International Journal of Inventive Engineering and Sciences (IJIES)
ISSN: 2319-9598 (Online), Volume-13 Issue-9, September 2026

A Systematic Accuracy–Efficiency Evaluation of YOLO11 for Fine-Grained Vehicle Detection in Urban CCTV Traffic

Shashank Mudgal*

Abstract: ***Fine-grained vehicle detection in urban CCTV imagery requires balancing recognition quality and inference efficiency. This paper presents a systematic empirical evaluation of YOLO11 on the BMD-45 urban traffic dataset using the provided experimental results. The study evaluates model scale, input resolution, augmentation configuration, hardware platform, and class-wise performance. For YOLO11n at 640×640, the reported precision, recall, mAP@0.50, and mAP@0.50:0.95 are 91.40%, 87.20%, 93.10%, and 68.20%, respectively, with a mean latency of 18.7 ms and 53.48 FPS on the reported T4 configuration. Scaling from YOLO11n to YOLO11l increases mAP@0.50:0.95 by 5.40 percentage points but increases latency from 18.7 to 48.6 ms. Increasing resolution from 416×416 to 768×768 increases mAP@0.50:0.95 by 5.30 percentage points while latency rises from 11.8 to 27.9 ms. The supplied hardware results range from 77.52 FPS on an L4 configuration to 5.43 FPS on the reported CPU baseline. Under the supplied conditions, the analysis identifies YOLO11n at 640×640 as a practical low-latency operating point [1]-[6].***

Keywords: YOLO11; BMD-45; Vehicle Detection; CCTV; Intelligent Transportation Systems; Edge AI; Real-Time Object Detection; Accuracy–Efficiency Trade-off.

I. INTRODUCTION

Urban traffic-monitoring systems increasingly rely on computer vision to obtain vehicle-level information from CCTV streams. Fine-grained detection is more demanding than generic vehicle detection because it must distinguish visually related categories under viewpoint, scale, occlusion, and illumination variation.

Real-time deployment introduces a second requirement: a detector must achieve sufficient accuracy while maintaining predictable inference latency. This creates an accuracy–efficiency frontier: larger models and higher input resolutions may improve detection quality but require more computation.

YOLO11 is a contemporary real-time object-detection family released by Ultralytics. BMD-45 is relevant because it contains large-scale urban CCTV imagery and a fine-grained vehicle taxonomy.

Manuscript received on 14 August 2026 | First Revised Manuscript received on 19 August 2026 | Second Revised Manuscript received on 02 September 2026 | Manuscript Accepted on 15 September 2026 | Manuscript published on 30 September 2026

*Correspondence Author(s)
Shashank Mudgal*, Department of Electronics and Communication Engineering, Roorkee Institute of Technology, India. Email ID: shashankcoder4321@gmail.com, ORCID ID: 0009-0000-2068-1059

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open-access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

This study characterises how model capacity, image resolution, augmentation, and hardware affect YOLO11 performance for fine-grained traffic detection.

A. Dataset and Experimental Setup

BMD-45 is an India-specific urban CCTV vehicle-detection dataset with 45,986 images, about 481,947 bounding boxes, and 14 vehicle categories. The reported split contains 35,792 training images and 10,194 validation images, with a separate test split whose annotations are withheld.

The primary configuration uses YOLO11n with pretrained initialisation, 640×640 input, a 10,000-image training subset, a 2,000-image validation subset, 100 epochs, batch size 8, seed 42, and the reported Colab T4 platform.

Table I: Primary Experimental Configuration

Parameter	Configuration
Model	YOLO11n
Input	640×640
Initialization	Pretrained
Training subset	10,000 images
Validation subset	2,000 images
Epochs	100
Batch size	8
Seed	42
Primary hardware	Colab T4

II. EVALUATION METRICS

We report performance using precision, recall, mAP@0.50, mAP@0.50:0.95, mean latency, P95 latency, and FPS. Derived quantities are calculated only from the supplied measurements. For latency in milliseconds, FPS follows the reciprocal relationship FPS = 1000/t. Parameter efficiency is used only as a descriptive ratio.

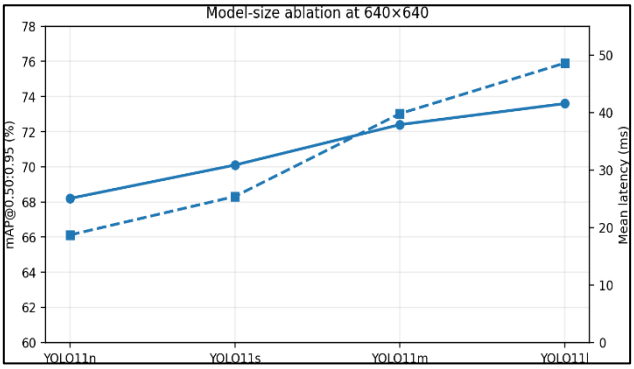
III. RESULTS

A. Primary YOLO11n Result

Metric	Value
Precision	91.40%
Recall	87.20%
mAP@0.50	93.10%
mAP@0.50:0.95	68.20%
Mean latency	18.7 ms
FPS	53.48

The primary YOLO11n configuration achieves 68.20% mAP@0.50:0.95 with 18.7 ms mean latency and a reported 53.48 FPS.

B. Model-Size Ablation

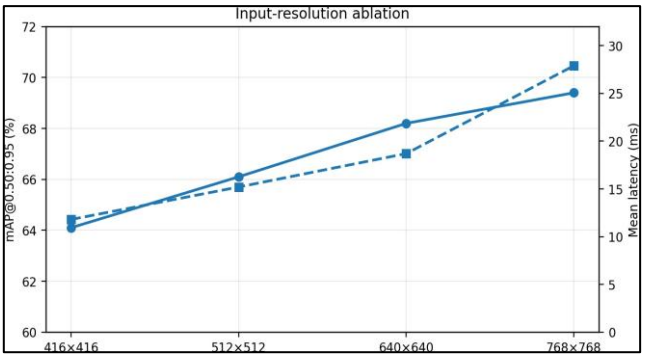


[Fig.1: Accuracy and Latency as YOLO11 Model Size Increases]

Model	Params (M)	mAP50-95	Latency (ms)	FPS
YOLO11n	2.6	68.20%	18.7	53.48
YOLO11s	9.4	70.10%	25.4	39.37
YOLO11m	20.1	72.40%	39.8	25.13
YOLO11l	25.3	73.60%	48.6	20.58

YOLO11l gains 5.40 percentage points of mAP@0.50:0.95 over YOLO11n, but mean latency increases by 29.9 ms. The supplied model-size sweep therefore shows a clear accuracy–latency trade-off.

C. Resolution ablation

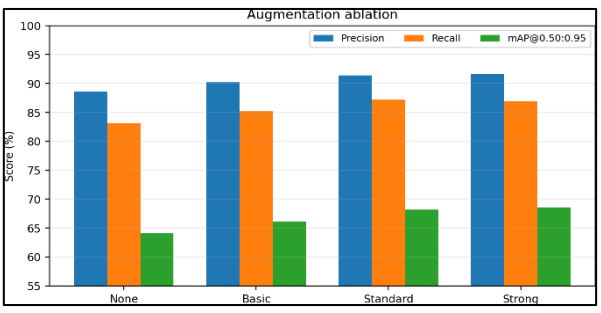


[Fig.2: Effect of Input Resolution on mAP@0.50:0.95 and Mean Latency]

Resolution	mAP50-95	Latency (ms)	FPS
416×416	64.10%	11.8	84.75
512×512	66.10%	15.2	65.79
640×640	68.20%	18.7	53.48
768×768	69.40%	27.9	35.84

Relative to 416×416, 768×768 improves mAP@0.50:0.95 by 5.30 percentage points but adds 16.1 ms latency. The 640×640 setting provides most of the observed gain before the sharpest latency increase.

D. Augmentation Ablation

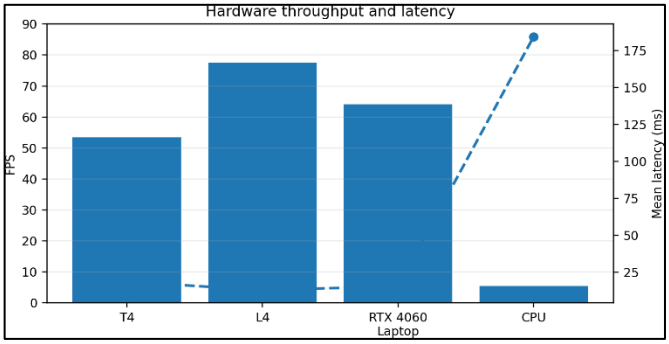


[Fig.3: Precision, Recall, and mAP@0.50:0.95 Across Augmentation Settings]

Augmentation	Precision	Recall	mAP50	mAP50-95
None	88.60%	83.10%	89.90%	64.10%
Basic	90.20%	85.20%	91.80%	66.10%
Standard	91.40%	87.20%	93.10%	68.20%
Strong	91.60%	86.90%	93.40%	68.60%

Strong augmentation produces the highest supplied mAP@0.50:0.95, while recall is slightly below the Standard setting. Augmentation-specific latency was not supplied.

E. Hardware Evaluation

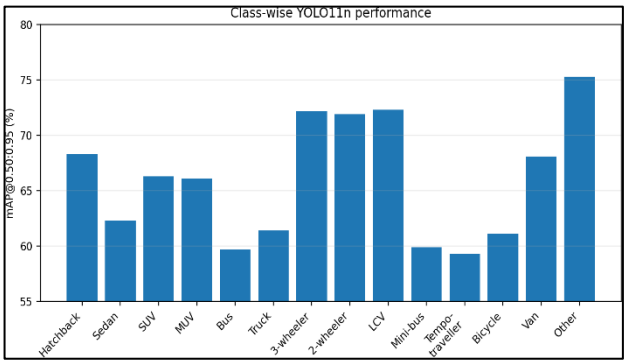


[Fig.4: Reported throughput and Mean Latency Across Hardware Configurations]

Platform	Mean latency	P95 latency	FPS
Colab T4	18.7 ms	23.4 ms	53.48
Platform	**Mean latency**	**P95 latency**	**FPS**
Colab L4	12.9 ms	16.7 ms	77.52
RTX 4060 Laptop	15.6 ms	19.9 ms	64.10
CPU (i7-12700H)	184.2 ms	231.6 ms	5.43

The L4 configuration reports 77.52 FPS, whereas the CPU baseline reports 5.43 FPS. The throughput ratio is 14.28×.

F. Class-wise Analysis



[Fig.5: YOLO11n Class-Wise mAP@0.50:0.95 Across the 14 BMD-45 Categories]

Class-wise mAP@0.50:0.95 ranges from 59.30% for Tempo-traveller to 75.30% for Other. The unweighted mean is 66.01%, with a range of 16.00 percentage points and a population standard deviation of 5.28 percentage points. This variation shows that aggregate mAP does not fully capture category-level behaviour.

IV. DISCUSSION

The supplied measurements show a consistent accuracy–efficiency trade-off. Larger models improve mAP but reduce throughput, while higher input resolution

Retrieval Number: 100.1/ijies.I115113090926
DOI: 10.35940/ijies.I1151.13090926
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

increases accuracy at a progressively larger latency cost. The 640×640 configuration is an attractive operating point in the supplied resolution sweep because the 640•768 transition produces only a 1.20-point mAP gain while adding 9.2 ms latency.

Hardware selection strongly affects deployment feasibility: the reported L4 configuration is 14.28× faster than the CPU baseline. Class-wise results further indicate meaningful heterogeneity across vehicle categories. However, the available results do not include image-level predictions, confusion matrices, object-size strata, or class-frequency statistics, so these data alone cannot establish causal explanations for class differences.

V. LIMITATIONS AND REPRODUCIBILITY

The current evidence is based on the supplied subset/validation measurements. The official test annotations are withheld, so the paper does not establish generalisation on held-out test data. The supplied results also lack repeated random-seed experiments, raw epoch logs, exact software-package versions, checkpoint hashes, and a complete latency-measurement protocol. Before final journal submission, the principal experiments should be repeated across independent seeds, with uncertainty intervals, training curves, confusion matrices, error examples, and, where available, official held-out-test evaluation.

VI. CONCLUSION

This paper presents an accuracy–efficiency characterisation of YOLO11 for fine-grained vehicle detection in urban CCTV traffic. YOLO11n at 640×640 reports 68.20% mAP@0.50:0.95 with 18.7 ms mean latency and 53.48 FPS on the reported T4 configuration. Larger models improve accuracy but incur substantial latency, while 640×640 provides a favourable point in the observed resolution trade-off. Hardware choice can change throughput by more than an order of magnitude. The study therefore provides a practical basis for selecting YOLO11 configurations based on deployment constraints, while clearly identifying the additional experiments needed to support stronger statistical and generalisation claims.

DECLARATION STATEMENT

As the article's author, I verify the accuracy of the following information.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organisations or agencies. This independence ensures the research is conducted objectively and without external influence.
- **Ethical Approval and Consent to Participate:** The data provided in this article are exempt from the requirement for ethical approval or participant consent. The ethical approval and consent are as follows: This study did not involve human participants, human subjects, personal data, or animal experiments. The research used publicly available datasets and therefore did not require ethical approval or informed consent.
- **Data Access Statement and Material Availability:** The data and materials in this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed solely.

REFERENCES

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," CVPR, 2016. https://ieeexplore.ieee.org/document/7780460
2. A. Wang et al., "YOLOv10: Real-Time End-to-End Object Detection," 2024. DOI: 10.48550/arXiv.2405.14458
3. G. Jocher and J. Qiu, "Ultralytics YOLO11," Ultralytics, 2024. https://docs.ultralytics.com/models/yolo11
4. A. Sharma et al., "BMD-45: Bengaluru Mobility Dataset for Large-Scale Vehicle Detection from Urban CCTV," CVPR Findings 2026. DOI: 10.48550/arXiv.2604.24419
5. Y. Zhao et al., "DETRs Beat YOLOs on Real-time Object Detection," 2023. DOI: 10.48550/arXiv.2304.08069
6. T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," CVPR, 2017. https://ieeexplore.ieee.org/document/8099589

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Publisher /Journal and/or the editor(s). The Publisher /Journal and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

© Copyright: All rights reserved.