JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A DEEP LEARNING ALGORITHM APPLICATION FOR AUTOMATIC DETECTION OF UNPREDICTABLE ACCIDENTS

Dr.K.Thenmozhi[1], A.Thananjeyan[2], M.Gokul Raj[3] ,S.Santhosh Kumar[4]

[1]Associate Professor, [2]Student, [3]Student, [4]Student, Department of Information Technology, Dr. N.G.P Arts and Science College.

Department of Information Technology, Dr. N.G.P Arts and Science College

Coimbatore, India

Abstract: This study will introduce and apply the Object Detection and Tracking System (ODTS) for automatic detection and tracking of unexpected events on CCTVs in tunnels, which are likely to include (1) Wrong-Way Driving (WWD), (2) Stop, (3) Person out of vehicle in tunnel, and (4) Fire. Additionally, introduced and used will be the Faster Regional Convolution Neural Network (Faster R-CNN) for Object Detection and the Traditional Object Tracking technique. Using the Bounding Box (box) findings, the Object Detection results for Bounding Box (box) are collected. In order to provide each moving and identified object a distinct ID number, it then compares the Box of the most recent and previous video frames. With this method, a moving object can be tracked in time, which is not currently possible.

Index term— a faster R-CNN for object detection, object tracking, object detection and tracking, unexpected event detection, and tunnel CCTV accident detection

I. INTRODUCTION

Finding the size and location of target items in still photos or moving movies has been made possible with the help of object detection technologies. Applications for self-driving cars, CCTV security systems, cancer detection, etc. have all become more prevalent. Another aspect of image processing that can be accomplished is object tracking, which involves tracking the positions of specified objects over time and performing unique identification. However, in order to track objects, it is first essential to establish object class and location in a static picture that has been provided. Therefore, it can be claimed that the effectiveness of the object detection used should have a significant impact on the outcomes of object tracking. This object tracking technology has been effectively applied to a variety of tasks, including the tracking of a targeted pedestrian and a moving vehicle, accident monitoring in traffic cameras, monitoring of local crime and security concerns, etc. This research conducts a case study in the realm of traffic control about the analysis and management of traffic conditions by automatic object detection. These summaries are provided. [1] Claims that a self-driving automobile system for on-road vehicle detection has been created. This system recognizes moving objects and uses Convolutional Neural Networks (CNN) to categorize the different types of vehicles. By adjusting the tracking centre point in accordance with the location of the detected vehicle object on the image, the vehicle object tracking algorithm tracks the vehicle object. Distance between the car's driver and the cars you can see in the image. This process of the system enables to objectively view the object's location inside the vehicle for the advantage of the self-driving system. It can therefore locate the vehicle object at the camera to within a tolerance of 1.5 m in the vertical and 0.4 m in the horizontal. In conjunction with CNN and Support Vector Machine (SVM), a further deep learning-based detection system was developed in [2] to track moving autos on highways or city streets via satellite. This system finds the vehicle Box by conducting binary classification using SVM using the satellite picture as an input value and CNN to extract the feature. Additionally, a method for determining vehicle speed, classifying vehicle types, and assessing traffic volume was developed by Garibaldi, Padano, and Gurus Inga [3].Box, which was found by object recognition in videos or photographs, is used by this method. The system's algorithm was compared to the faster RCNN and the Gaussian Mixture Model + SVM. The speedier R-CNN at that moment seems to have detected the position and kind of the vehicle with higher accuracy. In other words, it may be argued that the algorithm-based object detection system is inferior to the deep learning-based approach. The object detection-based monitoring systems used in this work's development examples to gather traffic data all exhibit excellent deep learning performance. However, they were all difficult to implement in terms of giving the found items unique IDs and tracking them by keeping the same ID throughout time. In order to create an object detection and tracking system (ODTS) that can track moving information about the target objects, object tracking algorithm and deep learning-based object detection method are combined in this study. In the part that follows, the complete ODTS methods (Figure 1) will be thoroughly explained. Additionally, the tunnel accident detection system within the ODTS framework will be taken into account [4, 7]. This method is designed to track a certain local area on CCTV and find accidents or unexpected events that happen to moving objects.

II. OBJECT DETECTION AND TRACKING SYSTEM BASED ON DEEP LEARNING

A. CONCEPT

The process of object recognition and tracking by the ODTS through time is depicted in Figure 1[7]. It is believed that ODTS has received sufficient training to correctly detect objects in a given image frame. The trained object detection system provides sets of coordinates for objects on the provided picture frame at the time T to the object detection system (ODTS), which gets the selected video frames at the present time interval (c). The associated type or class of each identified object

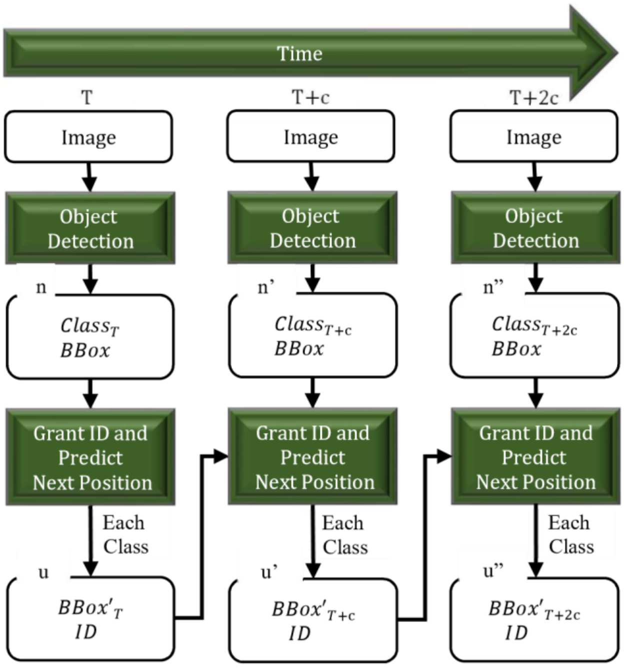


Figure 1. Object detection and tracking process of Object Detection-Tracking system over time. When class and BBox were obtained by object detection, object tracking algorithm grants ID and predicts next position using current and past BBox.

The object detection module simultaneously classifies the object. Then, based on the detected object information, a dependent the object tracking module is started in order to give each of the observed items a special ID number and forecast their future positions. Tracking BBox u has a different number than tracking BBox n. But the number of tracking BBox equals the total number of identified items if the past tracked BBox value is 0. For instance, if u is 0 in time T+c, u' equals n'. In other words, the present tracking BBox pulls from the identified items for each class while the former tracking BBox was absent. This object tracking module was built by introducing the SORT algorithm [5], an object tracking technique that employs the idea of Intersection Over Union (IOU) to trace identical objects with identical paths. The same object detection module employed at the time T is utilized to acquire and C on the freshly provided image at the subsequent time step Tic. Following that, IOU of all potential pairs between anticipated locations, ′ at time T, and detected object positions, at time T+c, are calculated. It will be considered that the closest items, or the pair with the highest IOU value, are the same object with the same ID. And any object in ′ that does not have an object pair with an IOU value greater than 0.3 will be regarded as having vanished from the area of interest (ROI). Similar to this, any item in that does not have an object pair with an IOU value greater than 0.3 will be regarded as newly created.

For object detection, this system uses a quicker RCNN learning algorithm [5], and for ID assignment and object tracking, it uses a SORT [6]. Sort [6] is known to support multi-object tracking 100-300fps degree speed. These system processes object Video frame interval c [7] had an impact on the capacity to track objects using the SORT [6] method, which is dependent on IOU value. By changing the object detection network's detection interval, video frame interval can gradually minimize the calculation amount. It was feasible to monitor the objects for up to six frame intervals [7] after experimenting with object tracking ability across the frame interval. The video frame interval should be optimized for the number of camera devices connected to a deep-learning server at once because increasing frame interval drastically affects object tracking performance.

B. TUNNEL ACCIDENT DETECTION SYSTEM

The ability to track objects utilizing the IOU-dependent SORT [6] approach was impacted by the video frame interval c [7]. Video frame interval can gradually reduce the calculation amount by altering the object detection network's detection interval. After experimenting with object tracking capability throughout the frame interval, it was possible to keep an eye on the objects for as many as six frame intervals [7]. Since increasing frame interval has a significant negative impact on object tracking performance, it is important to adjust the video frame interval for the quantity of camera devices connected to a deep learning server at once.

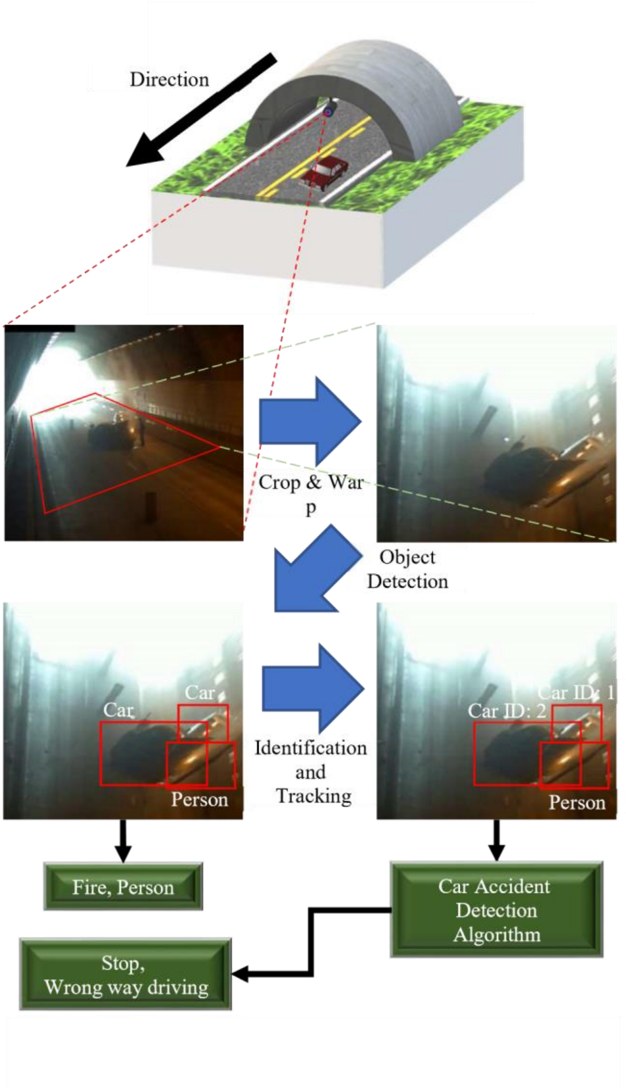
In the meanwhile, CCTVs are used in tunnels to monitor the target items and unexpected incidents. And with outstanding performance outside of tunnels, an automatic object detecting system for the targets would be employed for the purpose. In a tunnel, however, the mechanism is completely useless. It's because: (1) the tunnel video has poor illumination, thus the video was heavily influenced by the tail light or warning light of the moving car. (2) The tunnel video had a gloomy colour tone. In contrast to the road of the tunnel outside, it is a distinct colour. In contrast to the road of the tunnel outside, it is a distinct colour. Due to the two aforementioned factors, it seemed anticipated that the video surveillance system used on the roadways outside of the tunnels would not function properly inside of them. It is necessary to have an automated accident detection system that is tailored for road tunnels.

A deep learning-based Tunnel CCTV Accident Detection System was created in [7] to address the aforementioned issues. Faster R-CNN deep learning model was trained using this method. Additionally, the model on which this one was based learned from image datasets that included some tunnel accident instances. Then, only Car objects are tracked by ODTS's object tracking function, which is periodically used to determine Stop and WWD events using the Car Accident Detection Algorithm (CADA).

As seen in Figure 2, the CADA process finds an accident state. Region of Interest (RoI) should first be assigned to the CCTV screen in the tunnel, and then the extracted image should be cropped and warped using the ROI of the CCTV screen's original image. Although this process is similar to [1], its goal is to establish a consistent method for classifying Stop and WWD events. The picture extract makes it easier to train on the image unnecessary by creating objects that are the same size on the near and far sides, image region outside of ROI. These distinctions from [1] are made. Next, use a faster RCNN to recognize items like cars, fires, and people [5].

Next, an Extra 'No Fire' object was established by directly creating the object class to reduce the wrong answer for the Fire object in order to prevent erroneous detection of the Fire object. The No Fire object is designated for deceptive items like car taillights, tunnel lights, etc. Faster R-CNN training distinguishes the pre-defined object class from the background exception by taking into account the feature of those data. This approach might make it possible to reduce Fire misdetection on untrained data.

The Car object is then assigned a number by an object tracking algorithm, and at a predefined time, the current cycle BBox is compared with the cycle BBox from the previous cycle with the same ID. Image 2. Using tunnel CCTV to detect accidents.



With CADA, object detection and tracking can detect fire, person accidents, and wrong-way driving accidents over time. At this time, Stop was selected by IoU, and WWD was determined by Intersection over Line (IoL). IoU and IoL are concepts that refer to the ratio of overlapping lines. To determine inversion, use only the vertical value of BBox, as illustrated in the equation below.

$$\text{IoL} = \frac{h}{\qquad} \quad (1) \quad h$$

In order to determine the vertical velocity of the Box in the image, the WWD criterion is based on the warped tunnel CCTV image. IoL is said to be inverse if it is less than 0.75 and the vertical driving direction is the opposite.

The size and position of the Box should be taken into account regardless of the direction, hence IOU was utilized as a criterion for identifying the Stop event. It is considered a stop if it is 0.9 or higher.

The tunnel CCTV video accident detection system is used for an experiment with real object images and event videos received from tunnel monitoring sites through the process as indicated above.

III. EXPERIMENTS

The learning performance measurement of the constructed system and the system's overall performance in accident detection are the two main areas of experimentation in this work. Performance of object recognition has a significant impact on the SORT utilized in ODTS. Therefore, high performance object recognition by effective deep learning object detection network learning was needed to finish this system.

The complete system was then put to the test to determine if it can recognize the specified four accident events using the trained deep learning model. The system was evaluated for each image to see if it could detect each circumstance because in this scenario, both the object detection performance of the deep learning model and the discriminative ability of the CADA were necessary.

A. Deep learning training

Instead of using a movie to train the deep learning network, still photos were used instead. One epoch in this study serves as the definition of a training cycle for a complete dataset. The photos from accident events are part of the dataset to be learned. R-CNN [5] was trained using a faster version.

TABLE I. THE STATUS OF USED IMAGE DATASET

Number of Videos	Number of images	Number of objects		
		Car	*Fire*	*Person*
45	70914	427554	857	44862

The training dataset's current status is displayed in Table.1. By separating 45 videos into frames, this collection is made up of 70,914 video pictures. In contrast to the normal deep learning procedure, the deep learning training process did not separate learning data from inference data. This is because the dataset, in contrast to the publicly accessible datasets, used in this paper is that the images are continuous in each video. In other words, the images present in each video file have a backdrop image that is the same overall but vary according to the presence of objects. The object detection network's inference performance would exhibit identical performance whether the training data and the inference data were split up for each image. On the other hand, the stability of object detection over the full video may decline, which negatively impacts the accident's performance in terms of detection, making it challenging to evaluate the accident detection system's entire tunnel CCTV image recognition process. As a result, training involved gathering all accessible data, and learnt data are used to evaluate the efficacy of deep learning object detection.

The small amount of Fire objects is due to the rarity with which fire incidents occur in the tunnel. As a result, there is a considerable likelihood that a fire may be falsely detected or missed, and it is crucial at the tunnel control centre that false detection be less often than missed detection.

The reliability of the system is significantly reduced if the system that is put in the field is regularly informed that false detection has occurred even though there is no evidence of false detection. On the other hand, if the data was not detected, it was still able to use the enriched dataset in the time-lapse that is periodically added to the training dataset to automatically enhance the detection performance. Therefore, the experiment's main goal was to make sure that fewer false positives were made and that there were much more No Fire items than Fires.

R-CNN training advanced quicker with 10 epochs. Tensor flow 1.3.0 was the deep learning framework running on Linux [7]. Faster R-CNN training employs NVidia GTX 1070 as its graphics card. The training period lasted 60 hours, and average precision (AP) was used to assess each object class's inference performance.

TABLE II. INFERENCE RESULT OF DATASET

Number of images	Average Precision (AP)		
	Car	*Person*	*Fire*
70914	0.8479	0.7161	0.9085

The AP values for the three target objects may be seen in Table 2.. The largest item in the training dataset is a car, and when compared to other classes, the car object's AP value is extremely high. In other words, it was anticipated that the Car's deep running object detection performance would be quite trustworthy. On the other hand, because Person object has a long, tiny shape and is modest in size, AP for Person object in Table.2 results in a relatively low value.

The AP of the Fire item was as high as 0.9085, but because to the extremely low (857) number of training objects, incorrect identification of the object may be very likely.

B. Accident detection test using entire Tunnel CCTV Accident Detection System

The performance of the deep learning-based Tunnel CCTV Accident Detection System needs to be assessed in terms of accident detection based on the trained deep learning model. In order to do this, 4 movies were chosen and checked for the 4 occurrences listed in Table 3. To visualize the results of the detection on the video, a visualization application was created.

It was determined that it was recognized within 10 seconds of visual observation when the video frame interval was set to 6 frames per second at 30 frames per second [7]. Table 3 provides a summary of the video's length, its occurrence time, and its detected time.

TABLE III.

TIME OF EACH ACCIDENT AS DETAILED BY ACCIDENT

Accident video information	Item on video time		
	Video length	*Occurrence time*	*Detected time*
Stop	126s	5s	7s
Wrong Way Driving	29s	4s	12s
Fire	64s	29s	29s
person	72s	50s	50s

There is a delay between the Stop and WWD events in Table 3 and when they are detected. This is a characteristic of CADA, and in our experiment, it occurs every 2.4-sec cycle. The system, however, was able to distinguish between the Stop difference of 2 seconds and the WWD difference of 8 seconds. In contrast, images of the collision scene, including Person and Fire, showed swift detection. The visuals used in Table 3 were solely meant to be used for educational purposes, hence they are inappropriate for use in actual field installations. More test movies and the testbed application were consequently required.

IV. CONCLUSION

This research suggests a new method for ODTS that combines an object detection network powered by deep learning with an object tracking algorithm. It demonstrates how dynamic object information for a particular object class may be gathered and used.

On the other hand, the performance of object detection is crucial since SORT, which is utilized in ODTS object tracking, only takes data from Box and does not employ an image. Therefore, unless the object tracking technique is significantly reliant on object recognition performance, continuous object detection performance may not be as necessary.

Additionally, a Tunnel CCTV Accident Detection System based on ODTS was developed. The experiments on deep learning object identification network training and evaluation as well as system-wide accident detection were carried out. This system also includes CADA, which makes decisions about each cycle based on dynamic data about the car's objects. It was feasible to identify the incidents within 10 seconds after testing with the image that contained each incident.

However, deep learning training secured the object recognition performance of a trustworthy Car object, but Person displayed comparatively poor object detection performance. Due to the scarcity of Fire objects in the untrained movies, there is a high likelihood of false detection in the case of Fire. Nonetheless, by simultaneously training objects that are No Fire, it is possible to decrease the incidence of erroneous detections. Securing the Fire image later should enhance the deep learning object detection network's performance in detecting fire objects.

Although the ODTS can be used as an example of a Tunnel CCTV Accident Detection System, it can also be used in other industries that need to track the dynamic movement of a particular item, such as estimating vehicle speed or tracking illegal parking. Securing diverse photos as well as Fire and Person objects is required to improve the system's dependability. Additionally, the system's dependability might be increased by the use of the tunnel management site and ongoing monitoring of it.

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] 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.

[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] K. B. Lee, H. S. Shin, D. G. Kim, “Development of a deep-learning based automatic tunnel incident detection system on cctvs,” in Proc. Fourth International Symposium on Computational Geomechanics, 2018, pp. 140-141.