Object detection YOLO-v4 based smart car system with an integrated Virtual Assistant

Vignesh B R\textsuperscript{1}, Sandhip Kumar D\textsuperscript{1}, Gokul V\textsuperscript{1} and Priya R\textsuperscript{1}
Department of Electronics and Communication Engineering, Sri Krishna College of Technology, Coimbatore, India, 641042
r.priya@skct.edu.in

Abstract. Car accidents are a serious social problem that often results in both life loss and financial loss. Most car accidents are caused by a lack of safe distance between cars and also the awareness of the driver. To relieve this issue, in this paper we propose a real-time object detection and safety system. The proposed system consists of two units: a real-time object detection unit and a safety alarm unit. The system is supposed to apply in a normal driving scenario. As for the safety alarm module, it consists of three states: to detect the object; to calculate the safety factor; to determine the driving conditions. To justify the proposed system, a real experiment is conducted. The results show that the system can appropriately signify driving states: safe, dangerous, and warning. By the given experimental results, it implies that the system is feasible and applicable in real-time applications. The objective of this project is to detect objects in real-time and pass information to the user based on the data taken by object detection. The information can be provided to warn the user or to just inform the user based on the result of the sensors. The YOLO algorithm helps to design the output of the sensor based on the necessity of the user. The google assistant allows the user to provide voice commands. The car is integrated with all the above-mentioned.

1. Introduction

The automotive industry has been undergoing a lot of changes and has improved exponentially. A smart car is an artificial intelligence (AI) platform that can develop from its surroundings and experience to improve its skills. The extent of object detection is explored and implemented in real-time, for ensuring the safety of fellow humans. The proposed methodology has an efficient real-time object detection system using the YOLO network. You Only Look Once (YOLO) is a neural network for target recognition that employs deep learning (DL) algorithms. The latest version of the YOLO which is the YOLOv4 has been implemented for optimal speed and accuracy to recognize objects more precisely and faster than other recognition systems. Convolutional neural networks (CNNs), a form of deep learning algorithm, are used effectively. The execution of convolutional neural networks, work on both existing datasets and real-time datasets. The integration of virtual assistants is used as a way to interact with the provided system to have almost instantaneous voice feedback about the data and values the user needs. The intelligent system will assist humans in eliminating manual labour in day-to-day jobs as well as the possibility of losing precious human lives in dangerous circumstances. The Existing system uses the Tiny-YOLO-v3 for object detection. Tiny-YOLO-v3 employs a wide number of convolutional layers and filters, resulting in a large number of parameters, high storage requirements, and poor detection time in compressed settings.
Another issue with Tiny-YOLO-v3 is that its detection precision is low, and its detection accuracy is reduced as a consequence of the network's excessive compression techniques [1]. There are no sensors for distance measurement of the environment that includes real-time detection of the object. The processing time is slow because of the relatively weak graphical processing unit used for object detection.

2. Proposed Prototype

In order to solve the issues with Tiny-YOLO-v3, this paper proposes YOLO-v4, which emphasizes mostly on network efficiency of the model scale, detection speed, and accuracy. The intention of YOLO-v4 is to create a model that is smaller, quicker, and finer and can operate in restricted environments. To begin with, the network configuration is configured by reducing the amount of parameters in a reasonable manner rather than simply eliminating the hidden layer. The virtual assistant is one of the most used AI (Artificial Intelligence) in everyday life by people. As the virtual assistant has been normalized it is better to use that feature and embed it with the system, which opens a wide range of possibilities, that can involve many things from voice commands to displaying certain information to providing an interface for the user to entertainment purposes to getting real-time information. The possibilities of using a virtual assistant are limitless. But the previously existing systems have failed to implement that part of the system Figure 1.

The following are the elements in the system:

A. Sensor perception

B. Object Detection Phase

C. Interactive Virtual Assistant

![Figure 1. Block diagram](image)

Figure 1. Block diagram

A. Sensor Perception

Sensor nodes and cameras are used to determine an object's surface morphology, measure its location, and calculate its distance. To use ultrasonic sensors to calculate the velocity and structure of an object, organise certain sensor nodes parallel and then position the attribute within the detection region of these sensor nodes. The surface configuration of an entity may be amended and the acceleration of the object determined using data collected from all sensors.
B. Object detection phase
Supervised learning software is used at this point. Using the Open CV-library and Darknet, the software is written in Python. We may send the subject to the device in two forms for detection: first, directly by scripting, and second, through the camera. In this case, the datasets are preloaded on the file and the extraction is made from the preloaded images. The aim is to create a model that can identify the protest of defined differentiation using open-source hardware and that attacks the concept of visual details captured from a normal camera with fair lucidity. Using OpenCV and Darknet, the suggested measurement is carried out on the Raspberry Pi.

C. Interactive virtual assistant
Google Assistant is also integrated into the raspberry pi module which uses speech to answer queries and respond to tasks [2]. A speaker is connected to the module to enable audio streaming. Dynamic interactivity enabled.

3. Methodology of Virtual Assistant and Voice Recognition
The immersive platform used for Virtual Assistants is extremely sophisticated. They comprehend not only the vocabulary, but also for the context of the user's words. They are able to deduce meaning from situations, and their actions are unpredictable. They'll be able to get a longer personal interaction this way. They could also be programmed to do more complex tasks. Digital assistants will also comprehend slangs used for the everyday natural communications and analyse emotions using languages, enhancing an even greater range of interactive skills. Virtual assistants are more engaging than chatbots thanks to NLP. Via hard-coding, wildcard word matching, and time-consuming phrase preparation, virtual assistants can understand conversations, provide strong NLP capabilities, and carry out a small range of conversations. It's seen in activities like judgement and e-commerce as well. It can do things like share news, play music, provide ETA feedback, and even monitor digital devices. Apart from chatbots, virtual assistants improve over time. If citizens get used to using voice search and conversing with smart speakers, the possibilities become almost limitless. When we use smart speakers, our requests get longer by default, because we want to converse. This will be much more true as technology progresses and the consistency of the responses to our queries grows smoother and more accurate. In the meantime, whether on a cell computer, smart speaker, or car dashboard panel, users utilize their virtual assistants for a variety of tasks.

A computer or program's capacity to accept and interpret dictation or recognise and carry out spoken orders is known as voice or speaker recognition. With the advent of AI and interactive assistants like Amazon's Alexa, Apple's Siri, and Microsoft's Cortana, voice recognition has gained popularity and applications. Speech Recognition, or letting the machine recognise human speech, is the first thought that springs to mind when discussing voice control. Voice Recognition is a technique that allows a machine to recognise words spoken in a spoken language. Speech is a fantastic way to manipulate sensors and communicate [3]. Algorithms based on acoustic and language processing are used to recognise voices. Language modelling matches sounds with word sequences to better distinguish between terms that sound alike. Acoustic modelling is a representation of the interaction between speech linguistic units and audio signals; acoustic modelling indicates the relationships between linguistic units of sound and audio signals. Speech recognition technologies enable users to communicate with devices simply by listening to it, allowing them to make orders, set alerts, and perform other basic tasks without having to use their hands. Analog-to-digital transfer, also known as voice recognition applications on devices, involves a digital signal which is translated from analogue audio. A machine must provide a digital dictionary of terms or phrases, as well as a quick way to compare this information to signals, in order to perceive a signal. When the software is run, the speech rhythms are saved on the local
machine and stored into memory. A comparator compares the processed patterns to the A/D converter's performance, a process known as pattern recognition.

4. Comparative Study of YOLO

Object detection is a very important task and a necessity in many fields in this world. Some of them include the following medical diagnosis, robot navigation, AR (Augmented reality), autonomous driving of vehicles, and many other fields using object detection [4]. In these crucial scenarios, object detection plays a very important role. The YOLO-v4 shows greater advantages than many other traditional methods such as ATSS (Adaptive Training Sample Selection), ASFF (Adaptive Spatial Feature Fusion), EfficientDet, and Centre Mask [5]. YOLO-v4 is among the quickest object detection systems, with high precision and decent real-time efficiency. Since its inception, the YOLO-v4 has undergone many enhancements and revisions, including those to previous iterations such as the YOLO-v1, YOLO-v2, and YOLO-v3. The YOLO-v1 model has hit 1 GB of storage capacity, which is impractical with either device, limiting the user's ability to create even more useful knowledge and requiring a high-performance operating framework to preserve the consistency and speed of the proposed model. Attempting to use the next models, which had its own set of drawbacks. The completely structures are removed in YOLO-v2, and anchor boxes are introduced to better predict anchor boxes, rendering the YOLO predictor quicker and more stable than YOLO-v1. The more the system is robust the best we can make of it, but the next version of the YOLO had much more beneficiary improvements. YOLO-v3 exploits the residual framework to penetrate further through the network layer, resulting in a significant improvement in precision. YOLO and its enhancements have a large VOLUME since they operate on the popular GPU (Graphical Processing Unit) platform. YOLO: A Real-Time Object Recognition Method for Accurate and Fast Environment Recognition. These target detection algorithms' model sizes, on the other hand, are too small for the surroundings. Tiny- YOLO-v3 seems to be the most recent version of YOLO, with a tiny sample size that works in almost any area. However, the detection precision isn't very good, and real-time efficiency on low-power systems is also slow.

YOLO-v4 is introduced in this paper as a more effective object detection model for restricted settings, derived from Tiny-YOLO-v3. It reduces the model size while maintaining high detection precision and real-time efficiency. The YOLO-v4 is much more efficient and much faster than its predecessor and has been improved to a significant amount. The only downfall of using the YOLO-v4 is the higher GPU needed to compute the data. In this paper, the usage of the Google Colab free online GPU has made the processing of the huge ocean of data's in a higher pace with minimal expenses. Google Colab has an open-source network for this GPU for the usage of this huge pool of data. This paper comprises the system description and experimental values and the methodology used and taking the course of the overall accomplishment of this system. The CNN (Convolution Neural Network) has been an integral part of the proposed system. The CNN has been widely used for object detection and has the ability to extract features automatically. The number of frames and the average rate of getting the data for the process of the detection is one of the hardships that needed to be solved in this proposed system. The object detection algorithm works hand in hand with the CNN. The object detection algorithm is split into two types depending on the convolution neural network: two-stage and one-stage detectors Figure 2.
In this proposed system we have used the one-stage detector. The two-stage detection stages are straightforward; the first is identification, which progresses to producing candidate boxes and creating predictions based on those boxes. The two-stage convolution neural network has a fixed or set rectangle to detect objects even though the scale of the objects could be very small and tiny in other words. So, the reliability of the two-stage detectors has fallen significantly for the real-time object detections. As we need a quick and reliable source for object detection as the size of the detecting area is fixed it becomes more difficult for the algorithm to remove the unwanted spaces and detect the actual object much quicker. The two-stage detectors are developed which leads to the one-stage detectors. The independent end-to-end learning and independent reforming of the frame size lead to greater success this one-stage convolution neural network was proposed in 2016. The one-stage convolution neural network makes the object detection problems, a regression problem. YOLO-v4 was exceptionally swift and sufficient as compared to the two detectors. However, it had stability problems with previous YOLO models. The one-stage detector shapes itself according to the size and shape of the objects which gives us the accuracy desired by the user. The one-stage detector has been widely used more in recent years. The combination of all these lead to the proposed system and the efficiency of the proposed has been made to its limit. The raspberry pi is the heart of the proposed system which helps to display and is used as a tool to provide the connection to the user Figure 3 and Figure 4.

5. Experimental Outcome

![Figure 2. One-Stage object Detector](image)

![Figure 3. The proposed YOLO-v4 is related to other state-of-the-art object detectors.](image)
Figure 4. Using YOLO-v4 for real-time object detection with trust scores and bounding boxes

6. Conclusion

This system has multiple features which will give more safety to the user. The distance, virtual assistant, and real-time object detection are made possible and viewed on the dashboard of the car. There are varieties of sensors for analyzing the safety factor, a lot more can also be added according to the need. The user can taste the latest artificial intelligence technology that makes life a lot more secure and comfortable than before. The Virtual Assistant System has enormous scope in the future. Like Siri, Google Now and Cortana become popular in the automobile industry. This makes the adaptation smooth to a complete voice command system. It’s all about that frictionless experience that makes our lives much easier and safer.

References

[1] Wei Fang, Lin Wang and Peiming Ren, Tinier-YOLO: A Real-time Object Detection Method for Constrained Environments, IEEE Access, Vol. 8, 24 December 2019.
[2] Septimiu Mischie, Liliana Măiu-Iovan and Gabriel G presc, Implementation of Google Assistant on Raspberry Pi, International Symposium on Electronics and Telecommunications (ISETC), 8-9 November 2018.
[3] Bharath Sudharsan, Sree Prem Kumar and Rakesh Dhakshinamurthy, AI Vision: Smart speaker design and implementation with object detection custom skill and advanced voice interaction capability, 11th International Conference on Advanced Computing (ICoAC), 18-20 December 2019.
[4] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi: 10.4018/978-1-5225-5097-6.ch010
[5] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. 27, no. 4, p. 273, 2018.