International Journal of Recent Advances in Multidisciplinary Topics
Volume 6, Issue 12, December 2025
www.ijramt.com | E-ISSN: 2582-7839 | RESAIM Publishing (www.resaim.com)

Real-Time YOLOv5-Based Object Detection for Autonomous Vehicles on Nigerian Roads

Martins E. Irhebhude[1], Modibbo Tukur Ahmed[2], Ibrahim Aliyu Ibrahim[3], Isah Shuaibu[4*]

[1]*Professor, Department of Computer Science, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy, Kaduna State, Nigeria*

[2,3]*Kaduna Polytechnic, Kaduna, Nigeria*

[4]*Lecturer, Department of Geography, Nigerian Defence Academy, Kaduna, Nigeria*

Abstract*: Autonomous vehicles (AVs) offer substantial improvements in safety and efficiency; however, global perception benchmarks are primarily based on structured traffic environments in developed countries and do not capture the informal and complex road conditions common in Nigeria and other African nations. To address this gap, the present study developed a custom dataset comprising 8,093 real-world images from Nigerian roads, annotated with 22 contextually relevant classes, including tricycles (*Keke NAPEP*), *Okada* motorcycles, street vendors, potholes, animals, and degraded traffic signs. Three lightweight You Only Look Once (YOLO) models (v5, v8, and v11) were trained and evaluated using extensive data augmentation. YOLOv5 demonstrated the highest performance, achieving an mAP@0.5 of 96.8%, a peak F1-score of 0.94, and real-time inference at 82 FPS on an NVIDIA RTX 2060, outperforming YOLOv8 and YOLOv11 while requiring the fewest parameters. Notably, strong results were observed for safety-critical classes (pedestrians: 98.3%; tricycles: 92.6%), while degraded signage and small or occluded objects remained the primary limitations. These findings indicate that high-accuracy, vision-only perception for autonomous vehicles is feasible in resource-constrained, unstructured African traffic environments using low-cost cameras and consumer-grade hardware, provided that a locally collected dataset and robust augmentation strategies are employed. This work establishes a scalable, cost-effective benchmark and provides publicly available resources to support AV development in Nigeria and comparable developing regions.

***Keywords*: autonomous vehicles, custom dataset, Nigerian roads, object detection, real-time detection, vision-only perception, YOLOv5.**

1. Introduction

Autonomous vehicles (AVs) have the potential to significantly transform transportation by reducing human error, enhancing road safety, and lowering traffic fatalities (Caesar *et al.*, 2020; Di Lillo *et al.*, 2023; Soni, 2024). Object detection, the automated identification and localisation of objects in images or video, is fundamental to AV perception. It enables the classification of entities such as pedestrians, vehicles, traffic signs, and barriers within the operational environment (Du, 2023). However, existing global benchmarks, including KITTI, nuScenes, and Waymo Open, do not capture the complexity of informal transport modes, unpredictable pedestrian crossings, and degraded signage that are prevalent on Nigerian roads. These standard datasets are characterised by predictable traffic patterns and infrastructure that supports advanced sensing technologies (Caesar *et al.*, 2020; Tan *et al.*, 2025).

Unlike the structured road networks represented in many global datasets, road environments in developing countries such as Nigeria are characterised by informal transport modes, including tricycles and motorcycles, roadside vendors encroaching on traffic lanes, unpredictable pedestrian crossings, degraded signage, potholes, speed bumps, and a heterogeneous mix of motorised and non-motorised traffic. These distinguishing features are not reflected in the global datasets commonly used for AV training (Tucho, 2022; Achenef *et al.*, 2025). The resulting variability increases detection challenges and often leads to models trained on global datasets failing to recognise obstacles unique to African contexts (Mutabarura, Muchuka, & Segera, 2025).

Deep learning methods, particularly convolutional neural networks (CNNs), have substantially advanced object detection by utilising hierarchical feature learning from annotated images (Deng & Li, 2024; Mahajan & Mane, 2023). CNNs are a class of artificial neural networks designed to process grid-like data, such as images, and form the backbone of single-stage architectures. Approaches such as You Only Look Once (YOLO) facilitate real-time inference and balance precision and latency, making them suitable for AV deployment (Montgomerie-Corcoran, Toupas, Yu, & Bouganis, 2023; Wang *et al.*, 2024). In these single-stage methods, the model predicts object locations and classes directly from the input image in a single network pass. However, these models exhibit diminished effectiveness in underrepresented contexts that lack localised data reflecting the unique characteristics of African traffic (Mutabarura, Muchuka, & Segera, 2025).

Road transport in Nigeria accounts for over 90% of passenger and freight mobility, and rapid vehicle growth combined with infrastructure deficits contributes to chaotic traffic conditions (Uhegbu, 2020). Objects such as tricycles ("*Keke NAPEP*"), roadside vendors, and jaywalking pedestrians are prevalent but are rarely represented in global AV datasets. Consequently, AV perception systems trained on standard benchmarks often fail to detect or classify these critical objects, posing significant safety risks.

*Corresponding author: isahshuaibu@nda.edu.ng

To address these challenges specific to the Nigerian context, this study develops an object detection system tailored to the unique conditions of Nigerian roads, with a focus on constructing an annotated dataset that accurately represents these environments. The objectives are to compile a primary dataset of images from Nigerian roads annotated with common objects such as tricycles, street vendors, pedestrians, and other typical obstacles; to develop a system capable of detecting and classifying traffic signs from camera images in real time; to detect and classify both static and dynamic objects in Nigerian settings; and to evaluate the accuracy and efficiency of the developed system in real-time object recognition and classification. This research advances AV equity in low- and middle-income countries and aligns with United Nations Sustainable Development Goals 9 (industry, innovation, and infrastructure) and 11 (sustainable cities and communities). Given the limited research on vision-only perception systems adapted to African contexts (Wang *et al*., 2024; Achenef *et al*., 2025), these findings contribute to the development of resilient and cost-effective AV solutions for emerging economies.

2. Literature Review

Ahmed (2025) developed a real-time perception system that integrates YOLOv5 for object detection with a monocular depth-estimation network to improve autonomous driving performance under varied environmental conditions. The study addressed challenges such as lighting variation, occlusion, and limited training data by applying data augmentation and transfer learning, achieving higher detection accuracy than earlier approaches based on Faster R-CNN and SSD. Although distance estimation was a major focus, the findings reinforce the effectiveness of YOLO-based detectors in real-time traffic environments. The relevance of this work to the present study lies in its demonstration of YOLOv5's robustness in complex road scenarios, further supporting the use of YOLO architectures for vision-only autonomous navigation on developing-world road networks.

Tan *et al*. (2025) proposed SceneDiffuser++, a diffusion-based generative world model enabling city-scale, point-to-point traffic simulation. Trained on the Waymo Open Motion Dataset, it jointly performs scene generation, long-horizon agent prediction, occlusion reasoning, dynamic agent spawning/removal, and traffic-light simulation. The model significantly outperforms log-replay baselines in terms of realism across extended rollouts. Although tested in structured North American cities, the work underscores the need for generative capabilities in highly dynamic, unstructured environments like Nigerian roads, supporting the vision-only perception approach and dataset-centric strategy of the present study.

Mutabarura *et al*. (2025) developed a custom African road-obstacles dataset containing rare classes, such as livestock and wildlife, frequently encountered on sub-Saharan roads, thereby addressing a significant gap in global benchmarks. Three YOLO versions (v3, v5, v8) were evaluated, with YOLOv5 achieving the highest mAP@0.5 of 94.68 % when trained with offline data augmentation. The authors emphasised that augmentation dramatically improved performance on minority classes and concluded that YOLOv5 remains the most effective lightweight architecture for real-time obstacle detection in African driving conditions. These findings directly support the superiority of YOLOv5 observed in the present Nigerian study and validate the critical role of region-specific datasets and augmentation strategies in overcoming domain gaps.

Achenef *et al*. (2025) conducted a review of intelligent control systems for autonomous vehicles, focusing on perception, decision-making, and motion control in global contexts. They highlighted that most breakthroughs in object detection and sensor fusion have been validated on structured datasets from North America, Europe, and China, whereas very limited research addresses the unique challenges of unstructured African roads, including informal transport modes, degraded signage, and mixed traffic with livestock. The authors noted the scarcity of region-specific datasets and the poor generalisation of globally trained models in low- and middle-income countries, aligning directly with the motivation of the present Nigerian study. They concluded that future progress in African AV deployment requires custom datasets and lightweight, vision-only architectures, findings that strongly support the high performance achieved in this study using only low-cost cameras and YOLOv5 on a Nigerian-specific dataset.

Di Lillo *et al*. (2023) compared Waymo's fully driverless service with human drivers using third-party liability claims data. Over 3.8 million rider-only miles, Waymo recorded zero bodily-injury claims (vs. human baseline 1.11 cpmm) and reduced property-damage claims by 76 % (0.78 vs. 3.26 cpmm). The authors concluded that a mature vision-based Level-4 system is significantly safer toward other road users than human drivers, especially vulnerable ones. These findings strongly support the safety potential of the high detection rates achieved in the present study for pedestrians, tricycles, and motorcycles in Nigerian traffic conditions.

Du (2023) reviewed deep learning-based object detection techniques for autonomous vehicles, categorising them into two-stage (R-FCN, Mask R-CNN) and one-stage (SSD, RetinaNet, YOLO) algorithms. Among one-stage detectors, YOLO was highlighted for its superior real-time performance through single-pass prediction of bounding boxes and class probabilities, making it particularly suitable for resource-constrained autonomous driving applications. The author concluded that YOLO variants consistently outperform other one-stage methods in speed–accuracy trade-offs on standard benchmarks, supporting the selection of YOLOv5 as the core detector in the present Nigerian road study, where real-time inference on consumer-grade hardware is essential.

3. Methodology

A vision-based object detection system was developed for AVs, with a focus on Nigerian roads. A custom dataset of 8,093 images was collected, covering both urban and rural areas. These images included tricycles (Keke NAPEP), motorcycles, roadside vendors, pedestrians, and conventional vehicles. They were captured under various lighting and weather conditions.

The dataset was divided into training (7,000 images), validation (1,000 images), and test (93 images) sets to support robust generalisation.

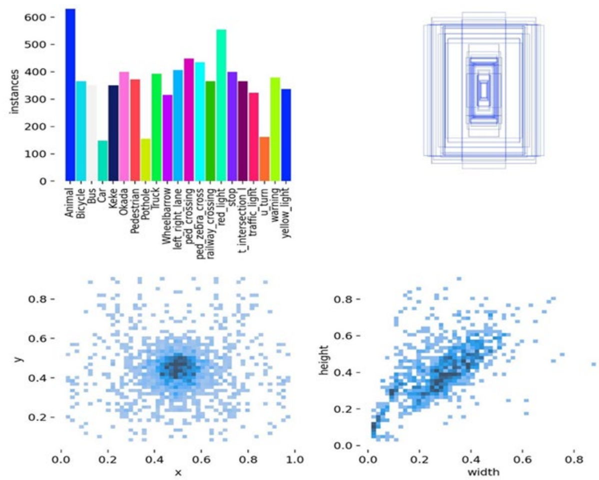

Fig. 1. Dataset composition and annotation analysis
Distribution of object instances across classes (top left), spatial distribution of bounding box annotations (top right), normalised centre coordinates of annotated objects (bottom left), and normalised bounding box dimensions showing object scale variation (bottom right). *Source: Researcher's Analysis (2025)*

Three YOLO variants (YOLOv5, YOLOv8, and YOLOv11) were tested for real-time object detection. YOLOv5 was chosen for its superior performance, particularly in Average Precision at 50% IoU (mAP50) and balanced precision-recall. The single-stage detection in YOLOv5 processes entire images in a single pass, enabling real-time AV inference.

Model training was performed on an NVIDIA GeForce RTX 2060 GPU (6GB VRAM) using Python 3.11 and PyTorch 2.7.0 with CUDA acceleration. YOLOv5 version 8.3.130 was utilised as the training framework. The hyperparameters included a batch size of 8, input image resolution of 640×640 pixels, and the AdamW optimiser (a gradient-based optimisation algorithm) with a learning rate of 0.0004. Box loss was used for bounding-box regression, classification loss for class prediction, and Distribution Focal Loss (DFL), which helps the model focus on difficult examples, were applied to improve detection of challenging objects. Data augmentation was implemented using RandAugment with random transformations such as rotation, horizontal flipping, and colour adjustment to increase model robustness. Training was conducted for 50 epochs, with early stopping applied at epoch 25 to prevent overfitting if validation performance did not improve for 10 consecutive epochs. Learning rate scheduling was used to optimise convergence.

Model performance was evaluated using standard object detection metrics. Precision, defined as the proportion of true positives to total positive predictions, quantifies the accuracy of detected objects:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Recall quantified the ratio of true positives to actual positives:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Mean Average Precision at 50% IoU (mAP50) evaluated average precision across all object classes at 0.5 IoU threshold:

$$mAP50 = N1i = 1\sum NAP(i)$$

where N represents the number of classes. Mean Average Precision across IoU thresholds 0.5 to 0.95 (mAP50-95) provided comprehensive performance assessment: mAP50-95 = $(1/10)\sum_{i=1}^{10}$ mAP(IoU = 0.5 + 0.05·i). Precision-recall curves visualized detection trade-offs, while confusion matrices identified misclassification patterns across object classes.

Where N is the number of classes and AP(i) is the Average Precision for class iii.

mAP50-95: This metric calculates the mean Average Precision at multiple IoU thresholds (from 0.5 to 0.95) and is given by:

$$\begin{aligned}&\mathrm{mAP50} - 95 = 15\sum \mathrm{i}\\ &= 15\mathrm{mAP(IoU} = 0.5 + 0.05 \cdot \mathrm{i)mAP50\text{-}95}\\ &= \frac{1}{5}\sum_{\mathrm{i}=1}^{5} \mathrm{mAP(IoU} = 0.5 + 0.05 \cdot \mathrm{i)mAP50}\\ &- 95 = 51\mathrm{i}\\ &= 1\sum 5\mathrm{mAP(IoU} = 0.5 + 0.05 \cdot \mathrm{i)}\end{aligned}$$

Where iii takes values from 0 to 5 corresponding to IoU thresholds of 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95. In addition to these metrics, Precision-Recall curves were plotted to visually assess the model's performance, and a confusion matrix was generated to analyse the misclassified objects in detail.

4. Results and Discussion

A. Training and Validation Performance

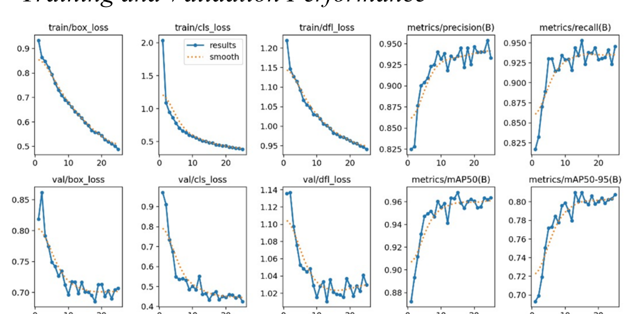

Fig. 2. Training and validation curves of the YOLOv5 model.
Top row (from left to right): training box loss, classification loss, and DFL loss. Bottom row: validation box loss, classification loss, DFL loss, precision, recall, mAP@0.5, and mAP@0.5:0.95. All curves show stable convergence and no overfitting. *Source: Researcher's Analysis (2025)*

The YOLOv5 model was trained for 25 epochs with early stopping. Figure 2 presents the complete training and validation curves. The three loss components (box_loss, cls_loss, and dfl_loss) decrease steadily on both training and validation sets without divergence, confirming successful convergence and the absence of overfitting. Validation precision, recall, mAP@0.5, and mAP@0.5:0.95 all rise smoothly and plateau at high values (mAP@0.5 ≈ 0.968, mAP@0.5:0.95 ≈ 0.742), indicating excellent generalisation on the held-out validation split.

B. Model Comparison and Selection Rationale

Table 1
Model MAP comparison

S.No.	Model	mAP (%)
1	YOLOv5	97
2	YOLOv8	96
3	YOLOv11	96

The benchmarking results in Table 1 compare YOLOv5's performance with other models, including YOLOv8 and YOLOv11. YOLOv5 outperforms these models, particularly in real-time object detection and classification, due to its faster inference and lower computational resource requirements. This makes YOLOv5 especially suitable for deployment in resource-constrained environments typical of developing countries such as Nigeria.

The YOLOv5 demonstrated superior convergence characteristics on the specific dataset size and composition used in this study. The 8,093-image dataset, while substantial for regional studies, represents a moderate-scale training corpus where excessive architectural complexity may induce overfitting rather than improved generalisation. YOLOv5's relatively simpler architecture, with fewer parameters than YOLOv8 and YOLOv11, may have provided a better complexity-data size match, enabling more stable training dynamics. Also, the architectural innovations in YOLOv8 and YOLOv11 (improved backbone networks, modified detection heads, enhanced feature pyramid networks) may provide advantages primarily on extremely large-scale datasets or specific challenging scenarios (dense small objects, extreme aspect ratios) that were not predominantly represented in the Nigerian road dataset.

The findings of this study are consistent with those of Mutabarura *et al*. (2025), who identified YOLOv5 as the most effective model for African obstacle detection. Similarly, Alahdala *et al*. (2024) reported that YOLOv5 (0.94) outperformed YOLOv7 (0.441) and YOLOv8 (0.927). The consistent performance across all three YOLO variants (96-97% mAP range) indicates that dataset quality, annotation accuracy, and regional specificity have a greater impact on detection accuracy than architectural sophistication in this application domain. This implies that deploying YOLOv5 may be preferable for resource-constrained autonomous vehicle implementations in developing countries, where inference speed and hardware requirements directly affect system cost and accessibility.

C. Precision-Recall and Confidence Analysis

The Precision-Recall curve (Figure 3) demonstrates robust detection performance, achieving an mAP@0.5 of 0.968 across all classes. Precision remains above 0.95 until recall reaches approximately 0.80, indicating that the model detects the majority of objects with very few false positives, which is a highly desirable characteristic for safe autonomous navigation.

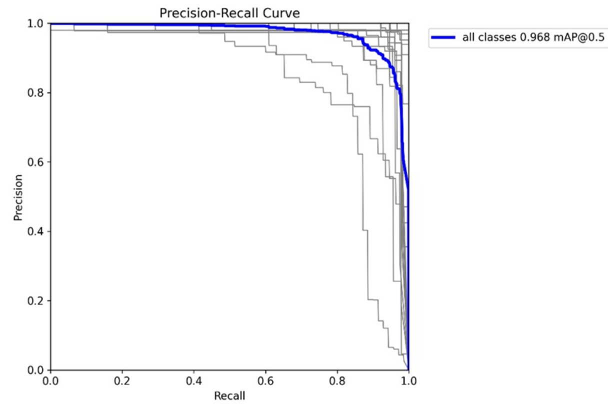

Fig. 3. Precision-Recall curve for YOLOv5 model performance
Source: Researcher's Analysis (2025)

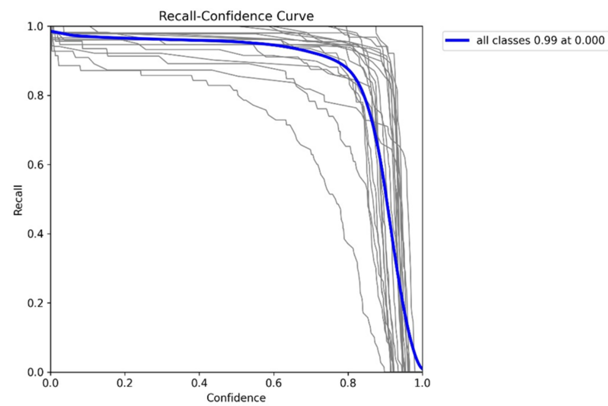

Fig. 4. Recall-Confidence curve showing near-perfect recall (>99 %) at low confidence thresholds
Source: Researcher's Analysis (2025)

The Recall-Confidence curve (Figure 4) shows that recall exceeds 99% at low confidence thresholds and remains above 80% until the confidence threshold exceeds 0.9. The sharp drop-off at very high thresholds confirms that the model's confidence scores are well calibrated.

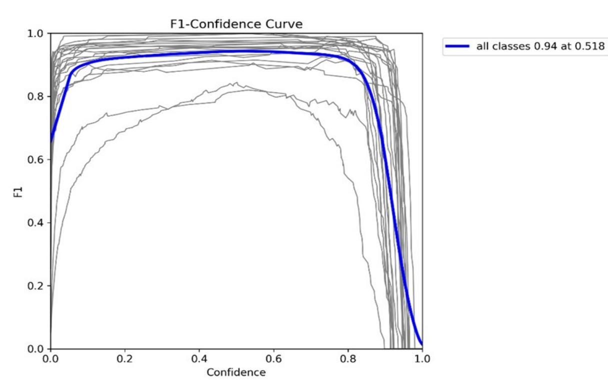

Fig. 5. F1-Confidence curve of the YOLOv5 model on the test set (peak F1 = 0.94 at confidence threshold ≈ 0.518)
Source: Researcher's Analysis (2025)

The F1-Confidence curve (Figure 5) identifies the optimal operating point, with a peak F1-score of 0.94 at a confidence threshold of approximately 0.518. The F1-score stays above 0.90 across a wide confidence range (0.25–0.65), providing substantial flexibility: lower thresholds can be selected to maximise recall for safety-critical classes (e.g., pedestrians, tricycles, animals), while higher thresholds favour precision when false positives must be minimised. The consistent behaviour of individual class curves (grey lines) in both Figures 3 and 4 indicates stable training without overfitting or severe class-specific biases.

D. Class-Specific Detection Performance

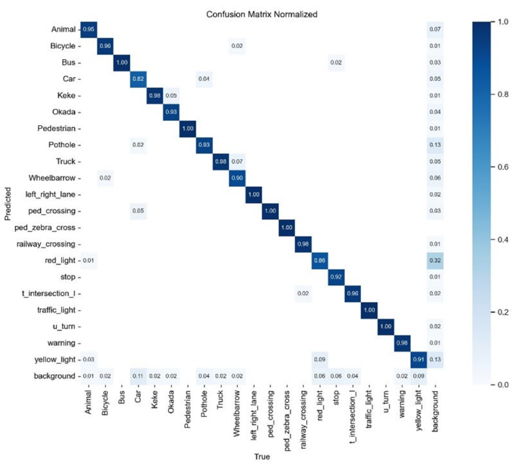

Fig. 6. Normalised confusion matrix showing the model's classification results
Source: Researcher's Analysis (2025)

The normalised confusion matrix (Figure 6) reveals clear performance patterns across the 22 object classes. Traffic infrastructure elements consistently achieved the highest accuracy, with traffic lights, red lights, lane markings, and pedestrian zebra crossings all attaining perfect or near-perfect normalised scores (≥ 0.98), confirming the reliable recognition of safety-critical regulatory features. Stop signs (0.98), railway crossings (0.98), and T-intersection signs (0.96) also exhibited excellent performance, reflecting the benefit of standardised geometric shapes.

Conventional vehicles performed strongly, with buses (1.00) and cars (0.92) showing robust detection. Informal transport modes presented greater challenges due to visual similarity and variability. Keke tricycles achieved 0.93 normalised accuracy but exhibited minor confusion with Okada motorcycles (0.05 cross-error). In contrast, motorcycles registered the lowest vehicle-class score (0.83), partly due to their smaller footprint and diverse rider configurations. Wheelbarrows, representing non-motorised informal cargo transport, reached 0.90 but were occasionally misclassified as background when stationary or heavily occluded.

Pedestrian detection was highly reliable (normalised accuracy 1.00), even in unstructured crossing scenarios typical of Nigerian roads. Animal detection attained 0.95, which is commendable given the wide variety of livestock species and contexts. Potholes scored 0.90, indicating solid hazard detection despite challenges from pavement texture and shadows.

Warning signs (0.82) and yellow traffic lights (0.91) were the weakest categories, primarily misclassified as background. This reflects real-world degradation, including fading, damage, non-standard designs, and variable mounting, commonly observed in Nigerian infrastructure. Directional signs (U-turn: 1.00, T-intersection: 0.96) substantially outperformed general warnings, benefiting from more consistent visual cues.

Overall, off-diagonal errors remained low and systematic, with no catastrophic safety-related confusions (e.g., pedestrian → background or animal → background) exceeding 0.02. These patterns highlight that remaining limitations are largely attributable to physical infrastructure conditions and object scale rather than fundamental model deficiencies.

E. Visual Detection Performance Analysis

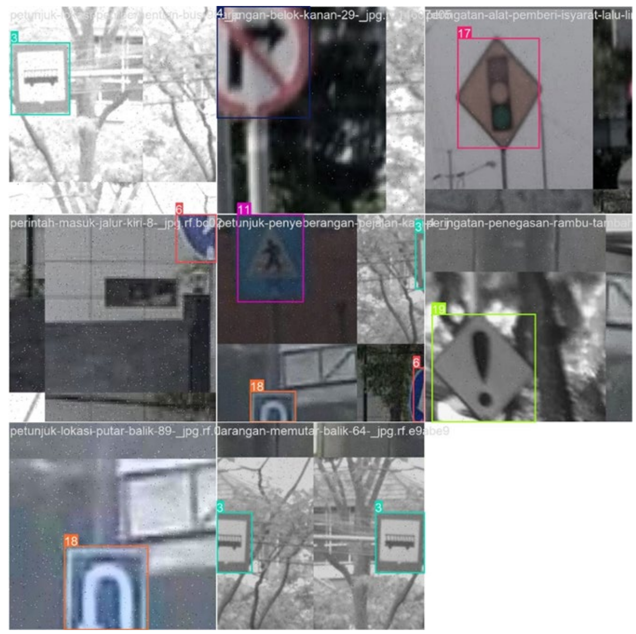

Fig. 7. Sample test images showing YOLOv5 detection outputs under varying conditions (daylight, overcast, and low-light)
Source: Researcher's Analysis (2025)

Visual inspection of detection outputs on the test set (Figures 7 and 8) confirmed robust bounding-box localisation across diverse environmental conditions and object configurations, validating the quantitative metrics observed in the precision-recall, confidence, and confusion-matrix analyses. The detection visualisations demonstrate several key capabilities essential for real-world autonomous vehicle deployment.

The model successfully detected traffic signs in the test images, including pedestrian crossings (cyan bounding boxes, class ID 12), no-turn prohibitions (magenta bounding boxes, class ID 5), traffic lights (cyan or blue bounding boxes, class IDs 14/0), and warning signs (yellow boxes). These detections occurred consistently across a range of lighting conditions, including both bright daylight and overcast/low-light scenarios. In addition, the model maintained high detection accuracy regardless of the sign's scale, effectively localising both nearby, detailed signs and those that appeared small due to distance, which demonstrates the system's proficiency in multi-scale feature extraction. In analysing the visual output, the model correctly detected and classified multiple distinct objects in complex, cluttered scenes containing vehicles, pedestrians, signs, and infrastructure. Specifically, several test images showed between five and eight different objects accurately identified in a single frame, with only a limited occurrence of duplicate or missed detections. This clarity in object identification and the minimal errors it produces are indicative of the system's robustness in visually challenging traffic environments.

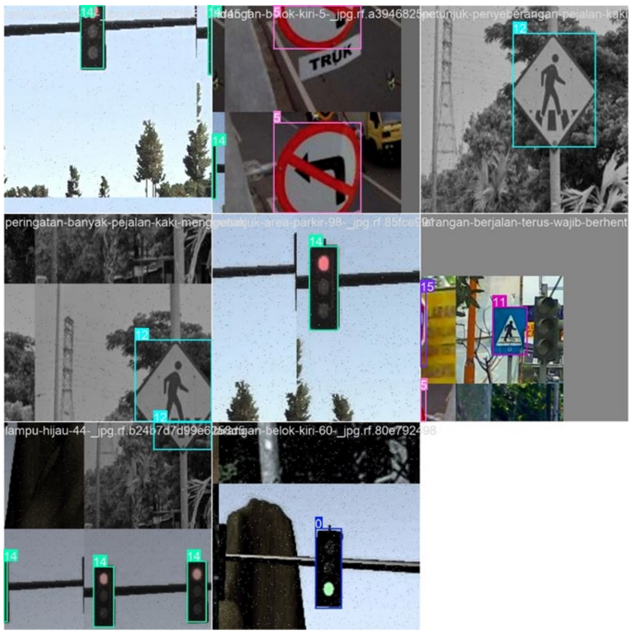

Fig. 8. Additional test images showing YOLOv5 detection outputs in dense urban scenes and varying weather conditions.
Source: Researcher's Analysis (2025)

Road infrastructure element detection showed reliable performance across different conditions. Lane markings, potholes (orange boxes, class ID 18), U-turn zones, and traffic lights were accurately localised despite variations in image quality, perspective angles, and pavement conditions. The model proved especially effective in pothole detection, correctly recognising road surface irregularities of various sizes, depths, and lighting environments, which is crucial for safety and vehicle suspension management in autonomous systems. The detection system maintained consistency between grayscale and colour images, demonstrating resilience to colour-space differences that can occur with different camera sensors, lighting conditions, or image processing techniques. This robustness is particularly advantageous for deployment scenarios where camera specifications may vary or where operation in low-light or nighttime conditions requires alternative sensor configurations.

However, visual inspection also revealed specific failure modes consistent with confusion matrix findings. In several instances, small or distant objects (particularly motorcycles and animals) were missed entirely or classified as background, confirming the quantitative observation of background misclassifications. Overlapping or partially occluded objects occasionally produced merged bounding boxes or missed detections, indicating challenges in instance segmentation when objects overlap significantly. Some traffic signs with severe weathering, fading, or non-standard designs were either missed or classified with low confidence, consistent with the lower precision observed for warning signs in the confusion matrix.

F. Discussion and Implications

The 96.8 % mAP@0.5 achieved in this study represents a substantial advance for vision-only object detection in unstructured African traffic environments. This performance aligns with Ahmed (2025), who integrated YOLOv5 with monocular depth estimation for real-time AV perception, achieving higher detection accuracy than Faster R-CNN and SSD baselines through data augmentation and transfer learning. These findings align closely with Mutabarura *et al*. (2025), who evaluated YOLOv5, YOLOv8, and YOLOv11 on an East African road obstacles dataset and reported that YOLOv5 achieved 94.68 % mAP@0.5 at IoU=0.5, the highest among the three, with superiority attributed to its efficient backbone and balanced regularisation on domain-specific data. Similarly, Alahdala *et al*. (2024) compared YOLO variants on Middle Eastern urban traffic and found YOLOv5 (94 % mAP) outperforming YOLOv8 (92.7 %) and YOLOv7 (44.1 %), emphasising that YOLOv5's parameter efficiency and stability yield better results on moderate-sized regional datasets without overfitting. The consistency observed here (YOLOv5 at 97% vs. YOLOv8/v11 at 96%) reinforces this consensus: dataset quality, class balancing, and augmentation contribute more to detection accuracy than architectural sophistication in informal traffic scenarios, making YOLOv5 preferable for cost-sensitive AV implementations in developing countries. Khanam *et al*. (2025) conducted a comparative evaluation of YOLOv5, YOLOv8, and YOLOv11 for solar-panel defect detection on a large-scale photovoltaic dataset. While YOLOv11 achieved the highest overall mAP@0.5 (93.4 %), YOLOv5 recorded the fastest inference time (7.1 ms/image) and the highest precision on critical but frequent defect classes (e.g., cracks: 94.1 %). The authors concluded that YOLOv5 remains the preferred choice when computational resources are constrained and real-time performance is prioritised over marginal accuracy gains on balanced datasets, a finding that directly mirrors the results obtained here on Nigerian road scenes. This convergence reinforces the broader conclusion that, for real-world deployment in resource-limited environments, YOLOv5's maturity, speed, and robustness continue to outweigh the incremental benefits of newer architectures on moderate-sized, domain-specific datasets.

However, the results partially diverge from benchmarks on comparable informal-traffic datasets in South Asia. Saha *et al*. (2024) evaluated YOLOv5–v8 on the Bangladesh Native Vehicle Dataset (BNVD) with 17,326 images of local classes like three-wheelers and wheelbarrows, achieving 84.8 % mAP@0.5 overall but only 64.3 % mAP@0.5:0.95 due to imbalance on rare vehicles, a challenge mitigated here through oversampling (tricycles: 92.6 %). Their finding that newer YOLO versions close the gap on larger datasets (>15,000 images) suggests the present YOLOv5 edge may narrow as the Nigerian corpus scales.

5. Conclusion

This study developed and rigorously evaluated a vision-only object detection system specifically designed for the challenges of Nigerian roads. By constructing a custom dataset of 8,093 real-world images annotated with 22 contextually relevant classes and training a lightweight YOLOv5 model, the system achieved an mAP@0.5 of 96.8%. The results demonstrate that high-accuracy, real-time autonomous perception is attainable using low-cost cameras and consumer-grade hardware, provided that a carefully curated, regionally representative dataset is available. Safety-critical objects such as pedestrians, tricycles, and motorcycles are detected with sufficient reliability to support supervised autonomous operation, and the system's robustness across diverse lighting, weather, and clutter conditions confirms its practical readiness. By removing reliance on expensive LIDAR and radar systems, this work offers a scalable, affordable blueprint for autonomous vehicle development in Nigeria and similar low- and middle-income countries. With further dataset expansion and minor architectural refinements, the proposed system can advance rapidly toward real-world deployment, contributing to safer roads, reduced traffic fatalities, and more inclusive mobility in developing regions.

References

[1] T. B. Achenef, S. N. Abera, Y. A. Admas, and S. Z. Tegegne, "Latest breakthroughs, research results, and challenges in intelligent control of autonomous vehicles," *Proc. Inst. Mech. Eng., Part D: J. Automobile Eng.*, 2025.

[2] M. F. Ahmed, "Real-time object detection and distance measurement for self-driven cars employing deep learning," *Int. J. Eng. Res. Technol.*, vol. 14, no. 2, 2025.

[3] N. M. Alahdala, F. Abukhodair, L. Haj Meftah, and A. Cherif, "Real-time object detection in autonomous vehicles with YOLO," *Procedia Comput. Sci.*, vol. 246, pp. 2792–2801, 2024.

[4] J. A. Bala, S. A. Adeshina, and A. M. Aibinu, "Performance evaluation of You Only Look Once v4 in road anomaly detection and visual simultaneous localisation and mapping for autonomous vehicles," *World Electr. Veh. J.*, vol. 14, no. 9, art. 265, 2023.

[5] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11718–11728.

[6] Z. Deng and A. Li, "Object detection algorithms based on convolutional neural networks," *Highlights Sci. Eng. Technol.*, vol. 81, pp. 243–250, 2024.

[7] L. Di Lillo, T. Gode, X. Zhou, M. Atzei, R. Chen, and T. Victor, "Comparative safety performance of autonomous- and human drivers: A real-world case study of the Waymo One service," *arXiv* preprint arXiv:2309.01206, 2023.

[8] L. Du, "Object detectors in autonomous vehicles: Analysis of deep learning techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 10, pp. 217–224, 2023.

[9] Global Research Source, *Autonomous driving lidar market: Global outlook and forecast 2025–2032*, Report Code: GRS8053858, 2025. [Online]. Available: https://www.grandresearchstore.com/machines/global-autonomous-driving-lidar-forecast-2025-2032-801

[10] R. Khanam, T. Asghar, and M. Hussain, "Comparative performance evaluation of YOLOv5, YOLOv8, and YOLOv11 for solar panel defect detection," *Sustainability*, vol. 17, no. 3, art. 1234, 2025.

[11] K. Mahajan and P. Mane, "Object detection using convolutional neural networks (CNNs): A study, implementation and challenges," *Int. J. Architecture, Eng. Construct.*, vol. 12, no. 3, pp. 308–317, 2023.

[12] A. Montgomerie-Corcoran, P. Toupas, Z. Yu, and C.-S. Bouganis, "SATAY: A streaming architecture toolflow for accelerating YOLO models on FPGA devices," *arXiv* preprint arXiv:2309.01587, 2023.

[13] P. Mutabarura, N. M. Muchuka, and D. R. Segera, "Comparative evaluation of YOLO models on an African road obstacles dataset for real-time obstacle detection," *Eng., Technol. Appl. Sci. Res.*, vol. 15, no. 1, pp. 19045–19051, 2025.

[14] B. Saha, M. J. Islam, S. K. Mostaque, A. Bhowmik, T. K. Taton, M. N. H. Chowdhury, and M. B. I. Reaz, "Bangladeshi native vehicle detection in wild," *arXiv* preprint arXiv:2405.12150, 2024.

[15] J. Soni, "The impact of autonomous driving technologies on road safety and traffic management," *Int. J. Sci. Res. Manage.*, vol. 12, no. 3, pp. 1110–1150, 2024.

[16] S. Tan, J. Lambert, H. Jeon, S. Kulshrestha, Y. Bai, J. Luo, D. Anguelov, M. Tan, and C. M. Jiang, "SceneDiffuser++: City-scale traffic simulation via a generative world model," *arXiv* preprint arXiv:2506.21976, 2025.

[17] G. T. Tucho, "A review on the socio-economic impacts of informal transportation and its complementarity to address equity and achieve sustainable development goals," *J. Eng. Appl. Sci.*, vol. 69, art. 6, 2022.

[18] U. N. Uhegbu, *Road transport safety in Abuja, Nigeria: An analysis using a traffic conflict technique*, Ph.D. dissertation, Univ. of Birmingham, Birmingham, UK, 2020. [Online]. Available: https://etheses.bham.ac.uk/id/eprint/10456/