Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37026

Comparative Evaluation of YOLO11n and RT-DETR-L for Car Detection under Overall and Nighttime Road Conditions

Chu Song

Undergraduate School of Civil, Environmental and Resources Engineering, Kyoto University, Kyoto, Japan
songchu.sc@outlook.com

Abstract. This study evaluated the performance of You Only Look Once (YOLO11n) and Real-Time Detection Transformer (RT-DETR-L) while dealing with car detection tasks based on the Berkeley DeepDrive 100K (BDD100K) dataset and further analyzed their performance under nighttime driving conditions. The models were trained and tested on the homogeneous training and validation sets extracted from BDD100K, with car as the only detection category. Precision, mean Average Precision (mAP), Recall, and inference speed were utilized as the major metrics for evaluation. The results show that RT-DETR-L achieves higher detection precision, recall, and mAP than YOLO11n under both overall and nighttime conditions, while YOLO11n demonstrates a clear advantage in inference speed. Under nighttime conditions, both models exhibit a certain degree of performance degradation compared with the overall validation results, with the mAP@0.5:0.95 metric showing the largest decline for both models. These findings indicate an apparent trade-off between the accuracy and efficiency while executing detection tasks for the two architectures. This study provides a reference for selecting appropriate object detection models under different road and lighting conditions in autonomous driving applications.

Keywords: BDD100K, YOLO11n, RT-DETR-L, car detection, nighttime detection

1. Introduction

In recent years, with the development of intelligent transportation and the influence of artificial intelligence, attention from multiple aspects has been grabbed in the field of autonomous driving. An autonomous driving system requires multiple perception modules, which serve as the foundation of safe driving, to understand the surrounding environment. In perception tasks, object detection is responsible for identifying and locating the key objects on the roads, such as pedestrians, vehicles, traffic signals, and other obstacles. It provides crucial information for the following path planning and decision-making control.

Traditional methods of object detection rely heavily on manual feature extraction methods based on color, texture, and shape information. The results were not able to be very stable. As a result, the Convolutional Neural Network (CNN) gradually became a major technical approach in the object detection field. In comparison, neural network models automatically examine high-level features from the image and significantly improve their performance. Among them, YOLO is one of the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37026

typical one-stage detection algorithms that transforms the detection process into a single prediction task, achieving high detection speeds. It shows advantages in a real-time autonomous driving scenario.

Even though object detection models have developed rapidly over the past few years, it remains somewhat difficult to compare the performance of different models. Current studies mainly focus on proposing new network architectures or improving upon existing models. For instance, to improve the detective performance by incorporating attention mechanisms, optimizing network architecture, or designing new loss functions. However, these studies typically conduct experiments on a single model only, and there are differences in the experimental environment, data processing methods, and evaluation metrics; therefore, direct comparisons between different models are not valid.

With regard to autonomous driving applications, solely focusing on a single evaluation metric, such as accuracy, is not sufficient. The actual driving environment includes different weather and lighting conditions and various road scenarios. Therefore, models should not only achieve high accuracy but also possess good adaptability to the environment and computational efficiency. Thus, it is more meaningful to conduct a comparative analysis of different types of object detection models under standardized conditions.

In order to evaluate object detection performance in autonomous driving environments more accurately, it is necessary to use large-scale datasets containing a wide variety of driving scenarios. As a famous transportation dataset, BDD100K is widely used in autonomous driving research, including real-world road images taken at different times of the day and under varying weather and driving conditions [1]. Compared with datasets collected in a laboratory setting, BDD100K provides a more accurate reflection of various complicated conditions that an autonomous driving system may face. Therefore, in the later part of this study, BDD100K will serve as a research platform for various models to evaluate their performance.

This paper established a BDD100K-based comparison framework for target spotting models and conducted a standardized evaluation of representative deep learning detection models. It further uses metrics such as mAP, precision, and inference speed to comprehensively analyze the detection performance of different models. Finally, conduct further research into the impact of different environmental conditions, such as changes in lighting, on model performance, thereby assessing the robustness of different models in real-world autonomous driving scenarios. Through the above analysis, this paper aims to provide some guidance on the selection and practical application of autonomous driving object detection models.

2. Related work

As a crucial and major aspect of computer vision research, object detection is receiving widespread attention. Early object detection methods primarily relied on manual characteristic representations and traditional machine learning classifiers. However, with the development of machine learning, especially the introduction of CNN, deep learning models were able to learn key features through large-scale datasets automatically and further improve their accuracy and robustness greatly. The attention mechanism introduced by the Transformer also pushes the development of object detection. In recent years, plenty of neural network-based frameworks have been introduced and widely applied in the field of autonomous driving.

Li et al. investigated vehicle and pedestrian detection using an improved RT-DETR model, aiming to enhance detection performance in complex road environments [2]. Their work optimized the RT-DETR framework to improve feature representation capability and detection accuracy while the surrounding conditions are challenging and tough for the models to work properly, such as

DOI: 10.54254/2755-2721/2026.37026

background interference, scale variation, and object occlusion. Their experimental results have shown that the improved model achieved more outstanding detection performance and, at the same time, maintained the inference speed.

In addition to Transformer-based detectors, the YOLO family has applied approaches for object detection at a real-time level. Rami et al. investigated real-time road object detection in smart city environments by introducing a quantized version of YOLOv11 [3]. Their study focuses on reducing model complexity and improving deployment efficiency through quantization techniques, enabling the detector to achieve faster inference while keeping outstanding detection results. Josdaan et al. and Lee et al. also evaluate the model performance on object detection based on the YOLO family [4, 5]. Besides, Liu et al. and Lei et al. also introduced some enhancement methods for the YOLO model in the object detection field [6, 7]. These studies have demonstrated the promising performance of different detection models for object detection tasks. For some specific scenarios in transportation object detection, such as extreme weather, Yue et al. then investigated the performance of YOLOv8 in rainy and foggy conditions, which show challenges for the model [8]. By changing the detection target, Chen et al. evaluated the small object detection tasks of the models [9]. However, although similar comparison of different object detection models was conducted in other fields, such as dental medicine by Eftimie et al., a comprehensive, systematic and in-depth comparison among the latest different model architectures for transportation object detection is still needed [10].

3. Methodology

3.1. Research framework

The target of this study is to compare the autonomous driving performance of different object detection models. Uniformed dataset, data division, and evaluation indices are utilized to train and test the two different object detection models: YOLO11 and RT-DETR. The two models represent two different technical routes. YOLO11 adopts a single-stage detection framework based on convolutional neural networks. While RT-DETR achieves a transformer architecture based on an end-to-end object detection framework.

The experiment is divided into two levels. First, the two models are evaluated based on the general dataset regardless of different scenarios based on indices such as Precision, Recall, and mAP. Based on the overall performance, the study further divides the dataset based on the scenarios in the labels of the BDD100K dataset and carries out a conditional performance comparison for the dark driving environment. Finally, combine the overall detection accuracy, robustness in the dark environment, and the computing efficiency to comprehensively evaluate the advantages, limitations, and adaptability of YOLO11n and RT-DETR-L.

3.2. Dataset and data preparation

This study utilizes the BDD100K dataset as the resource for training and evaluation of object detection models. BDD100K includes many actual driving images covering divergent road and weather conditions throughout the day, which is beneficial to reflect the performance of models in complicated visual perception tasks. Considering that this study mainly focuses on the different performances between different target detection models, and the limitation of experimental calculation resources and training costs, the study does not use the complete BDD100K dataset but randomly extracts some images from it.

DOI: 10.54254/2755-2721/2026.37026

This study limits the detective task to a single category of car detection, and only the car category in the label is retained. Selecting cars as the study target is because cars are the main targets in the driving environment, and this category is widely distributed in the images in the BDD100K dataset, providing a stable foundation for training and testing.

In the formal experiment, 5000 images are extracted randomly from the BDD100K dataset, and 1000 images are utilized as validation data. For each image, only the category of the car in the label is extracted. Original BDD100K data is screened and formatted to meet the training requirements of the target detection framework used.

To ensure the equality of different models, the same dataset is utilized for training and validation. Input images are processed uniformly according to the model training requirements, while maintaining consistent data preprocessing within the feasible range.

3.3. Object detection models

In order to compare different models' performance in detecting objects, this study selects YOLO11 and RT-DETR as the experimental models. Both can achieve real-time object detection, but they show differences in core framework and mechanism. YOLO11 continues the one-stage target detection framework based on CNN. RT-DETR, however, is based on a transformer framework. Therefore, under the same conditions and detection tasks, the difference in precision, performance, and efficiency of the two models can be analyzed.

YOLO11 is a real-time object detection model from the YOLO family. The model uses a backbone network to sort out multi-level visual characteristics from images in the dataset and further integrates features at different scales to predict object classes and bounding boxes. In this study, YOLO11 was applied to single-class car detection on road images from the BDD100K dataset. Transfer learning was performed using pretrained weights, and the model was fine-tuned on the vehicle detection dataset constructed for this study. Only the car class was retained during training, allowing the model to specifically pay attention to the learning vehicle features in road scenes.

Unlike the traditional YOLO family, which mainly relies on convolutional operations for feature extraction and prediction, RT-DETR introduces the Transformer mechanism into this field and uses an encoder and decoder to model the relationships between images and objects. In this study, RT-DETR was trained using the same BDD100K dataset as YOLO11. After training, its performance was examined in the same way as YOLO11. By controlling the training data, detection category, and evaluation method, this study compares the features of the above-mentioned models on the same road vehicle detection task and further analyzes the characteristics of CNN and Transformer-based detection architectures.

3.4. Experimental setup

The research was executed on the device equipped with Apple M5, and the models were trained and tested based on PyTorch 2.8.0 and the Ultralytics framework. The research utilized YOLO11n and RT-DETR-L as representative models. The two models both utilized the same BDD100K dataset, including a random 5000 training images and 1000 validation images. The detection target was 'car'. The input images of the two models were all set to 640 by 640 pixels, and the random seed was set to 42.

Since the computational resources of the two models show diversity, YOLO11n's batch size was set to 8, while for RT-DETR-L, it was 4. Each model was trained for 10 epochs. Adam with Weight decay (AdamW) optimizer was applied to RT-DETR-L, and the initial learning rate was set to

DOI: 10.54254/2755-2721/2026.37026

0.0001. For the two models, the best weights were kept for later evaluation, instead of directly utilizing the weights.

3.5. Evaluation metrics

The study utilized Precision, Recall, mAP@0.5, and mAP@0.5:0.95 as major evaluation metrics for car detection. Precision measures the proportion of correct detections of predicted vehicles. Recall measures the proportion of actual vehicles that are successfully detected by the models. mAP@0.5 means the mean Average Precision at an Intersection over Union (IoU) threshold of 50%, whereas mAP@0.5:0.95 calculates the average performance across multiple IoU thresholds. In addition to the accuracy of the object detection, speed is also a very important metric. The inference time of every image during the validation stage was also recorded to compare the computational efficiency of YOLO11n and RT-DETR-L.

3.6. Overall performance evaluation

This study compares YOLO11n and RT-DETR-L in terms of both overall performance and nighttime performance. First, the best weights saved during the training process of each model were utilized for evaluation on the validation dataset. Based on the overall performance evaluation, the study further utilized the time-of-day attributes provided by BDD100K labels and constructed a nighttime validation subset. No additional training was conducted. Instead, the best model weights obtained from the overall training were directly evaluated on this subset. By comparing the performance changes between the complete validation set and the nighttime subset, the vehicle detection capability and performance stability of YOLO11n and RT-DETR-L under low-light conditions were further analyzed.

4. Results and analysis

4.1. Overall detection performance

Table 1 demonstrates the integrated detection performance of YOLO11n and RT-DETR-L on the BDD100K car-only validation set. The validation set contains 1000 driving images with 10222 car instances. The results of RT-DETR-L were better than YOLO11n in all four detection metrics, including Precision, Recall, mAP@0.5, and mAP@0.5:0.95. YOLO11n, however, achieved 0.760 in the Precision part. The Recall of its result was 0.552, whereas RT-DETR-L achieved 0.807 and 0.688, respectively.

In terms of average precision, YOLO11 obtained an mAP@0.5 of 0.644 and mAP@0.5:0.95 of 0.371, while RT-DETR-L achieved 0.767 and 0.452, respectively. Compared with YOLO11n, RT-DETR-L improved mAP@0.5 by 0.123 and mAP@0.5:0.95 by 0.081. As a result, RT-DETR-L provides stronger overall car detection performance and maintains better localization performance under the current experimental situation.

Table 1. Overall detection performance comparison between two models

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Speed ms/image
YOLO11n	0.760	0.552	0.644	0.371	2.6
RT-DETR-L	0.807	0.688	0.767	0.452	35.4

DOI: 10.54254/2755-2721/2026.37026

One of the notable differences between these two models was observed in Recall. The Recall of RT-DETR-L was 0.688, which was 0.136 higher than the 0.552 achieved by YOLO11n. This states that RT-DETR-L spots a higher rate of ground-truth vehicles. At the same time, the Precision increased from 0.760 for YOLO11n to 0.807, indicating that the higher Recall was not accomplished through a substantial increase in false detections but was accompanied by an improvement in overall detection precision.

Although RT-DETR-L obtained higher detection accuracy, YOLO11n showed its advantage in detection efficiency. During validation, the average inference time of YOLO11n for each image was 2.6 ms, compared with RT-DETR-L's 35.4 ms per image. Based on the recorded information in the current experimental environment, YOLO11n was approximately 13.6 times faster per item. Therefore, the results demonstrate a clear trade-off between precision and speed while detecting objects for the two detection models.

Overall, RT-DETR-L shows stronger ability in detection accuracy and recall, whereas the low detection latency of YOLO11n provides an advantage for applications when computational resources are limited, or the requirements of real-time are stronger.

4.2. Nighttime condition performance

To further test the performance of the two models while detecting objects under nighttime conditions, images labeled with 'timeofday=night' were selected from the original 1000-image validation set to construct a nighttime validation subset. The subset contains 388 nighttime road images with 3476 car instances. Both of the two models utilized their best trained weights obtained from the previous stages. Therefore, this experiment reflects the generalization performance of the same trained models under nighttime conditions.

Table 2. Detection performance comparison under nighttime conditions

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Speed ms/image
YOLO11n	0.719	0.541	0.638	0.336	3.9
RT-DETR-L	0.768	0.658	0.738	0.395	35.5

Table 2 shows the results under nighttime conditions. The results indicate that RT-DETR-L still shows more outstanding performance across all major detection metrics than YOLO11n. Under nighttime conditions, the Precision of YOLO11n was 0.719, and the Recall was 0.541. For mAP, 0.5 was 0.638, and 0.5:0.95 was 0.336. In comparison, the results of RT-DETR-L were 0.768, 0.658, 0.738, and 0.395, respectively. These results state that RT-DETR-L remains very strong in detecting cars even under more challenging nighttime conditions, with noticeable advantages in Recall and mAP.

Compared with the overall validation results, both models showed a certain degree of degradation in performance under nighttime road conditions. The mAP@0.5 of YOLO11n decreased from 0.644 to 0.638, representing a reduction of only 0.006, while its mAP@0.5:0.95 decreased from 0.371 to 0.336, a reduction of o.o35. For RT-DETR-L, mAP@0.5 reduced from 0.767 to 0.738, while mAP@0.5:0.95 dropped from 0.452 to 0.395. Both models exhibited a larger reduction in the mAP@0.5:0.95 metric, showing that nighttime conditions may have a stronger effect on bounding-box localization quality.

One of the noticeable things is that the minor decrease in mAP@0.5 of YOLO11n shows a relatively stable car detection capability of the model under a less strict threshold. Although RT-DETR-L experienced a larger performance reduction, its nighttime detection results remained

DOI: 10.54254/2755-2721/2026.37026

consistently higher than those of YOLO11n. Therefore, the nighttime experiment demonstrates the different characteristics of the two models: RT-DETR-L maintains an advantage in detection accuracy and target recall, and YOLO11n shows relatively stable basic detection performance together with higher detection efficiency.

5. Conclusion

The study tested the performance of YOLO11n and RT-DETR-L on car detection tasks based on the BDD100K dataset and further evaluated the nighttime condition performance. The results show that the Transformer-based model RT-DETR-L is much stronger in detection precision in every accuracy aspect, and remains the advantage under nighttime conditions. In comparison, YOLO11n shows higher detection efficiency even though its precision is relatively lower. Besides, both models show relative performance degradation under nighttime conditions, especially for mAP@0.5:0.95. Overall, the results of the experiments show an apparent trade-off between the two models on precision and the speed of detection. This indicates that the selection of the models should be made comprehensively according to the requirements for accuracy, real-time, and environmental adaptability.

The current study and its experiments were conducted using a relatively lightweight setup. Future studies should further enlarge the training dataset to examine more about the stability and reliability of the current results, and also may involve some other scenarios, such as snowy, foggy, or other complicated road conditions, to comprehensively understand the environmental adaptability of different models. Meanwhile, the models selected this time remain diverse in scales and computational complexity, so models with more similar scales can be selected for comparison, and their efficiency can be tested on different hardware platforms, so as to examine the applicability of different architectures in the field of object detection more comprehensively.

References

[1] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., & Darrell, T. (2020). BDD100K: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2636–2645).

[2] Li, J., & Wang, J. (2025). Research on vehicle and pedestrian detection based on improved RT-DETR. *International Journal of Advanced Network, Monitoring and Controls, 10*(2), 85–93.

[3] Rami, C., Lamaakal, I., Ouahbi, I., El Makkaoui, K., & Maleh, Y. (2025). Quantized YOLOv11 for real-time road object detection in smart city environments. In *2025 International Conference on Circuit, Systems and Communication (ICCSC)* (pp. 275–280).

[4] Josdaan, J., Tamsil, V. C., & Harefa, J. (2025). CBAM-enhanced YOLO11n: A lightweight and efficient model for solid waste detection compared to YOLOv9n, YOLOv10n, and RT-DETR. In *2025 International Conference on Information Technology Research and Innovation (ICITRI)* (pp. 1–6).

[5] Lee, Y.-H., & Lee, W.-B. (2026). From YOLOv1 to YOLO26: A unified comparative review of the evolution, architecture, and applications of YOLO-based object detection frameworks.

[6] Liu, Z., Li, L., Hu, Y., & Hu, S. (2025). Enhanced road object detection with DFPD-YOLO: Focusing on small and occluded targets. *The Journal of Supercomputing, 81*(14), 1329.

[7] Lei, Y., Shan, C., & Ge, W. (2025). EdgeCaseDNet: An enhanced detection architecture for edge case perception in autonomous driving. *PLOS ONE, 20*(12), e0338638.

[8] Yue, X., Yu, J., Huang, H., & Xu, J. (2026). Experimental design for intelligent traffic sign detection in rainy and foggy weather based on YOLOv8. *Experimental Technology and Management, 43*(5), 85–91.

[9] Chen, C., & Wu, B. (2026). Research on small object detection algorithm based on cross-scale frequency guidance and dual-pathway collaboration. Available at SSRN 7190857.

[10] Eftimie, S., Ileni, T., Iacob, L., Hedeșiu, M., & Dioșan, L. (2025). Tooth-level detection and mapping of dental pathologies on panoramic radiographs using YOLOv11 and RT-DETR. *MethodsX, 15*, 103696.