may be used to find symmetrical regions and spatial information. Philipsen et al. (2015) discussed traffic light recognition using several image processing techniques for scene capture. By integrating image processing with a Hidden Markov Model (HMM)–developed approach, they were able to ascertain the traffic signal's present state. The testing results showed that the technology achieved an accuracy of 90.55 percent, which is much higher than the 78.54% accuracy achieved by prior image processing techniques. Using visual data that included prior knowledge of the size, location, aspect ratio, and context of the traffic lights, Li et al. (2017) developed the Traffic Light Recognition method. They built structure and fine-tuned channel characteristics for all types of light by merging the detectors. They also improved the results' accuracy by including interframe

information analysis. Testing was conducted using the VIVA data set, and the model achieved an accuracy of 71.50% on red traffic lights, 27.67% on red left traffic lights, 52.16% on green traffic lights, and 40.47% on green left traffic lights using the VIVA validation dataset. With the use of deep learning and high dynamic photography, Wang and Zhou (2018) created an identification method to find possible candidates for traffic lights in low and dark frames. The system achieved accurate image recognition and categorization in a wide range of lighting conditions by combining context information in well-lit frames with the saliency model, which takes into account both color and shape information in darker frames. The employment of a multiclass classifier in the bright channel helped to decrease the number of false positives. Performance was enhanced via temporal trajectory tracking, and search regions were minimized for quicker recognition by using a prior mask.

Lu et al. (2018) presented a new attention two-stage architecture consisting of Accurate Locator and Recognizer (ALR) and Attention Proposal Modeller (APM) to generate probable regions with the aim of finding and categorizing small items. Using Tencent Street View as a standard, they collected 15,000 traffic light samples and created a dataset. On the Tsinghua Tencent 100K test dataset, the model achieved an AP of 91.7%, an AR of 83.4%, and a mAP score of 87.0% at an IOU of 0.5, as shown in the trials. The model achieved mAP=86.2%, AR=83.6%, and AP= 84.7% on the TTTL test set.

To enable object recognition with a pixel size less than 10, Müller and Dietmayer (2018) redesigned SSD. The stride value for following network layers reduced as InceptionV3 took over for VGG16 during earlier box creation for small object identification. We utilized the NMS method to make sure that the identical item wouldn't be detected more than once. The experiments demonstrated false positive rates of between 0.1 and 10 Frame Positives Per Image (FPPI) using the DriveU Traffic Light Dataset (DTLD), with recall rates of 95% for smaller items and 98% for larger ones, respectively. Using deep learning and the outlier detection approach, Munoz-Organero et al. (2018) were able to identify odd driving locations using GPS-derived data. At each anomaly, data on acceleration and speed was automatically extracted in order to classify different types of traffic signals, roundabouts, and street crossings and to generate traffic maps. Classification accuracy was 0.88 and recall was 0.89 based on the model's experimental data.

Using Faster R-CNN InceptionV2 and Single Shot Multibox Detector (SSD) MobileNetV2, Janahiraman and Subuhan (2019) were able to recognize traffic signals in Malaysia. Compared to the SSD-based model, which achieved an accuracy of 58.209% in traffic light identification, faster R-CNN achieved a far higher accuracy of 97.015%. Possatti et al. proposed YOLOv3, an object recognition system based on deep learning that contains both online and offline components (2019). In contrast to the realtime nature of traffic light detection, earlier map construction and annotation were performed offline. The studies included distinguishing between red-yellow and green traffic lights on predetermined routes using the DTLD, LISA-TLD, and IARA-TLD datasets. With parameters IoU = 0.5 and two threshold values of 0.2 & 0.5, the model attained a mAP score of 55.21% on IARA-TLD and a recall of 62.28% at a threshold of 0.2.

Wang et al. (2020) proposed a detection approach that includes an area recommendation and classification stage.

At the area proposal stage, zones were generated according to traffic light color, intensity, and shape using Gaussian filtering, grayscale processing, and the TopHat transform. We further selected the regions according to the hue component after converting the image from RGB to HSI color space. The investigations used 6,804 images, with an average accuracy, recall, and precision of 99.6%, 99.2%, and 98.5%, correspondingly, for the model.

Lee and Kim (2019) proposed a DNN-based encoder-decoder network that, to recognize small traffic lights, makes use of concentrated regression loss. The model used freestyle anchor boxes to detect objects close to the grid's boundary. There was a 19.86% rise to 49.16% in the Area Under Curve (AUC) and a 7.19% increase to 42.03% in the mAP score on the Bosch-TL dataset, as shown in the experimental results.Biswas et al. (2019) found a solution to the problem of estimating traffic density for automated traffic monitoring. No matter the size, shape, or angle of view, SSD can handle everything. Using data collected from fifty-nine independent traffic cameras, they compared MobileNet-SSD against SSD. With an accuracy of 92.97%, SSD outperformed MobileNet-SSD, which reached 77.30%. By using bilinear interpolation and creating a novel improved loss function based on IoU, Cao et al. (2019) enhanced the position information. Results showed an improvement in accuracy and memory when it came to the little things.

A portable aid system for the visually impaired to identify traffic signals using the AdaBoost algorithm was developed by Wu et al. (2019) utilizing a parallel architecture constructed using FPGA. All parts of the system worked together to perform calculations on the integral representation of the image. The MATLAB program was used to simultaneously compute the confidences of weak classifiers before they were recognized on the FPGA platform using Vivado design suites. The traffic light footage used in the experiment were able to achieve 30 FPS. Heuristic algorithms have low accuracy and excessive power consumption, according to Ouyang et al. (2019), whereas deep learning methods have the opposite problem. Their lightweight, real-time traffic light system used a convolutional neural network (CNN) classifier in conjunction with a heuristic-based module to select possible detection locations. In GPU-based simulations, the model achieved an accuracy of 99.3 percent on the Nvidia Jetson TX1 and 99.7 percent on the TX2 with the detector module. Yeh et al. (2019) demonstrated a system that included detection and identification steps. For the detection process, map data was used. To further aid with distant traffic signal detection, two cameras with varying focal lengths were used.

Aneesh et al. (2019) trained and evaluated a RetinaNet-based model on the 1280 x 720 pixel traffic signal photos included in the Bosch Traffic signal Dataset (BTLD). Using connected component-based labeling with an 8-connected neighborhood, Kim et al. (2019) presented a segmentation approach that found the coordinates of the bounding boxes to identify potential regions and a convolutional network. Compared to the standard Faster R- CNN, the proposed technique performed better in the simulations.Modules such as residual, densenet, and normal comprised the Multi-Backbone Network (MBBNet) presented by Ouyang et al. (2020). Reducing computations was achieved by the use of channel compression. The results shown that the system achieved an accuracy of over 94% on low-power edge devices while processing high-resolution images.To locate minute objects, Hassan

ISSN: 2057-5688

et al. (2020) compared two approaches. The first approach used a typical color-based segmentation method to identify objects based on Hue, Saturation, and Value attributes; the second method used mask R-CNN to detect traffic lights.

In order to recognize and categorize photographs in the Bosch Small Traffic Light Dataset, Gokul et al. (2020) evaluated the model architecture and parameters of Faster RCNN and YOLO detectors. In terms of accuracy, the experiment demonstrated that the Faster RCNN model outperformed the YOLO model. The goal of Wang et al. (2022) was to identify and recognize traffic signals. To improve detection accuracy, we used shallow features and added a Gaussian model to the bounding box prediction to make it more uncertain. Results from the VIVA Challenge Competition showed that the proposed strategy raised the area under the PR curve by 7.09%. An algorithm called YOLOv4 was surpassed by 2.86% by the mAP score.
Using the YOLOv5 algorithm and K-means for anchor generation, Yan et al. (2021) were able to solve issues based on color and geometric characteristics. The experiments yielded detection speeds of 143 FPS and 63.3% AP on the BDD100K dataset, correspondingly.

CONCLUSION

This paper reviewed the literature of researched and experimented techniques, models, architectures used by researchers for detecting different objects. It also discussed different models, architecture, and techniques most prevalent for Traffic Light Detection. On the basis of the literature survey, the work in this thesis identified the scope for development of the technique for Traffic Light Detection.

REFERENCES

[1] Aneesh, A.N., Shine, L., Pradeep, R. and Sajith, V., 2019. Real-time Traffic Light Detection and Recognition based on Deep RetinaNet for Self-Driving Cars. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 1, pp.1554-1557.

[2] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K. and Ghayvat, H., 2021. CNN variants for computer vision: History, architecture, application, challenges and future scope. Electronics, 10(20), p.2470.

[3] Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., Liu, Y. and Wu, Z., 2019. An improved faster R-CNN for small object detection. IEEE Access, 7, pp.106838- 106846.

[4] Fairfield N. and Urmson C., 2011. Traffic light mapping and detection. In 2011 IEEE International Conference on Robotics and Automation, pp. 5421-5426.

[5] Girshick, R., 2015. Fast r-cnn. In Proceedings of the IEEE International Conference on computer vision, pp. 1440-1448.

[6] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861

[7] Lin, X., Zhao, C. and Pan, W., 2017. Towards accurate binary convolutional neural network. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 344-352.

ISSN: 2057-5688

[8] Razakarivony, S. and Jurie, F., 2016. Vehicle detection in aerial imagery: A small target detection benchmark. Journal of Visual Communication and Image Representation, 34, pp.187-203.

[9] Redmon, J. and Farhadi, A., 2017. YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271.

[10] Ren, S., He, K., Girshick, R. and Sun, J., 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp.1137-1149.

[11] Road Accidents in India 2019. Available at (https://morth.nic.in/sites/default/files/RA_Uploading.pdf) (Accessed: 24 July 2020).

[12] Tuzel, O., Porikli, F. and Meer, P., 2006, May. Region covariance: A fast descriptor for detection and classification. In European conference on computer vision, pp. 589-600. Springer, Berlin, Heidelberg.

[13] Uijlings, J.R., Van De Sande, K.E., Gevers, T. and Smeulders, A.W., 2013. Selective search for object recognition. International journal of computer vision, 104(2), pp.154- 171.

[14] Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, pp. I-I.