Enhancing YOLO with an Improved Module for License Plate Detection

Dian Ning, Dong Seog Han*

School of Electronic and Electrical Engineering, Kyungpook National Univ.

ningdian@knu.ac.kr, dshan@knu.ac.kr*

개선된 모듈을 이용한 YOLO 번호판 탐지 모델 구현

닝디엔, 한동석*
경북대학교

Abstract

This paper discusses a more effective method for identifying car license plates using a modified version of the you only look once v9 (YOLOv9) model. Existing approaches face difficulties in accurately detecting license plates under crowded and cluttered environments, mainly because of insufficient spatial awareness and ineffective handling of multi-scale objects. The main change is the addition of a spatial and depth module to YOLOv9, which helps reduce unnecessary spatial information and enhances the model's ability to perceive features at different scales. Tests showed that our modified YOLOv9 achieved 98.5% accuracy on the UFPR-ALPR dataset. making the model faster and more suitable for real-time applications in intelligent.

Ⅰ. Introduction

License plate recognition (LPR) is significant for modern transportation systems. It assists with tasks such as tracking vehicles, monitoring traffic, enforcing laws, and managing parking systems automatically. One key job in LPR is accurately identifying license plates in pictures. Machine learning approaches have been proven to be more effective at handling poor-quality images and diverse environments compared to traditional methods that relied on basic image features.

However, it has focused on several problems, such as the lack of datasets and the need for more efficient models. Many current workers utilize data from fixed cameras in parking lots or for security purposes. This makes it hard to apply these methods to actual driving situations. We need to consider the model's performance at the same time. Models created using this type of data often struggle when there are numerous cars, changing lights, and cluttered backgrounds.

To address these issues, this paper proposes a more effective license plate detection model utilizing a refined you only look once v9 (YOLOv9) [1] setup. We enhance YOLOv9's neck structure by introducing a spatial and depth part, which facilitates the mixing of features across different scales and reduces wasted space, rather than relying on a new dataset. Tests on the UFPR-ALPR data indicate that our enhanced setup finds license plates more precisely and faster than the standard YOLOv9. This suggests its potential application in real-world driving LPR applications.

Ⅱ. Method

License plate recognition involves capturing frames that often include multiple objects and background areas. The key challenge is pinpointing the license plate within this complex scene. This work uses the YOLOv9 framework as the main detection model to isolate license plates.

The YOLOv9 structure has four parts: input, backbone, neck, and prediction. The input part adjusts images to 640×640×3 and employs various data augmentation methods. The backbone network extracts high-level features through convolution and the RepNCSP-ELAN4 structure (RepNCSP is the reparametrized net with cross-stage partial connection and ELAN is efficient long-range attention network), which combines RepNBottleneck blocks with a generalized efficient layer aggregation network (GELAN) to enhance gradient flow and utilize features more effectively. The neck part utilizes multi-scale path aggregation network (PAN) layers to enhance feature mixing across different resolutions. The prediction part outputs bounding boxes using intersection over union (IoU) based loss functions.

To improve the feature extraction capability of the YOLOv9-t backbone and neck, the traditional AConv block has been replaced with a spatial and depth extractor module (SDE). The original AConv structure uses average pooling, followed by a convolution layer. This reduces spatial resolution, which can result in the loss of fine-grained details that are crucial for detecting small objects, such as license plates.

In contrast, the changed module reorganizes spatial information into the channel dimension before applying a lightweight convolution. This transformation preserves local texture and boundary information while reducing the number of spatial operations. Consequently, the new neck strengthens multi-scale feature representation and improves gradient flow across layers.

Ⅲ. Experiments

We use the UFPR-ALPR dataset [3] for our experiments, which comprises 4,500 images captured in Brazil. The UFPR-ALPR has 40% of its images in the training dataset, 20% in the validation dataset, and 40% in the testing dataset.

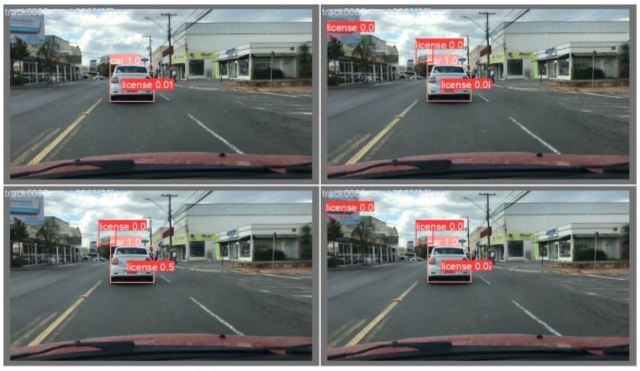
We selected Yolov9-T as the base model. Our input image size is 640 x 640 x 3, and we disabled various image enhancement features during training to ensure a fair comparison. Training consists of 100 epochs, and all other settings are the same as the official YOLOv9 hyperparameters unless otherwise specified. All experiments were conducted on a PC with a Nvidia GTX 4080 GPU.

To test the performance of the designed neck architecture, we used partial and complete two-variant models. The partial one replaces the three AConv modules with the SDE modules in the backbone, while the complete one replaces all five AConv modules in the backbone and neck. Evaluation results of Yolov9-T and the two variants based on the UFPR-ALPR dataset are shown in Table 1.

TABLE I.
COMPARING THE EVALUATION METRICS BETWEEN TWO MODELS USING THE YOLOV9-T MODEL
.

Backbone	Precision	Recall	mAP50	mAP50:95
ACONV	0.985	0.951	0.982	0.786
SDE	0.987	0.969	0.991	0.804

It shows the improvement achieved by introducing the new module into the YOLOv9-T backbone and neck. Compared with the original YOLOv9-t, it has a notable increase in both precision and recall. Besides, mAP50 rises from 0.982 to 0.991 (+0.9%), and mAP50:95 improves from 0.786 to 0.804(+1.8%). The visualization is shown on Fig.1.



(a) The original model

(b) The new model with SDE modules.

Fig. 1. The visualization comparison of two models (a) and (b).

IV. Conclusion

In this study, we use efficient modules to improve the YOLOv9 model and achieve better performance. We will strive to enhance performance and speed further in the future.

ACKNOWLEDGMENT

This work was supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korea government(Ministry of Science and ICT) (RS-2025-25444562).

REFERENCES

[1] Jiang, Y., Tan, Z., Wang, J., Sun, X., Lin, M., & Li, H. (2022). GiraffeDet: A heavy-neck paradigm for object detection. *arXiv preprint arXiv:2202.04256*.

[2] Wang, C. Y., Yeh, I. H., & Mark Liao, H. Y. (2024, September). Yolov9: Learning what you want to learn using programmable gradient information. In *European conference on computer vision* (pp. 1-21). Cham: Springer Nature Switzerland.

[3] Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., & Menotti, D. (2018, July). A robust real-time automatic license plate recognition based on the YOLO detector. In *2018 international joint conference on neural networks (ijcnn)* (pp. 1-10). IEEE.