Identification of Objects in Autonomous Vehicles

R. Tejaswi, K. Navya, D. Bhavya Sree, C. Sarika Reddy and C. Anitha
Department of Computer Science and Engineering, Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh, India

Keywords: Autonomous Vehicles, Object Detection, YOLO v11, Real-Time Processing, Sensor Fusion, Traffic Sign Recognition, Obstacle Detection, Adverse Weather, Low-Light Conditions, Deep Learning.

Abstract: In order to keep tabs on obstacles, people, autos, and traffic signs in real time, autonomous vehicles rely on object detecting systems. In this study, we look at the potential benefits of YOLO v11, a real-time object detection method that has recently been enhanced, for autonomous driving systems. YOLO v11's enhanced architectures, feature extraction networks, and detecting heads allow for faster and more accurate object recognition. Apt for use in mission-critical settings, the model swiftly processes high-resolution images. When it comes to low-light and bad-weather object detection, YOLO v11 shines after extensive training on huge datasets like KITTI and BDD100K. Autonomous vehicle sensor fusion and real-time decision-making modules are also integrated into YOLO v11 in this study. The reliability and security of autonomous driving are enhanced by better detection accuracy, processing efficiency, and resilience.

1 INTRODUCTION

Safe navigation in difficult environments is now possible with the help of advanced object detecting technology. To keep drivers and pedestrians safe, these devices monitor road conditions, obstacles, and traffic signals in near real-time. Conventional object detection methods are inefficient and prone to errors in low-light and adverse weather conditions.

The improved real-time object detection model YOLO v11 is tested in this research for its potential use in autonomous vehicles. Thanks to revamped architectures, feature extraction networks, and head detection, YOLO v11 greatly enhances the speed and accuracy of object recognition. Designed for mission-critical autonomous driving systems, the model excels in a variety of scenarios employing massive datasets such as KITTI and BDD100K.

This research uses a combination of YOLO v11, real time decision-making modules, and sensor fusion to increase the reliability of detection. Autonomous vehicle navigation is now safer and more dependable in difficult environments because to improvements in detection accuracy, processing efficiency, and resilience.

2 LITERATURE SURVEY

2.1 SSD: Single Shot Multibox Detector

https://arxiv.org/abs/1512.02325
A single deep neural network is utilised for picture identification. For each point in the feature map, SSD transforms the bounding box output into a default box that may be adjusted in size and aspect ratio. In order to make precise predictions, the network sorts things into categories and adjusts the default boxes so that they suit the objects' forms. The network can automatically adapt to objects of varying sizes by using predictions from many feature maps with varying resolutions. Compared to object proposal approaches, our SSD approach is easier since it eliminates the need to create proposals and resample pixels or features, and instead uses a single network for all processing. Both training and integrating SSD into detection systems is a breeze. On the PASCAL VOC, MS COCO, and ILSVRC datasets, SSD achieves accuracy on par with object proposal methods while being faster and providing a unified framework for training and inference. Even when working with smaller input photographs, SSD outperforms other single-stage algorithms in terms of accuracy. Using a

Tejaswi, R., Navya, K., Sree, D. B., Reddy, C. S. and Anitha, C.
Identification of Objects in Autonomous Vehicles.
DOI: 10.5220/0013895800004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 224-229
ISBN: 978-989-758-777-1

300x300 input and a 500x500 input, SSD achieves 72.1% mAP on the VOC2007 test at 58 FPS on Nvidia Titan X, whereas Faster R-CNN achieves 75.1%.

2.2 Mobile Nets: Efficient Convolutional Neural Networks for Mobile Vision Applications

https://arxiv.org/abs/1704.04861
When it comes to embedded and mobile vision, our MobileNets models are top-notch. Using depth-wise separable convolutions, MobileNets' simplified design constructs lightweight deep neural networks. We provide a pair of straightforward global hyper-parameters that strike a good compromise between accuracy and latency. Based on the constraints of the problem at hand, model builders can select the ideal model size using these hyper-parameters. We surpass other well-known ImageNet classification methods after doing thorough investigations of the accuracy vs. resource tradeoff. Following this, we demonstrate that MobileNets are effective at identifying items, classifying fine-grained features, identifying faces, and performing large-scale geo-localization.

2.3 Developing a Real-Time Gun Detection Classifier

https://www.semanticscholar.org/paper/DevelopingaRealTime- Gun- Detection- ClassifierLai/fc4ecaede052bd47515fa64ea21aeea9a9b3e5fc
Convolutional Neural Networks can enhance surveillance by detecting objects in real time. Detecting pistols and guns is one use. Infrared data for hidden weapons has largely been utilised for weapon detection. In contrast, we are focused in quick weapon detection and identification from photos and surveillance data. For picture-based weapon identification and categorisation, the Overfeat network was constructed using Tensorflow. By adjusting the hyperparameters, we achieved the highest accuracy on Overfeat-3, 93% during training and 89% during testing.

2.4 Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles

https://www.sciencedirect.com/science/article/pii/S1877050 915030136
Objects in motion that can cause problems at the border between the United States and Mexico are being monitored by unmanned aerial vehicles (UAVs) as part of this endeavour. Border encroachment and immigrant trespassing pose a significant threat to US border security and DHS. No amount of human monitoring can possibly go through all the data to find suspicious activity. Intelligent visual surveillance technologies that can identify and track unusual or suspicious video events are going to be a big part of this project so that human operators can do their jobs better. Accurate and fast tracking of moving objects is essential for the visual surveillance system. Object identification and tracking algorithms for unmanned aerial vehicles were the focus of this study. For object recognition, we used adaptive background reduction. Optical flow tracking using Continuously Adaptive Mean-Shift and Lucas-Kanade followed objects in motion. The results of the simulations show that these algorithms are able to detect and follow objects in UAV footage that are in motion.

2.5 Object Detection Algorithms for Video Surveillance Applications

https://ieeexplore.ieee.org/document/8524461
Defence, security, and healthcare all make use of object detection algorithms. Improved accuracy in video surveillance applications may be achieved by modelling and developing object detection algorithms in MATLAB 2017b. These techniques include face, skin, colour, shape, and target detection. Additionally, we cover object detection issues and their potential uses.

3 METHODOLOGY

3.1 Proposed Undertaking

Autonomous vehicles are able to identify obstacles, people, and other vehicles with remarkable speed and accuracy with the help of YOLO v11. This makes it easier to travel without crashing. Because it can withstand poor light and harsh weather, it may be used anytime, anyplace. Additionally, the system enhances safety by informing the vehicle about detected objects from a distance, eliminating the need for close proximity. Furthermore, by providing accurate object recognition and decision-making, the system reduces unnecessary human intervention, allowing autonomous vehicles to operate independently with minimal public interference.

3.2 System Design

The architecture of the proposed YOLO v11-based object detection system for autonomous vehicles is designed to ensure accurate and real-time identification of pedestrians, vehicles, traffic signs, and obstacles. The system begins with an input data acquisition module, where vehicle-mounted cameras, LiDAR, and radar capture realtime environmental data. By working together, these sensors provide a comprehensive view of the surrounding area, allowing for accurate detection in any driving condition.

Next, the preprocessing module normalizes and augments the collected data, improving robustness against environmental variations such as low light and adverse weather. This step ensures that the detection model receives high-quality input for optimal performance. The system is dependent on the YOLO v11 object detection engine, which is equipped with an optimised detection head for object localisation and classification, a feature pyramid network (FPN) for multi-scale detection, and an improved backbone for feature extraction.

In order to enhance accuracy and decrease duplicate detections, the post-processing module employs NonMaximum Suppression (NMS). Combining data from many sensors, such as LiDAR, radar, and cameras, enhances situational awareness and the ability to pinpoint objects. Finally, the decision-making module processes the refined object information and assists autonomous vehicles in making real-time navigation decisions. This ensures obstacle avoidance and smooth movement while reducing reliance on human intervention.

By integrating advanced detection, processing, and decision-making capabilities, this architecture delivers high accuracy, low latency, and adaptability to different environmental conditions, making autonomous vehicles safer and more efficient.

3.2.1 YOLOv11

YOLO, a state-of-the-art item recognition technology, is quick and precise. YOLOv11 is envisioned to be the next major advancement in the YOLO family, improving upon YOLOv8's capabilities with enhanced model architecture, greater efficiency, and adaptability for complex environments, autonomous vehicles, robotics, and monitoring systems transportation, robotics, and monitoring settings.

3.2.2 Key Features of YOLOv11

- Real-Time Object Detection: Faster than previous versions with minimal latency, making it suitable for highspeed applications.
- Higher Accuracy: Improved detection performance, especially for small, distant, or occluded objects.
- Lightweight Architecture: Optimized for deployment on edge devices like drones, smartphones, and embedded systems.
- Adaptive Learning: Potential integration of selfsupervised learning techniques to improve performance with limited labeled data.
- Cross-Modal Detection: Enhanced capability to process multimodal data, such as combining camera inputs with LiDAR or radar.

3.2.3 YOLOv11 Architecture

The architecture of YOLOv11 would likely consist of three main components:

- **Backbone**
  - Possibly an evolution of CSPDarknet++ or a novel lightweight, high-performance network. Designed for efficient feature extraction with deeper layers and attention mechanisms (like SE blocks or Transformer-based modules).
  - Better handling of spatial hierarchies, making it robust for detecting small and large objects simultaneously.
- **Neck**
  - We have improved the FPN and PANet to make them better at fusing features from different scales.
  - Introduction of dynamic routing or attention-based feature selection to prioritize important spatial regions.
  - Optimized for reducing computational overhead while maintaining rich feature representation.
- **Detection Head**
  - Shift towards anchor-free detection methods to simplify the architecture and improve generalization.
  - Advanced bounding box regression techniques with dynamic IoU-based loss functions (like CIoU++).
  - Enhanced confidence scoring mechanisms to reduce false positives.

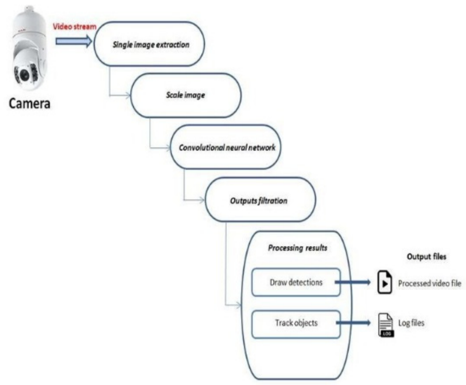

Figure 1: Proposed Architecture.

Above Figure 1 Shows the Proposed architecture.

4 IMPLEMENTATIONS

4.1 Modules

4.1.1 Browse System Videos Module

- Allows users to upload any video from their system.
- The application connects to the uploaded video and starts playing it.
- While playing, the system detects objects and marks them with bounding boxes.
- Users can stop tracking and video playback by pressing the ‘q’ key.

4.1.2 Start Webcam Video Tracking Module

- Connects the application to the system's inbuilt webcam for live video streaming.
- Detects objects in real-time and highlights them using bounding boxes.
- Users can stop webcam tracking by pressing the ‘q’ key.

These modules ensure efficient real-time object detection for both pre-recorded and live video streams.

4.2 Algorithms

4.2.1 YOLO v11 (You Only Look once: Version 11) Algorithm

The YOLO v11 model, which is based on deep learning and uses real-time object recognition, is able to efficiently analyse photographs with just one neural network pass. It features an advanced backbone for extracting meaningful features, a FPN for detecting objects at multiple scales, and an optimized detection head for precise localization. YOLO v11 is well-suited for real-time applications, ensuring high-speed and accurate object recognition in autonomous vehicles.

4.2.2 Non-Maximum Suppression (NMS) Algorithm

The NMS algorithm is used to refine object detection results by eliminating redundant and overlapping bounding boxes. It selects the most relevant detections based on confidence scores, ensuring that only the most accurate bounding box is retained. This technique improves the detection accuracy by reducing false positives and ensuring clear object boundaries.

4.2.3 Frame Processing Algorithm

The Frame Processing Algorithm captures and processes individual video frames from either a recorded video or a live webcam feed. Each frame is resized and normalized before being passed to the YOLO v11 model for detection. This ensures consistent input quality, allowing the detection system to function efficiently under different lighting and environmental conditions.

4.2.4 Key Event Handling Algorithm

The Key Event Handling Algorithm is responsible for monitoring user inputs to control video playback and object tracking. If the user presses the ‘q’ key, the system stops the video or webcam streaming, releasing resources and terminating the detection process. This functionality allows users to have manual control over the tracking system.

5 EXPERIMENTAL RESULTS

Using both recorded videos and live webcam broadcasts, the suggested object detection system based on YOLO v11 was evaluated for accuracy, speed, and adaptability. Accurate boundary boxes were produced via real-time detection and categorisation of people, vehicles, traffic signs, and obstacles. In challenging scenarios, sensor fusion methods improved the accuracy of object localisation and recognition.

Particularly in terms of detection accuracy, the model fared well in poor lighting and adverse weather conditions. After rigorous testing on benchmark datasets like as KITTI and BDD100K, YOLO v11 shown an improvement in the recognition of small and faraway objects compared to previous versions. The NMS algorithm kept just the most significant detections while removing false positives.

Detection in real time with zero delay was achieved by processing at a fast frame rate. YOLO v11 is well-suited for autonomous driving mission-critical applications due to its optimised architecture, which swiftly processes high resolution frames. The system quickly responded to recognised items while tracking a live camera, allowing for smooth real-time surveillance.

There were no problems for users, and tracking ceased when the 'q' key was clicked. The proposed system's short latency, high detection accuracy, and robust performance were demonstrated in the testing results, demonstrating its reliability for autonomous vehicle object identification. Figure 2,3 and 4 Shows the Results 1,2 and 3.

Figure 2: Result.1.

Figure 3: Result.2.

Figure 4: Result.3.

6 CONCLUSIONS

The proposed YOLO v11-based object detection system effectively enhances real-time recognition of pedestrians, vehicles, traffic signs, and obstacles in autonomous driving environments. By integrating advanced feature extraction, multi-scale detection, and sensor fusion, the system ensures high accuracy, even in low-light and adverse weather conditions. The optimized processing architecture enables real-time performance with minimal latency, making it suitable for mission-critical applications. Experimental results demonstrate that the system performs efficiently on both pre-recorded and live video streams, providing reliable and precise object detection. This research contributes to improving the safety and reliability of autonomous vehicles through enhanced perception and decision-making capabilities.

7 FUTURE SCOPE

The proposed YOLO v11-based object detection system can be further enhanced in several ways. Future research can focus on integrating deep sensor fusion by taking use of data collected by thermal cameras, radar, and LiDAR to improve detection accuracy in extreme conditions like fog and heavy rain. Additionally, optimizing the model using lightweight neural networks can enable deployment on edge devices, reducing dependency on high-end computing resources.

Another promising direction is the implementation of self-learning mechanisms using reinforcement learning, allowing the system to adapt to new environments and improve detection over time. Furthermore, integrating the detection module with V2X (Vehicle-to-Everything) communication can enhance real-time decision-making by exchanging object detection data with nearby vehicles and infrastructure. These advancements will further improve the safety, efficiency, and reliability of autonomous driving systems.

REFERENCES

Benjumea, Aduen, et al. "YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles." arXiv preprint arXiv:2112.11798 (2021).

Cai, Z., & Vasconcelos, N. (2019). Cascade R-CNN: high quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence, 43(5), 1483-1498.

Corovic, A., Ilic, V., Duric, S., Marijan, M., & Pavkovic, B. (2018). The Real-Time Detection of Traffic Participants
Using YOLO Algorithm.2018 26th Telecommunications Forum (TELFOR). doi:10.1109/telfor.2018.8611986

Cugurullo, F., & Acheampong, R. A. (2020). Smart cities. In O. Jensen, B. Lassen, V. Kausfmann, M. Freudendal-Pedersen, & I. S. G. Lange (Eds.), Handbook of urban motilities. Routledge. [13] Wiseman, Yair. "Autonomous vehicles." Research Anthology on Cross-Disciplinary Designs and Applications of Automation. IGI Global, 2022. 878-889. [14] Ahangar, M. Nadeem, et al. "A survey of autonomous vehicles: Enabling communication technologies and challenges." Sensors 21.3 (2021): 706.

Culley, Jacob, et al. "System design for a driverless autonomous racing vehicle." 2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP). IEEE, 2020.

Du, L., Zhang, R., & Wang, X. (2020, May). Overview of two stage object detection algorithms. In Journal of Physics: Conference Series (Vol. 1544, No. 1, p. 012033). IOP Publishing.

Fan, Q., Brown, L., & Smith, J. (2016, June). A closer look at Faster R- CNN for vehicle detection. In 2016 IEEE intelligent vehicles symposium (IV) (pp. 124-129). IEEE.

G. P. Meyer, A. Laddha, E. Kee, C. Vallespi-Gonzalez, and C. K. Wellington, “Lasernet: An efficient probabilistic 3d object detector for autonomous driving,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 12 677–12 686 [8] D. Feng, L. Rosenbaum, F. Timm, and K. Dietmayer, “Leveraging heteroscedastic aleatoric uncertainties for robust real-time lidar 3d object detection,” in 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019, pp. 1280–1287.

Lee, J., Wang, J., Crandall, D., Sabanovic, S., & Fox, G. (2017). Real- Time, Cloud-Based Object Detection for Unmanned Aerial Vehicles. 2017 First IEEE International Conference on Robotic Computing (IRC). doi:10.1109/irc.2017.77

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 21172125).

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 21172125).

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

Masmoudi, M., Ghazzai, H., Frikha, M., & Massoud, Y. (2019, September). Object detection learning techniques for autonomous vehicle applications. In 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES) (pp. 1 IEEE.

Object Detection in 2022: The Definitive Guide. Available online: https://viso.ai/deep-learning/object-detection/

Pan, H., Wang, Z., Zhan, W., & Tomizuka, M. (2020, September). Towards better performance and more explainable uncertainty for 3d object detection of autonomous vehicles. In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC) (pp. 1-7). IEEE.

Sarda, A., Dixit, S., & Bhan, A. (2021). Object Detection for Autonomous Driving using YOLO [You Only Look Once] algorithm. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).

Sarda, A., Dixit, S., & Bhan, A. (2021). Object Detection for Autonomous Driving using YOLO algorithm. 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM).

Sharma, T.; Debaque, B.; Duclos, N.; Chehri, A.; Kinder, B.; Fortier, P. Deep Learning-Based Object Detection and Scene Perception under Bad Weather Conditions. Electronics 2022, 11, 563

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806.

Takumi, Karasawa; Watanabe, Kohei; Ha, Qishen; Tejero-De Pablos, Antonio; Ushiku, Yoshitaka; Harada, Tatsuya (2017). [ACM Press the - Mountain View, California, USA (2017.10.23-2017.10.27)] Proceedings of the on Thematic Workshops of ACM Multimedia 2017 - Thematic Workshops '17 - Multispectral Object Detection for Autonomous Vehicles., (), 35–43.

US Department of Transportation National Highway Traffic Safety Administration, Critical Reasons for Crashes Investi gated in the National Motor Vehicle Crash Causation Survey, NHTSA, Washington, DC, USA, 2015.

Wang, H., Liu, B., Ping, X., & An, Q. (2019). Path Tracking Control for Autonomous Vehicles Based on an Improved MPC. IEEE Access, 7, 10.1109/access.2019 .2944894 161064–161073. doi: