

A Study of Autonomous Driving Object Detection Models Using Computer Vision and Deep Learning

Mumukshu Bhatt^{1*}, Subodh Prasad², Rajesh Singh³ and Pooja Tamta⁴

^{1,2,3}Department of Information Technology, College of Technology, Govind Ballabh Pant University of Agriculture & Technology, Pantnagar, UK, India

⁴Department of Extension Education and Communication Management, College of Community Science, Govind Ballabh Pant University of Agriculture & Technology, Pantnagar, UK, India

mumukshu09bhatt@gmail.com, subodhprasad3@gmail.com, rs01234@gmail.com, pooja_tamta20@yahoo.com

Abstract

Object detection (OD) is a computer technique that enables localisation in computer vision, allowing the location and classification of objects in images/videos. With the help of deep learning (DL), OD has seen significant advancements, delivering higher accuracy with greater efficiency outcomes than traditional methods. This study provides an overview of work on object detection models. This study focuses especially on work in the field of autonomous driving over the last few years. The various models utilised in autonomous driving applications include LM-CNN-SVM, YOLOv4, and Bayesian Neural Networks (BNNs). In the discussed models, VGG16 yielded better results than the other models in this study. This study compares different models and algorithms in tabular form.

Keywords: *Autonomous Driving, CNN, Deep Learning, Efficiency, Neural Network, Object Detection.*

1. Introduction

Object detection (OD) involves identifying and localising object features in images or videos. It is a basic task for applications such as autonomous driving, medical diagnostics, and remote sensing. Traditional methods are based on handcrafted features and linear classifiers, but deep learning (DL) has revolutionised the field by enabling end-to-end learning of hierarchical features.

This study gives an overview of OD techniques for autonomous vehicles. These are built on prior work that emphasises the need for object detection models, algorithms and technologies used in previous studies, model variants for autonomous vehicles, probabilistic approaches in autonomous driving, handling imbalance issues, and early trainable systems.

2. Understanding the need for object detection

Object detection is one of the most important problems in autonomous driving. When we perform probabilistic modelling in object detection, we need to estimate the probability that an object belongs to the classes of interest. Object detection algorithms fall into two categories: one-stage and two-stage. YOLO and SSD are one-stage detection methods that directly convert the detection problem into a unified regression problem, and they are faster than two-stage methods. The object detection structure is optimised based on the YOLOv4 [1] model, significantly reducing computational complexity while improving detection accuracy [1].

3. The need for object detection in moving vehicles

Object detection in moving vehicles is crucial because it operates in dynamic, unpredictable real-world environments, where understanding "what is where" in real time is challenging for safety, navigation, and task execution.

3.1. The human body experiences fatigue during long drives.

When a driver gets tired, their response time slows down. It may also miss pedestrians, vehicles, and obstacles, leading to major security issues. Using an object detection system in an autonomous vehicle can reduce this security issue by issuing an alert.

3.2. Disturbance inside the vehicle

Like phone calls, kids, eating, etc., these distractions cause drivers to take their eyes off the road for seconds, which is enough to cause an accident. Object detection enables the ADAS (advanced driver assistance system) feature, which actively alerts when the driver is distracted.

4. How can assistive driving help us to overcome this problem?

Fusion of CNN (convolutional neural network), SVM (support vector machine), and AI (artificial intelligence) models can reduce accident rates across diverse weather conditions, such as foggy/dusty/or rainy. CNNs can be designed using different combinations of convolutional, pooling, and fully connected layers, with point-wise nonlinear activation functions [1]. SVM allows for misclassified intra-margin training. The SVM formulation with an inequality constraint aims to minimise both the training and generalisation errors.

4.1. This algorithm will be an appropriate technology in our studies for vehicle driving on the road.

Hybrid local machine system (LM-CNN-SVM) based on Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) [1]

- M-HMP: Multipath Hierarchical Matching Pursuit model that extracts expressive features from images through multi-pathways, similar to deep learning [1]
- HIKSVM: It is a fast model that uses the Histogram Intersection Kernel (HIK) and SVM classifiers. [1]
- HOGSVM: It is a fundamental model that uses both the HOG and the SVM for object detection. [1]
- JointDeep: It is a unified deep model realised in combination that has fast feature cascades, all-feature extraction, occlusion, deformation handling, and classification. [1]
- YOLO v4: - Self-driving Assistive driving mainly includes dynamic objects like pedestrians and different kinds of vehicles, and static objects like traffic lights and stop signs. By which we can localise various objects, using an optimised YOLOv4 model. And then, based on the detected objects, it is essential to clarify the intentions of road users who are performing safe self-driving without causing traffic accidents.

4.2. Exclusion from disturbances originating from the passengers or surroundings inside/outside the vehicle

The AI model will alert the driver by sending a message/alert tone while the driver is distracted. Even when an object is near the car, the model will also provide a signal or message to the driver, which can definitely reduce daily road accidents.

5. Challenges faced by moving vehicles for object detection

Assistive driving relies on object detection to increase security. Object detection uses sensors to detect pedestrians, vehicles, and obstacles on streets. Moving vehicles pose a major safety concern for drivers driving at high speeds. High-speed motion leads to camera image blurring, thereby reducing feature extraction accuracy.

5.1. Weather and lighting

Different weather conditions, such as snow/fog/rain/or low light, reduce the model's accuracy. Even models like LMCNN-SVM achieve 92.80% [2] accuracy in their parametric performance across different weather conditions, whereas other models have significantly lower parametric performance.

5.2. Obstruction and heavy traffic

Objects obscured in traffic cause false navigation. In urban areas with high traffic volume, small objects may be missed during detection.

6. Methodology that has been used previously for such detections

This study presents the models and algorithms used in previous studies.

The algorithm that is used is applied in the hybrid local machine system (LM-CNN-SVM) in eight steps as follows.

- All images are divided into local regions.
- Each patch is wrapped to a fixed size, namely 64 x 64 x 3, and converted to a grey image.
- Each patch is imported into the pre-trained AlexNet and the proposed CNN, shown in Figure 1.
- All networks are trained by stochastic gradient descent.
- The salient features obtained from the final fully connected layer of the networks are saved.
- Principal Component Analysis (PCA) is deployed to reduce the dimension of the saved features.
- SVM classifiers are applied to the reduced features. The one-against-rest method is used for solving multi-class classification problems.
- Decision fusion is used to combine the outputs of multi-class SVM classifiers [2].

For more, refer to Figure 1.

Type	Filter-stride	Feature maps
Input	64×64	1
Convolution 1	$3 \times 3-1$	64
Convolution 2	$3 \times 3-1$	64
Average Pooling 2	$2 \times 2-2$	64
Convolution 3	$3 \times 3-1$	96
Max Pooling 3	$2 \times 2-2$	96
Convolution 4	$5 \times 5-1$	256
Max Pooling 4	$3 \times 3-2$	256
Convolution 5	$3 \times 3-1$	384
Convolution 6	$3 \times 3-1$	384
Convolution 7	$3 \times 3-1$	256
Max Pooling 7	$3 \times 3-2$	256
Fully connected 8	$6 \times 6-1$	4096
Fully connected 9	1×1	$1 \times 1 \times \text{Class number}$

Figure 1: This figure shows the algorithm [1].

The YOLOv4 model is modified in terms of network structure [3]. The network architecture of YOLOv4 improves learning of small objects, but the study found that detection performance on some large objects is poor in real-world testing.

- **INTENTION RECOGNITION:** Based on the monocular pedestrian detection, the method of using pose estimation as the main information to predict whether pedestrians have the intention of crossing the road is explored in this section.
- **POSE ESTIMATION:** In the pose estimation stage, the PAFs model is adopted to acquire the skeleton features [4].
- **INTENTION RECOGNITION:** After the pose estimation stage, the skeleton features are acquired and fed into a CNN model to recognise the intention of pedestrians.

7. Efficiency/efficacy that has been delivered so far.

Compared with YOLOv3, YOLOv4 achieves nearly an 8% increase in mAP, but the inference speed is not significantly improved [3]. They claim that this paper challenges the limitations of object recognition, pedestrian detection and autonomous vehicles. They observed that the LM-CNN-SVM system achieved the highest values of accuracy, 89.80 6 0.50 and 92.80 6 0.5 for 15 and 30 images per class [1]. They have used the Brier score, along with calibration errors, to evaluate the regression and classification probability distributions of in-distributions. Detection results predicted by our probabilistic detectors.

The improvement in efficiency that their research claims is that the proposed approach achieves a high mAP of 96.92% on the Car, Truck, Pedestrian, and Cyclist classes of the KITTI dataset, along with high-speed detection at 61 fps (on the Titan RTX 2080Ti GPU) and a minimum depth estimation error margin of 0.915 m [4].

Table 1: Comparison of different Models

S.No	Paper Name	Author(s) name	Algorithm used	Model Developed	Accuracy/Efficiency	Description of work carried out during research

1	Object recognition and detection with deep learning for autonomous driving applications [3]	Aysxegu Ucxar, Yakup Demir, and Cu'neyt Gu'zelisx	Scale Invariant Feature Transform (SIFT), Histogram of Gradient (HOG), Binary Robust Independent Elementary Feature (BRIEF), the Speeded up Robust Feature (SURF), and Fast Retina Keypoint (FREAK)	LMCNN-SVM system.	The LM-CNN-SVM system achieved the highest values of accuracy, 89.80 6 0.50 and 92.80 6 0.5 for 15 and 30 images per class, respectively.	The LM-CNN-SVM system significantly surpassed the state-of-the-art methods in terms of excellence in terms of higher recognition performance.
2	DeepRange: deep-learning-based object detection and ranging in autonomous driving. [4]	Yashrajsinh Parmar, Sudha Natarajan, Gayathri Sobha	Deep learning, SIFT, SURF, HOG, and feature extraction combined with SVM & KNN.	CNN was enhanced by adding a module for handling distance determination.	CNN, along with the VGG-16 base network, achieves a detection speed of 61 fps, with an mAP of 96.92%.	Cost-effective detection and ranging methodology suitable for high-speed highway driving.
3	A Review and Comparative Study on Probabilistic Object Detection in Autonomous Driving. [5]	Di Feng, Member, IEEE; Ali Harakeh, Member, IEEE; Steven L. Waslander, Senior Member, IEEE; and Klaus Dietmayer, Member, IEEE	Monte Carlo Dropout (MC-Dropout).	Bayesian Neural Networks (BNNs)	Brier Score, alongside calibration errors, is used for evaluating both regression and classification probability distributions.	Carried out a survey and comparative study on probabilistic object detection using deep learning approaches in the field of autonomous driving.

4	A Deep Learning-Based Hybrid Framework for Object Detection and Recognition in Autonomous Driving [2]	YANFEN LI, HANXIANG WANG, L. MINH DANG, TAN N. NGUYEN, DONGIL HAN (Member, IEEE), AHYUN LEE, INSUNG JANG, AND HYEONJOON MOON	YOLO, SSD	YOLO v4,	YOLOv4 achieves faster processing speed and higher detection accuracy than the previous model.	The RISE algorithm was used in the classification of objects by making the corresponding saliency maps for each image.
---	---	--	-----------	----------	--	--

8. Conclusion & Future Scope

In autonomous vehicle applications, driving models used for safety purposes may behave differently depending on the vehicle's headlight shade. This condition may occur while the driver is driving at night. Under yellow/blue/white lighting, this may cause incorrect object recognition and detection, which can lead to dangerous events.

The weather is still a challenging factor in the testing process due to poor quality and poor illumination conditions. Foggy/dusty/rainy weather might be affecting these models. In assistive driving, if the driver is driving in this diverse weather, the driver will not be able to recognise objects within a few meters of the vehicle.

References

- [1] Uçar, A., Demir, Y., & Güzeliş, C. (2017). Object recognition and detection with deep learning for autonomous driving applications. *Simulation*, 93(9), 759–769.
- [2] Li, Y., Wang, H., Dang, L. M., Nguyen, T. N., Han, D., Lee, A., ... & Moon, H. (2020). A deep learning-based hybrid framework for object detection and recognition in autonomous driving. *IEEE Access*, 8, 194228–194239.
- [3] Chen, Y., Zhao, D., Lv, L., & Zhang, Q. (2018). Multi-task learning for dangerous object detection in autonomous driving. *Information Sciences*, 432, 559–571.
- [4] Cui, J., Liu, S., Tian, Z., Zhong, Z., & Jia, J. (2022). Result: Residual learning for long-tailed recognition. *IEEE transactions on pattern analysis and machine intelligence*, 45(3), 3695–3706.
- [5] Feng, D., Rosenbaum, L., & Dietmayer, K. (2018, November). Towards safe autonomous driving: Capture uncertainty in the deep neural network for lidar 3d vehicle detection. In *2018, the 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 3266–3273). IEEE.

- [6] Feng, D., Harakeh, A., Waslander, S. L., & Dietmayer, K. (2021). A review and comparative study on probabilistic object detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 9961–9980.
- [7] Feng, D., Rosenbaum, L., & Dietmayer, K. (2018, November). Towards safe autonomous driving: Capture uncertainty in the deep neural network for lidar 3d vehicle detection. In *2018, the 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 3266–3273). IEEE.
- [8] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1), 142–158.
- [9] Hall, David, Feras Dayoub, John Skinner, Haoyang Zhang, Dimity Miller, Peter Corke, Gustavo Carneiro, Anelia Angelova, and Niko Sünderhauf. "Probabilistic object detection: Definition and evaluation." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1031–1040. 2020.
- [10] Han, Y., Liu, X., Sheng, Z., Ren, Y., Han, X., You, J., ... & Luo, Z. (2020). Wasserstein loss-based deep object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 998-999).
- [11] Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8), 832–844.
- [12] Janai, J., Güney, F., Behl, A., & Geiger, A. (2020). Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends in Computer Graphics and Vision*, 12(1-3), 1–308.
- [13] Kim, H., Hong, S., Son, H., Roska, T., & Werblin, F. (2003, May). High-speed road boundary detection on images for an autonomous vehicle with a multi-layer CNN. In *2003, IEEE International Symposium on Circuits and Systems (ISCAS)* (Vol. 5, pp. V-V). IEEE.
- [14] Lin, W. C., Tsai, C. F., Chen, Z. Y., & Ke, S. W. (2016). Keypoint selection for efficient bag-of-words feature generation and effective image classification. *Information Sciences*, 329, 33-51.
- [15] McAllister, R. T., Gal, Y., Kendall, A., Van Der Wilk, M., Shah, A., Cipolla, R., & Weller, A. (2017, August). Concrete problems for autonomous vehicle safety: Advantages of Bayesian deep learning. International Joint Conferences on Artificial Intelligence, Inc.
- [16] Meyer, G. P., Laddha, A., Kee, E., Vallespi-Gonzalez, C., & Wellington, C. K. (2019). Lasernet: An efficient probabilistic 3d object detector for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern*

recognition (pp. 12677–12686).

- [17] Parmar, Y., Natarajan, S., & Sobha, G. (2019). DeepRange: deep-learning-based object detection and ranging in autonomous driving. *IET Intelligent Transport Systems*, 13(8), 1256–1264.
- [18] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
- [19] Uçar, A., Demir, Y., & Güzeliş, C. (2016, August). Moving towards object recognition with deep learning for autonomous driving applications. In *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)* (pp. 1–5). IEEE.
- [20] Uçar, A., Demir, Y., & Güzeliş, C. (2017). Object recognition and detection with deep learning for autonomous driving applications. *Simulation*, 93(9), 759–769.
- [21] Wang, J. G., & Zhou, L. B. (2018). Traffic light recognition with high dynamic range imaging and deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 20(4), 1341–1352.
- [22] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10), 3349–3364.
- [23] Pathak, G., Pandey, D., Sonkar, N., & Kohli, P. (2025). Seeing Beyond Words: Leveraging Computer Vision For Interviewee Analysis In Ai-Driven Video Interviews. Available at SSRN 5250720.
- [24] Zeiler, M. D., & Fergus, R. (2014, September). Visualising and understanding convolutional networks. In *European Conference on Computer Vision* (pp. 818–833). Cham: Springer International Publishing.
- [25] Zhang, S., Wen, L., Bian, X., Lei, Z., & Li, S. Z. (2018). Single-shot refinement neural network for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4203–4212).