DESIGN AND IMPLEMENTATION OF DL BASED MOVING OBJECT DETECTION AND TRACKING WITH CANNY EDGE DETECTION

[1]P.SUNANDA, [2]K.ASHARANI,[3]G.ALEKHYA ,[4]S.SHASHIKALA

[1,2,3,4] Assistant Professors, Department of Computer Science & Engineering,
G Pulla Reddy Engineering College (Autonomous), G Pulla Reddy Nagar,
Nandyal Road,Kurnool,Andhra Pradesh, India-518007.
Author mail id's: [1,2,3,4]psunandareddy@gmail.com, asharani.ecs@gprec.ac.in,
alekhyagurramgadda.ecs@gprec.ac.in,sshashikala.ecs@gprec.ac.in

ABSTRACT: As object detection and object tracking are two connected aspects of video surveillance, they are both regarded as crucial steps in computer vision for video surveillance, traffic analysis, and public safety. The first step to recognize objects in movies, which is necessary before moving on more difficult tasks like tracking. A strong programming model known as deep learning neural networks teaches to represent and abstract data like pictures, sounds, and text at various levels of abstraction. In this paper, a well-known deep learning network, Faster Regional Convolution Neural Network (Faster R-CNN), is applied to introduce the Object Detection and Tracking System (ODTS) for object detection and conventional object tracking algorithm for automatic detection and monitoring of unexpected events in Closed Circuit Televisions (CCTVs) in tunnels. To collect Bounding Box (BBox) results through Object Detection, ODTS receives a video frame in time as an input. Then it compares the bounding boxes of the present and previous video frames to assign a unique ID number to each moving and identified object. This technique allows it to be possible to follow a moving item in real time, which is difficult in object detection frameworks that follow conventional methods. After that, four accident videos featuring each accident were used to evaluate the ODTS-based Tunnel Closed Circuit Television (CCTV) Accident Detection System based on a trained deep learning model. Consequently, the system is able to identify every accident within ten seconds. The most significant aspect is that as the training dataset becomes more extensive, ODTS's detection capacity can be automatically enhanced without modifying the program codes. Accuracy, Recall and F1-Score are used parameters for performance analysis.

KEYWORDS: Deep Learning, Object Detection and Tracking System (ODTS), Faster Regional Convolution Neural Network (Faster R-CNN), Computer Vision.

I.INTRODUCTION

The automatic detection, classification, and tracking of numerous objects has applications in security, surveillance, traffic control, and object occlusions, random motions, complex backgrounds, varying illumination, and human-computer interaction have all been considered challenging tasks for a wide range of applications, making real-time tracking a problem that persists and a topic of active research [4]. Motion detection and motion estimation are two methods for detecting moving objects. While objects are moving, motion detection uses the fixed camera to recognize different areas from video frames. The following factors make object detection with fixed camera difficulties: Objects in a scene do not move consistently from frame to frame. The items in the scene might stop for a while before continuing to move. The scene's objects move at different speeds. Recognition issues increases because moving objects in the scene may not fill the entire frame.

To determine the size and location of target objects seen in videos or images, object detection technology has been effectively used. A number of applications have emerged, primarily for surveillance, security cameras, cancer detection, self-driving automobiles and other fields have all increased in frequency. Another aspect of image processing that can be performed is object tracking, this includes performing individual identification while monitoring the locations of recognized objects over the period of time. Moreover, in order to tracking objects, object class and location must first be determined using object detection in static images. Consequently, it can be concluded that the performance of the object detection should have a significant impact on the outcomes of object tracking.Wide-ranging applications have successfully utilized this object tracking technology, including the monitoring of targeted passengers, moving vehicle activity, accident monitoring in traffic cameras, local security and criminal activity monitoring, etc. In this research, a case study on the analysis and control of traffic conditions by automatic object identification has been conducted in the field of traffic control.

Recognizes that a technology for the self-driving vehicle to identify on-road vehicles has been created [1]. This system recognizes a moving object and categorizes the type of moving object using a Convolution Neural Network (CNN). Modifying the tracking centre point to match the position of the identified vehicle object on the images, the algorithm for monitoring moving objects in vehicles follows moving objects. After that, the system determines the distance between the driving vehicle and the observed vehicle objects and displays a localized image on the monitor that appears to the taken from a bird's viewpoint. The self-driving system is supported by this system procedure's objective view of a vehicle object's position. Another deep learning-based detection method for monitoring moving vehicles on highways or urban roadways by satellite was developed in collection with Support vector machine (SVM) and CNN [6]. The purpose of deep learning-based object detection was identify instances of objects corresponding to predefined classes using previously trained deep learning models. Pre-processing, inference, and post-processing are the

Three sections of the deep learning-based object identification application. The image data is gathered from various video streams during the pre-processing stage and resized to match the deep learning model's pre-defined input shape. At the inference stage, the resized data is input to the pre-trained deep learning model, to extract a set of bounding boxes containing the positions and types of objects predicted to be present in the captured image. The post-processing phase of the method then modifies and combines the bounding box data to increase final prediction accuracy. It is important to note that multi-core Central processing units (CPUs) perform the pre- processing and post-processing stages, whereas hardware accelerators process the inference phase. The position and kind of the vehicle appear to have been more precisely detected by the faster R-CNN at that moment. To put it another way, an algorithm-based technique of object recognition is better than the deep learning. In this analysis, object detection-based monitoring systems are used in all of the development cases to gather traffic information. It is dangerous to drive in a road tunnel due to the lack of space for movement in comparison to regular highways. In the case of a tunnel emergency, drivers should be alerted [2]. An efficient R-CNN deep learning model was used for training as well as a deep learning model for tunnel CCTV accident detection to solve the mentioned problems. Also, the model on which this one came acquiring from image datasets that included some tunnel accident instances. Then, only Car objects are tracked by ODTS object tracking function, which, using the Car accident detection algorithm (CADA) is regularly utilized to determine Stop and Wrong way driving (WWD) events. The original image can be used as input for convolutional neural networks (CNN) without any previous image processing. The convolutional neural networks (CNN) let use the original image as an input without having to process it first. In order to increase accuracy, researchers use parameters to improve the CNN model structure. The majority of the improved models require extra training and testing time. Since CNN models use a lot of data, the training images can't be too few. In this analysis, deep learning neural network-based tracking and background subtraction-based object detection are both utilized, because they have many advantages. The remaining sections are arranged as follows: The literature survey is discussed in Section II. In Section III explains Design And Implementation Of DL Based Moving Object Detection And Tracking With Canny Edge Detection. Section IV includes the experimental result analysis and finally paper concludes with Section V.

II. LITERATURE SURVEY

Huang, M and Yen S, et al., [19], demonstrate a computer vision system for monitoring traffic that is real-time and color-based. Among the many options for object detection, these are optical flow, the frame differential approach, background removal, and frame differential. The threshold function used in the frame differential approach, determines the changes after comparing the absolute differences between the following video frame comparisons. This technique is designed to recognize specific moving or certain pixels in an image. The issue of producing images with spot noises also occurs with the frame differential methods described when the threshold is set incorrectly. In order to distinguish moving objects from the static background, background subtraction is performed. This technique creates a background model using statistical data from pixels in video frames or earlier image information. The frame pixels are made up of two primary parts. Background pixels make up the first component with the most variance, whereas foreground pixels make up the Second. The background model has to be changed after the components have been matched.

J J Hyeok, L Myung-jae and Young-guk Ha, et al., [7] explains an integrated convolutional neural network-based learning system for object detection from images. It is intended to implement deep learning neural network architecture along with an online AdaBoost framework, which has the ability to learn multi-level features from a collection of auxiliary images. The reduction of hard negatives is performed by the use of an object proposal network with bounding box candidates. The use of CNN to organize an object identification system that learns and improves from images is proposed. The proposed learning framework, which automatically collects images into ordered classes and learns images with high accuracy, can be shared by multiple On-Board PCs. They present a fast, fully parameterizable Graphics Processing Unit (GPU) implementation of Convolutional neural network variations.

He, X., Ren, K., Zhang, S., and Sun, J. et al. [9] Deep learning method could be used to address the issue of object recognition and classification. To recognize hundreds of object classes, they trained millions of datasets. With their proposed deep learning method, most of them were successful in achieving accuracy rates of greater than 80% for both 3D and Red, Green, Blue(RGB) descriptors. Explains that a deep learning system can recognize objects in Image Net Classification with a level of accuracy that exceeds human ability. They can currently create and use a deep learning system in a mobile device due to technological advances (e.g. Tensor flow in Android and CoreML in iOS).

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. et al.,[10] uses deep learning to address the issue of scene recognition and achieved accuracy with more than 7 million datasets. Deep learning has recently achieved significant advances in computer vision, particularly in the areas of object detection and categorization. Deep learning techniques, sometimes known as the "black box approach," significantly enhance the problem of object detection and classification. Using a large number of images, deep learning allows the learning architecture to learn the significant aspects that identify the object.

J Gu et al.,[8]presents the most current convolutional neural network advancements.Deep learning is necessary for many applications due to its performance, including language processing, speech recognition, and visual recognition. Among the various kinds of deep learning neural networks, convolutional neural networks are the ones that are used and studied the most. During early days, there was a challenge in training convolutional neural networks without overfitting because there was a lack of training data and insufficient computational capability of systems. Training and validating convolutional neural networks became simple as a result of the growing amount of annotated data and the power of graphics processors, which attracted researcher's interest and allowed them to produce advanced results on a wide range of tasks.

B. Karasulu, S. Korukoglu, et al., [14] describes the use of the processed background subtraction approach in videos for moving object detection and tracking. A network of five convolutional layers and 60 million pixels, and deep learning was used to train the original Red, Green Blue(RGB) pixel space depth convolution neural network model, which has 650000 neurons, to turn into the best at arranging pictures on ImageNet. This model was 100 percentage points more accurate when using traditional manual design features than the runners-up.

III. DESIGN AND IMPLEMENTATI ON OF DL BASED MOVING OBJECT DETECTION AND TRACKING WITH CANNY EDGE DETECTION

In this paper, Design and Implementation of DL Based Moving Object Detection and Tracking with Canny Edge Detection is presented. The Fig. 1 shows the architecture of presented model.

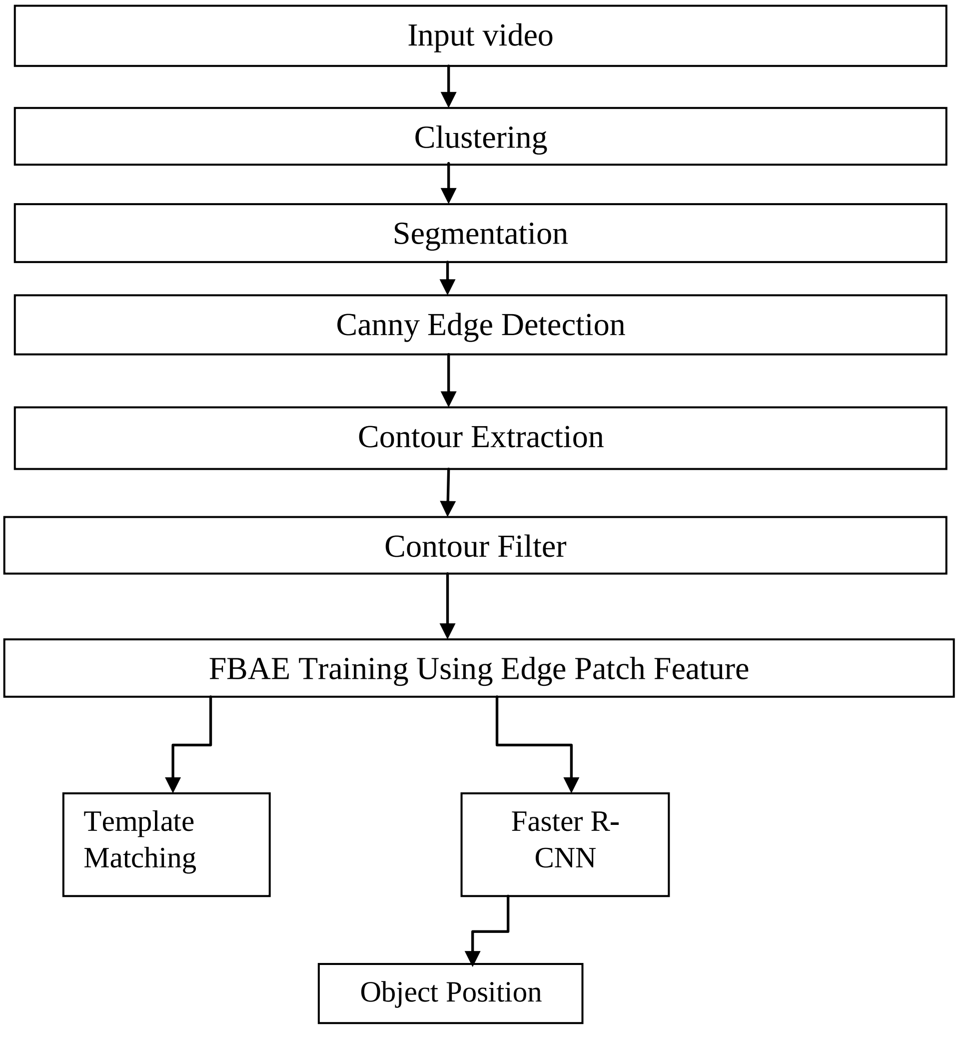

Fig.1The Architecture of Presented Model.

Fig. 1 shows the presented approach. Every moving object in the input video frames is found by the foreground detector that is included. To decide whether the distinguished moving object is a human or a vehicle, it is compressed and provided as a contribution to a previously created deep convolution neural network. If it is recognized and matches into one of the predefined classifications, it is put into a correlation filter, which will continue to follow it until it disappears from the video frame.

The vector quantization approach known as the clustering algorithm, and the attempts to cluster a number of initial input data, where the closest mean of the group to which each data point belongs to its own distinct individuals. This algorithm is a well-liked, unsupervised machine learning algorithm that is simple effective. After analyzing the dataset, this number is either determined automatically using certain techniques, such as the elbow method or the portrait analysis method, or, depending on the dataset and the needs of each research, it is manually established by users for each dataset. Centroids are chosen at random by the clustering algorithm and utilized as the starting points of each cluster in the dataset. Due to the random process of this stage, those points may represent different or same data from the dataset.

Since the threshold's size is dependent on empirical value and is not constant across a range of environmental conditions, the simple method for calculating the value is not preferable. With the explain that a segmentation algorithm would automatically choose a minimum value suitable for segmenting the contrast images into target and background, people add a segmentation technique into binary processing of a contrast image. For instance, adaptive threshold segmentation was used.

Usually, the boundaries of the target region in the generated binarization images is not smooth when the threshold value is used to segment the binarization of the difference image. Insufficient classification has left certain holes in the target area, and incorrect judgement has also left some textured noises in the background. The condition can be much improved, and the desired impact can be attained, by continuously opening and closing. However, it is important to remember that the treatment outcome is strongly impacted by the selection of an appropriate expansion/corrosion or combinations of open and close operations.

John Canny implemented the Canny edge detector. The canny detector attempts to conform to the following performance targets: Good detection: both failing to label true edge locations and incorrectly recording non-edge points should have low probabilities. Good localization: the operator's selected edge points should be as near as practicable to the true edge's centre. A single edge receives a single response. The basic steps of the Canny edge detection algorithm are categorized as follows: Using a Gaussian filter, increase the input image. Calculate the images of gradient magnitude and angle. To the image of the gradient magnitude, apply non-maxima suppression. To find and connect edges, use connectivity analysis and double thresholding.

Computer vision includes an area object detection that allows a system to look for objects in an image. Detection of pedestrians, surveillance, and industrial inspection are just uses for object detection. Taking the image of the object and then extracting its characteristics is the process of object detection. When a query images can be provided, the feature extraction technique is used on the given image and the modified query image is then searched for the features of the object that needs to be recognised. The object is considered, to be present in the query image if a suitable match is identified.

Using bootstrap samples from the training set, Floating-bagging-adaboost ensemble learning (FBAE) is used to create bagging classifiers, which are then combined to produce a bagged classifier. These classifiers' training may be done in parallel on different machines because they are completely independent of one another. Another method for combining weak classifiers to make a more grounded order rule is considered adaboost. In boosting, data sets and classifiers for the algorithm's training are progressively gathered. Through integrating the Bagging and Adaboost techniques, they create a new classifier ensemble strategy known as the Bagging-Adaboost Ensemble (BAE) approach. It combines the advantages of both approaches. Initially, BAE uses repeated bootstrap samples to create a set of new training datasets $\{X1, X2, \ldots, XB\}$ from the initial training dataset (X). The Adaboost technique is then used by BAE to create a set of classifiers $\{C1, C2, \ldots, C B\}$ that correspond to the collection of new training datasets. The output of the base classifiers are averaged to produce the final classifier. The lowest available error rate is always the main objective of applications. This new method is known as the Floating-bagging-adaboost ensemble algorithm (FBAE). After Bagging-AdaBoost, FBAE utilizes a backward technique to eliminate classifiers that have greater error rates. By using fewer classifiers overall than AdaBoost or bagging, the final FBAE classifier actually creates decreased error rates.

$C = (\sum_{i=0}^{B} Ci)/B$…….(1)

The minimum available error rate is always the main objective of applications. It is possible to post-optimize equation (1) to get the lowest error rate. For optimization, they utilize Floating Search.
The Floating-bagging-adaboost ensemble (FBAE) algorithm is the identifier they give to this original technique.

After obtaining the images with edge features, a template matching technique is performed. First, the template image's autocorrelation is determined. For template matching, the correlation response between the processed query and template images is determined and the position of the object is understood to be the point in the resultant matrix with a correlation value greater than a specific percentage of the maximum of the autocorrelation. The use of the edge detector, to extract edge attributes and the resulting utilization of connection to match layouts are the paper's essential concepts. This section primarily describes the current state of research into target identification using both conventional machine learning and deep learning.

The fundamental object identification architecture is faster R-CNN. As a result, the Faster R-CNN split into three networks in successively. The first network is the base network, which utilizes a pre-trained CNN for the classification function and is used to produce feature maps from the input images. Using the weights of a different pre-trained network on a larger dataset, this model is frequently utilized when training a classifier with limited datasets. In this technique, the local edge feature, a part-based feature representation is implemented. As a result, the patch's position and aspect ratio were randomly selected. This can be reduced for practical reasons by decreasing the edge patch's aspect ratio. Of course, it may be expanded by using a different aspect ratio, but it could be difficult to acquire a new aspect ratio and more knowledge about the topic may be needed.

IV. RESULT ANALYSIS

In this section, Design And Implementation Of DL Based Moving Object Detection And Tracking With Canny Edge Detection is presented. They demonstrate experimental results in this analysis, to demonstrate the effectiveness of the presented technique for improving deep learning object identification applications in Canny edge computing. The performance and tracking effectiveness of the presented method are evaluated. Using real-world images, this method obtains high detection rates and has a very quick target detection rate. False Alarm Rate (FAR) and True Positive Rate (TPR) are used to analyze performance.

The ratio between the total number of false positives, and the total number of both true positives and false positives in a frame is known as the false alarm rate, or FAR.

$$FAR = \frac{FP}{TP + FP} \ldots (2)$$

The ratio of all true positives to all objects in a frame is known as the True Positive Rate (TPR).

$$TPR = \frac{TP}{TP + FN} \ldots (3)$$

The true detection percentage of an object in a video is measured using the TPR metric, and the accurate detection of the objects in the image is referred to as True Positive (TP). The accurate detection of an object as negative is known as True Negative (TN). False Positive (FP) detection refers to selecting the incorrect object as positive. An incorrect object is detected as a False Negative (FN).

Accuracy: It is described as being the percentage of properly identified instances to all instances, and it is provided as

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

Recall: The recall evaluates how well the positive outcomes compare to the overall outcomes.

$$Recall = \frac{TP}{(TP + FN)} \ldots \ldots (5)$$

A weighted average of recall and precision is the F1 score. It must be one for the classification method to work well, and zero for the algorithm to perform poorly, as shown in equation 6 below.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \ldots (6)$$

From the first frame to the last frame, several moving objects enter and exit the monitor while being tracked. At certain times within the frame, some objects replay. Thus, the deep learning neural network should maintain the object information appropriately to track objects effectively. Every object detected by the deep learning neural network has a label assigned in the presented methodology. As an object leaves the frame, its label is removed. Classifier-based evaluation and detector-based evaluation are the two primary categories of object detection evaluation methods. The correct result of the input image must be known by both evaluation methods, and the effective solution must be contrasted with either the classifier's predictions or the object detector's identification.

Table.1: Performance Analysis

Classifiers	Accuracy (%)	Recall (%)	F1-Score (%)
SVM	65.25	69.63	75.36
Deep Learning	75.12	78.20	81.92
R-CNN	93.35	94.82	96.81

The Fig. 2 shows the Accuracy comparison between the SVM, DL based approach and presented R-CNN approach.

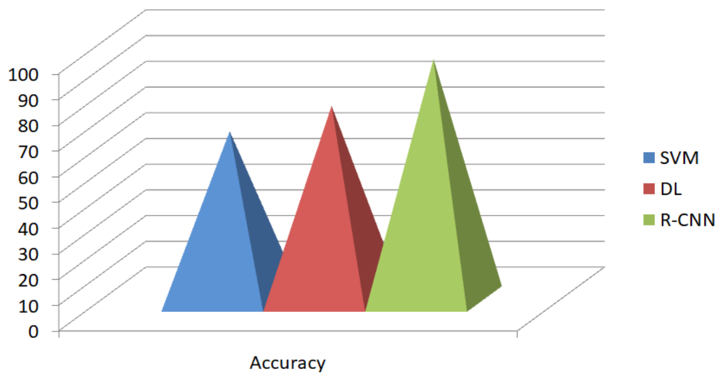

Fig. 2: ACCURACY COMAPARISON GRAPH

The Fig. 3 shows the Recall comparison between the SVM, DL, based approach and presented R-CNN approach.

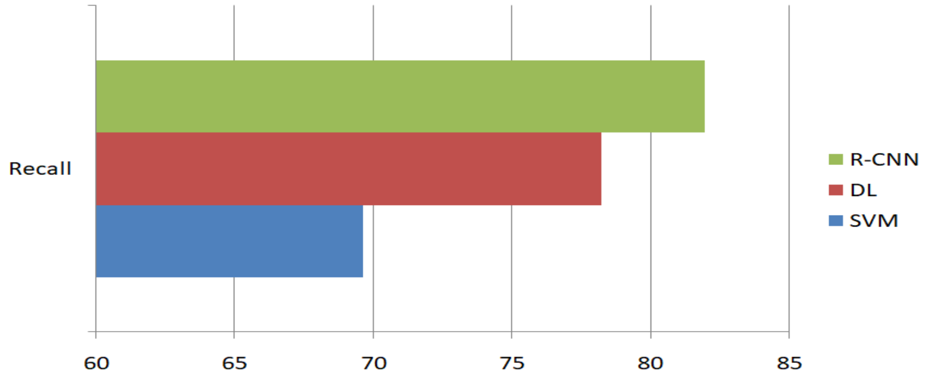

Fig. 3: RECALL COMAPARISON GRAPH

The Fig. 4 shows the F1-Score comparison between the SVM, DL, based approach and presented R-CNN approach.

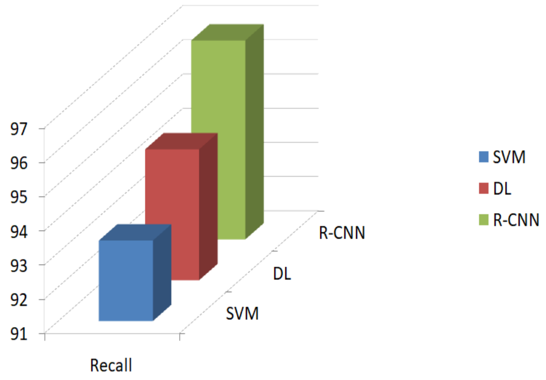

Fig. 4: F1-SCORE COMAPARISON GRAPH

V. CONCLUSION

In this section, Design And Implementation Of DL Based Moving Object Detection And Tracking With Canny Edge Detection is presented. It has been demonstrated that object tracking is an essential step in computer vision for traffic analysis, public safety, and video surveillance. It has been observed that video surveillance systems include two related parts: object detection and object tracking. Before addressing more difficult tasks like tracking, The initial phase is reported to identify objects in videos. This analysis indicates utilizing a deep learning neural network to detect and track moving objects. The recognition accuracy of deep learning neural networks makes them useful for transfer learning. As long as the object tracking algorithm does not depend largely on the way the objects are identified, it might not be as crucial to have continuous object detection performance. Using well-known datasets, the proposed method is tested for its performance in object tracking, and the results are then evaluated using probabilistic methods. The outcomes demonstrate that the proposed approach is successful at tracking moving objects. The presented R-CNN architecture performs better in terms of F1-Score, recall, accuracy, when compared to other based architectures.

VI. REFERENCES

[1] E. S. Lee, W. Choi, D. Kum, “Bird’s eye view localization of surrounding vehicles :Longitudinal and lateral distance estimation with partial appearance,” Robotics and Autonomous Systems, 2019, vol. 112, pp. 178-189.

[2] K. B. Lee, H. S. Shin, D. G. Kim, “Development of a deep-learning based automatic tracking of moving vehicles and incident detection processes on tunnels,” Korean Tunnelling and Underground Space Association, 2018, vol. 20, no.6, pp. 1161-1175.

[3] A. Arinaldi, J. A. Pradana, A. A. Gurusinga, “Detection and classification of vehicles for traffic video analytics,” Procedia computer science, 2018, vol. 144, pp. 259-268.

[4] Mengjie Hu, Zhen Liu, Jingyu Zhang and Guangjun Zhang, “Robust object tracking via multi-cue fusion”, Signal Processing, vol. 139, Oct.2017, pp. 86–95.

[5] Tianming Liang, Xinzheng Xu and Pengcheng Xiao, “A new image classification method based on modified condensed nearest neighbor and convolutional neural networks”, Pattern Recognition Letters, vol. 94, July 2017, pp-105-111.

[6] L. Cao, Q. Jiang, M. Cheng, C. Wang, “Robust vehicle detection by combining deep features with exemplar classification,” Neurocomputing, 2016, vol. 215, pp. 225-231.

[7] Hyeok J J, Myung-jae L and Young-guk Ha, “Integrated Learning System for Object Recognition from images based on Convolutional Neural Network”, In Proceedings of International Conference on Computational Science and Computational Intelligence, 2016, pp. 824-

828.

[8] Gu J, "Recent Advances in Convolutional Neural Networks", CoRR, vol. abs/1512.07108, 2015.
[9] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).
[10] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. In Advances in neural information processing systems (pp. 487-495).
[11]. Peng. S, "Flow detection based on traffic video image processing", Journal of Multimedia, vol. 8, no. 5, Oct 2013, pp.519-526.
[12]. Mao S et al., "Rapid vehicle logo region detection based on information theory", Computers and Electrical Engineering, vol. 39(3), 2013, pp.863-872.
[13] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
[14] B. Karasulu, S. Korukoglu. Moving object detection and tracking by using annealed background subtraction method in videos: Performance optimization[J]. Expert Systems with Applications, 2012, 39(1), pp:30-35.
[15]X.X. Niu and C.Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits", Pattern Recognition, vol. 45, 2012, pp. 1318-1325.
[16] Dan C.C, Ueli Meier, Jonathan Masci, Luca M.G and Jurgen S, "Flexible, High Performance Convolutional Neural Networks for Image Classification", Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, vol. 2, 2011, pp.1237-1242.
[17]David S. Bolme J. Ross Beveridge Bruce A. Draper and Yui Man Lui, Visual Object Tracking using Adaptive Correlation Filters, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010, pp.2544-2550.
[18]. F.H. Lin, Y.T. Zhou, X.S. An, I.You, K.R. Choo, Fair Resource Allocation in an Intrusion-Detection System for Edge Computing: Ensuring the Security of Internet of Things Devices, in IEEE Consumer Electronics Magazine, vol. 7, no. 6, 2010, pp: 45-50. doi: 10.1109/MCE.2018.2851723.
[19] Huang, M and Yen S, "A real-time and color-based computer vision for traffic monitoring system", IEEE International Conference on Multimedia and Expo (ICME 2004), vol. 3, 2004, pp. 2119–2122.
[20]. Hu W, Tan T, Wang L,. A visual surveillance of object motion and behaviors[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2004, 34(3): 334-352.