

An extensive analysis of 2D object recognition and distance estimation for autonomous driving using LIDAR

Md Ashiqur Rahman

1393169

md.rahman3@stud.fra-uas.de

Suva Sarkar

1431131

suva.sarkar@stud.fra-uas.de

Abstract—Effective vehicle detection is a top priority in the field of real-time applications like surveillance and autonomous driving. This environment has changed due to the You Only Look Once (YOLO) framework, which is well-known for its quick object detection. This study explores the integration of Yolo v11 and real-time vehicle recognition, refining its architecture for quick and accurate identification. The model's exceptional accuracy of 97.9% on mAP50 and 91.3% on mAP50-95, as demonstrated by empirical validation, highlights its strength. This accomplishment confirms the model's real-time processing capabilities in addition to attesting to its accuracy. The consequences have the potential to revolutionize urban planning and transportation safety. This study highlights YOLO v11's strength in intelligent object recognition, which has the potential to transform and improve a number of fields.

Keywords—*YOLO Framework, mAp50-95, Vehicle Detection, Object Detection.*

I. INTRODUCTION

It is impossible to overestimate the significance of vehicle detection in the modern world. Travel times have considerably increased due to the growing number of vehicles on the road, which presents difficulties for effective transportation [1]. One of the main causes of this issue is urbanization, which increases traffic density in cities [2]. The need for cutting-edge technologies that can address these issues becomes crucial in such a setting. Strong vehicle detection systems that can function well in a variety of lighting conditions are essential since accidents are a constant hazard, both during the day and at night [3][4][5]. As cities continue to struggle with the effects of urban expansion and increased congestion, addressing this urgent demand is essential for maintaining road safety and improving traffic management.

The need for advanced vehicle identification techniques in traffic situations has increased due to the quick internet connectivity and boom in technical improvements [6], which is in line with the Sustainable Development Goals of the UN, which are meant to promote innovation and resilient infrastructure [7]. Vehicle movement monitoring via entity recognition, computer vision-based recognition and tracking, and image segmentation for complex classifications are just a few of the many potential uses for this automation wave [8][9][10]. Image segmentation technology is essential for precisely interpreting pixel-level information about lanes, driving zones, and impediments in the field of autonomous driving. These

developments address issues of urbanization and traffic while also bringing transport systems into line with sustainability objectives, resulting in safer and more precise vehicle interactions.

II. LITERATURE SURVEY

As computer vision techniques advance, the field of vehicle identification and categorization has seen a revolutionary shift. The emergence of deep learning, particularly the You Only Look Once (YOLO) framework, transformed the area after the introduction of more conventional techniques like Haar cascades and HOG. This literature review follows this development, emphasizing the move from traditional methods to cutting-edge deep learning models like YOLO v11. The context for understanding the present research on effective real-time vehicle detection and classification using YOLO v11 is established by this survey.

When Zuraimi et al. [1] examine different YOLO models for tracking and vehicle recognition, they find that YOLOv4 performs best, with an astounding 82.08% AP50 on their unique dataset. The popularity of this model is ascribed to its sophisticated architecture, which exhibits advancements over earlier versions. In addition to its exceptional accuracy, YOLOv4 strikes a balance between speed and precision by achieving an immediate processing speed of roughly 14 FPS on a GTX 1660ti graphics card. The usefulness of YOLO models in real-time applications, especially in fields like vehicle tracking, is highlighted by this study.

Gorodokin et al.'s research [2] creatively uses machine vision to improve adaptive traffic light regulation, tackling congestion and traffic flow efficiency. The SORT open-source tracker and YOLO v4 neural network are combined to provide an impressive 92% accuracy in vehicle recognition and categorization. This study shows that integrating cutting-edge deep learning and tracking methods for efficient traffic analysis is feasible. The study emphasizes how data-driven approaches have the ability to transform adaptive traffic management systems and urban mobility.

During their search, Miao et al. [3] make a substantial contribution to the detection of low-light vehicles. Their

Creative method combines customized YOLO v3 neural network training with the MSR algorithm for image enhancement. This adaptation runs at a fast 30.03 fps frame rate and achieves an amazing 93.66% accuracy. Their approach successfully tackles the problem of nighttime detection, showcasing the adaptability of the YOLO v3 architecture and its potential for practical uses.

As an advancement of YOLOv3-tiny for better object identification, Chen et al.'s [4] study presents YOLO v3-live, a novel neural network structure. With a detection speed of 28 frames per second, YOLO v3-live achieves 87.79% mAP precision prior to quantization and retains 69.79% mAP after quantization. This development has intriguing applications across a range of fields and represents progress in finding a compromise between accuracy and real-time processing.

III. METHODOLOGY

A. Dataset Description

The collection of annotated car photos used in this study came from a variety of sources, such as web scraping, the Stanford Car Dataset, and Kaggle. Each class is represented by 500 photographs out of a total of 3000 images, making the dataset balanced and thorough. The dataset is prepared according to YOLO's unique format, which uses TXT files to contain the annotation data. The following discrete classifications are included in the dataset: Car, Three-Wheel, Bus, Truck, Motorcycle, and Van. The suggested vehicle detection and classification model is trained and assessed using this carefully chosen set of annotated photos.

B. Dataset Preprocessing

To guarantee peak performance, the dataset was carefully preprocessed before model training. A number of crucial procedures are included in the preprocessing pipeline. To ensure uniform input, photos are first scaled to a 800x800 resolution, standardizing their proportions. Data enrichment techniques like as flipping, rotational variations, and brightness modifications are used to create enriched versions of the original photos in order to increase the resilience of the model. The YOLO format annotations are modified appropriately to align with the enhanced photos.

In order to facilitate effective data administration, the dataset is organized by classifying photos and annotating labels into distinct folders. A deliberate data split is used, setting aside 30% of the dataset for testing and 70% for training in order to guarantee comprehensive model evaluation. By strengthening the model's dependability, this technique improves accuracy in practical applications like surveillance and driverless cars.

C. Dataset Analysis

Vehicle detection is essential for improving traffic control, enabling efficient surveillance systems, and enabling the operation of autonomous cars. This feature is essential for tasks like real-time monitoring, reducing traffic congestion, and avoiding collisions. Therefore, reliable vehicle detection serves as the foundation for many important features and advancements.

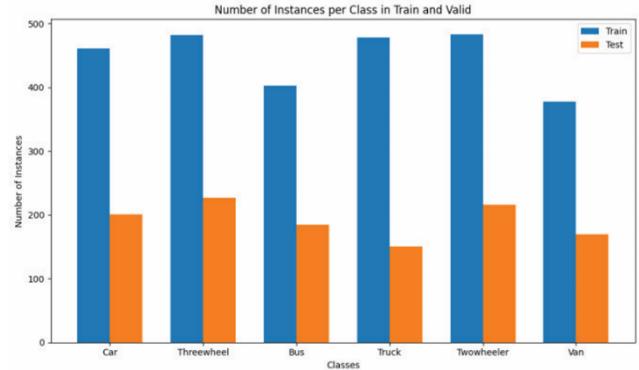


Fig. 1. Distribution of vehicle classes in the training and testing set

It is possible to identify vehicle patterns and features by using the extensive dataset. The dataset is divided into two training and testing folders and includes different vehicle types, including cars, three-wheelers, buses, trucks, two-wheelers, and vans (see the figures. 1 and 2).



Fig. 2. Classification of Vehicles Using Bounding Boxes

D. YOLO v11 Framework

In order to efficiently detect objects, the YOLO (You Only Look Once-) Framework divides a picture into a grid and uses the network output to directly forecast bounding boxes and item classes (see Fig. 3).

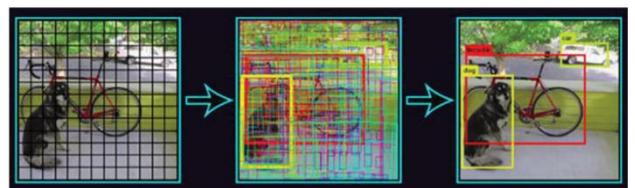


Fig. 3. Bounding box prediction with the YOLO Framework

YOLO is well suited for real-time activities like video analysis and autonomous driving since it can quickly identify objects in a single network run. By approaching object detection as a regression problem, it effectively predicts attributes and classes, which accounts for its speed and accuracy.

The architecture can be divided into several levels, such as detection, pooling, and convolutional layers. Through the acquisition of significant patterns from the input image, the convolutional layers carry out feature extraction. While maintaining key features, pooling layers minimize spatial dimensions. Bounding box predictions and class probabilities are produced by the detection layers. Two completely

connected layers come after 24 convolution layers in the network architecture of the YOLO framework. According to Fig. 4, changing 1x1 convolution layers shrinks the features space from earlier layers.

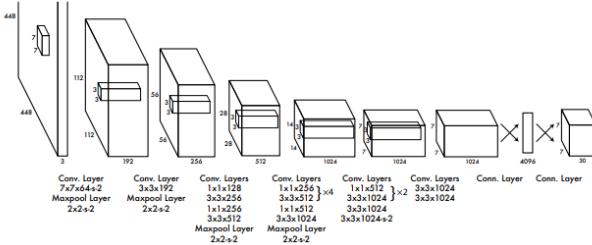


Fig. 4. The architecture of YOLO Framework

The central coordinates (x, y) and the bounding box's dimensions (width w , height h) are used to calculate the modified bounding box coordinates ($x_{min}, y_{min}, x_{max}, y_{max}$):

$$\begin{aligned} x_{min} &= \hat{x} - (\hat{w}/2) \\ y_{min} &= \hat{y} - (\hat{h}/2) \\ x_{max} &= \hat{x} + (\hat{w}/2) \\ y_{max} &= \hat{y} + (\hat{h}/2) \end{aligned}$$

IV. EXPERIMENTAL RESULT

Here, we provide the experimental results and talk about them using the YOLO v11 framework. We can learn more about the efficacy and performance of our technique in this setting by analyzing these outcomes.

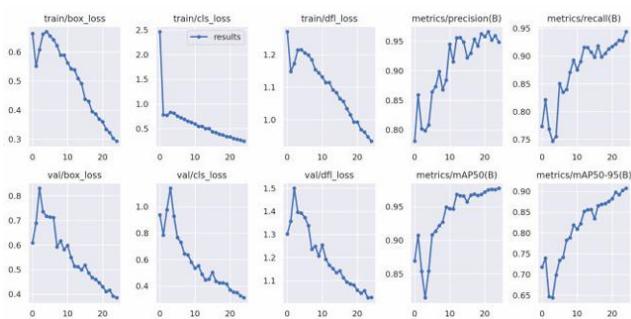


Fig. 5. Performance in training

The training procedure used the SGD optimizer with a learning rate of 0.01 and comprised 25 epochs with a batch size of 16. Separate parameter groups (57 for 0.0 weight decay, 64 for 0.0005 weight decay, and 63 for biases) were set up for weight decay and biases. The flexibility of the model was improved by augmentation approaches such blur, median blur, greyscale conversion, and CLAHE with a 0.01 probability. The YOLO model's object detection ability was enhanced by this through training

process , guaranteeing accuracy and resilience in real-world situations. (See Figure 5).

Class	mAP50	mAP50-95
Car	0.972	0.936
Three Wheeler	0.989	0.926
Bus	0.988	0.954
Truck	0.982	0.917
Two Wheeler	0.962	0.785
Van	0.983	0.957

Table 1. YOLO v11 model performance across all classes

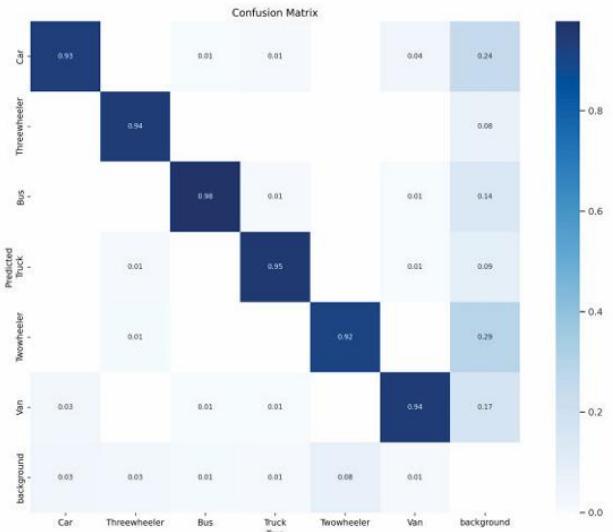


Fig. 6. Confusion Matrix

Important metrics and curves are used to assess object detection models. The model's sensitivity at various thresholds is illustrated by the Recall-Confidence Curve, which clarifies the relationship between recall and confidence thresholds (see Fig. 7). By balancing precision and recall, the Precision-Recall Curve illustrates how they interact at different confidence levels (see Fig. 8). Referring to Fig. 9, the Precision-Confidence Curve clearly illustrates how consistent precision is over a range of confidence levels. In the meanwhile, the F1-Confidence Curve captures the trade-off between recall and precision by connecting the F1-score to confidence thresholds (see Fig. 10). We provide a comprehensive evaluation of the model's performance over a range of thresholds and indicators by incorporating these curves into our presentation.

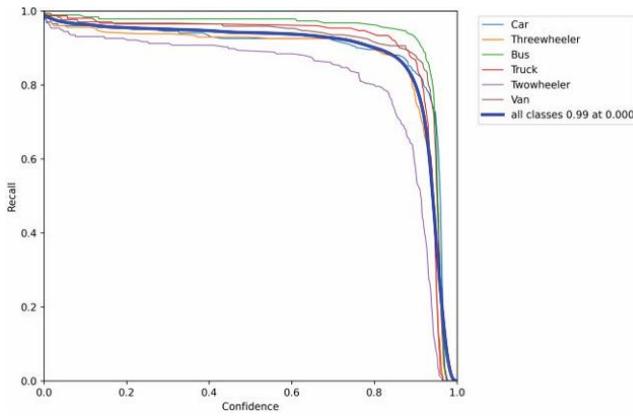


Fig. 7. Recall-Confidence Curve

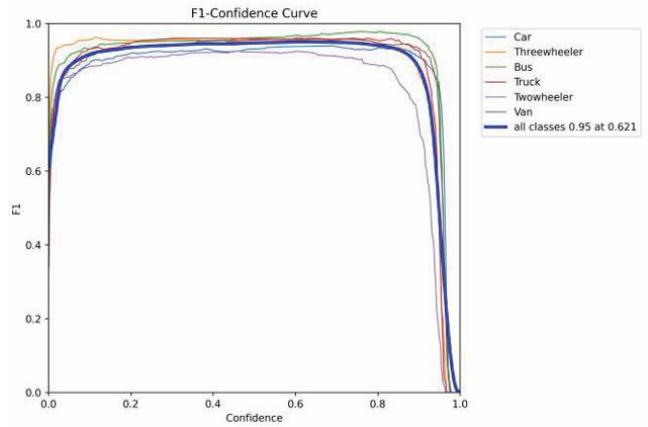


Fig. 10. F1-Confidence Curve

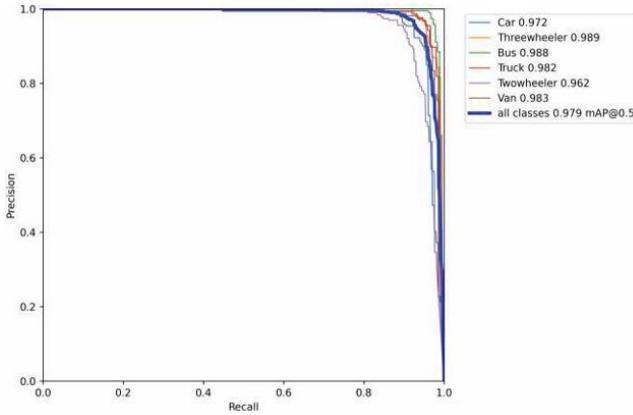


Fig. 8. Precision-Recall Curve

The model had very impressive performance, attaining a recall rate of 93.5% and a precision rate of 96.6%. Additionally, the model's ability to handle a range of item sizes and placements was demonstrated by the strong result of 97.9% for the average precision across several categories (mAP) at IoU 0.50. Further confirming the model's resilience in object detection and localization tasks, the mAP computed across IoU thresholds from 0.50 to 0.95 produced a respectable 91.3%.

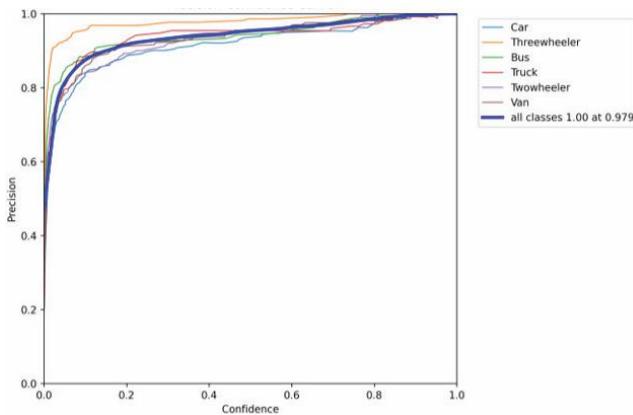


Fig. 9. Precision-Confidence Curve

V. CONCLUSION

In summary, significant discoveries and developments have resulted from the investigation into effective vehicle detection and categorization utilizing the YOLO v5 framework. Particularly noteworthy is the YOLO v11 invention, which has shown remarkable accuracy of 97.9% on mAP50 and 91.3% on mAP50-95. This accomplishment highlights YOLO v11's revolutionary potential in real-time applications and establishes it as a key instrument in the transformation of vehicle detection systems.

In the future, these methods will have an impact on more than only vehicle identification; they will influence the development of several technologies. YOLO v11's accuracy and real-time capabilities serve as an example of a larger trend in technical improvement. These methods provide the basis of advancements in self-driving car technology, which makes transportation networks safer. Additionally, their influence is felt in areas like smart cities, which allow for effective traffic control and real-time surveillance for better urban living. These developments go beyond transportation and have an impact on industries like robotics and healthcare, highlighting the revolutionary potential of these methods in fostering a more connected and effective future.

REFERENCES

- [1] M. A. Zuraimi and F. H. K. Bin, "Vehicle detection and tracking using YOLO and DeepSORT," in 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE, 2021.
- [2] V. Gorodokhin, S. Zhankaziev, E. Shepeleva, K. Magdin, and S. Evtyukov, "Optimization of adaptive traffic light control modes based on machine vision," Transp. Res. Procedia, vol. 57, pp. 241–249, 2021.
- [3] Y. Miao, F. Liu, T. Hou, L. Liu, and Y. Liu, "A nighttime vehicle detection method based on YOLO v3," in 2020 Chinese Automation Congress (CAC), 2020.
- [4] S. Chen and W. Lin, 2019 IEEE 3rd advanced information management, communications, electronic and automation control conference (IMCEC). IEEE, 2019.

- [5] X. Dong, S. Yan, and C. Duan, “A lightweight vehicles detection network model based on YOLOv5,” *Eng. Appl. Artif. Intell.*, vol.113, no. 104914, p. 104914, 2022.
- [6] H. Song, H. Liang, H. Li, Z. Dai, and X. Yun, “Vision-based vehicle detection and counting system using deep learning in highway scenes,” *Eur. Transp. Res. Rev.*, vol. 11, no. 1, 2019.
- [7] O. O. Khalifa, “Vehicle detection for vision-based intelligent transportation systems using convolutional neural network algorithm,” *Journal of Advanced Transportation*, vol. 2022, pp. 1–11, 2022.
- [8] C. Chen, “Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [9] Z. Chen et al., “Fast vehicle detection algorithm in traffic scene based on improved SSD,” *Measurement (Lond.)*, vol. 201, no. 111655, p.111655, 2022.
- [10] S. Chadwick, W. Maddern, and P. Newman, “Distant vehicle detection using radar and vision,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8311–8317.
- [11] H. Wang, Y. Yu, Y. Cai, X. Chen, L. Chen, and Q. Liu, “A comparative study of state-of-the-art deep learning algorithms for vehicle detection,” *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 2, pp. 82–95, Summer 2019.
- [12] M. Hassaballah, M. A. Kenk, K. Muhammad, and S. Minaee, “Vehicle detection and tracking in adverse weather using a deep learning framework,” *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4230–4242, 2021.
- [13] P. Premaratne, I. Jawad Kadhim, R. Blackridge, and M. Lee, “Comprehensive review on vehicle detection, classification and counting on highways,” *Neurocomputing*, no. 126627, p. 126627, 2023.
- [14] Laxmikant, Kumar, R. Bhuvaneswari, and B. Natarajan. "An Efficient Approach to Detect Diabetes using XGBoost Classifier." *2023 Winter Summit on Smart Computing and Networks (WiSSCoN)*. IEEE, 2023.
- [15] A. Padmavathi, A. Bansal, A. Gupta, A. Nambiar, and K. Laxmikant, “Heart Disease Prediction using Supervised Learning Algorithms - A Comparative Study,” *International Journal of Soft Computing And Artificial Intelligence (IJSCAI)*, no. 10, pp. 32–37, 2022.