INTELLIGENT TRAFFIC SIGN DETECTION USING YOLOV9

Afu Ichsan Pradana[1*], Harsanto[2], Joni Maulindar[3]
Universitas Duta Bangsa Surakarta, Indonesia[1,2,3]
*Correspondence Email : afu_ichsan@udb.ac.id

ABSTRACT

This research examines the automatic detection and classification of traffic signs using artificial intelligence (AI) and computer vision technologies. As urban traffic increases, quickly and accurately recognizing traffic signs becomes a challenge, especially under adverse conditions such as bad weather and limited visibility. Conventional technologies that rely on human vision are prone to errors, so an automated solution is needed. This research uses the YOLOv9 algorithm for real-time traffic sign detection, utilizing the Generalized ELAN (GELAN) architecture that combines the advantages of CSPNet and ELAN for efficiency and accuracy. The dataset used consists of 1924 images processed through various stages, including data augmentation and normalization. The model was trained for 15 epochs with fairly high accuracy results in the prohibitory, danger, and mandatory sign categories. However, there were still some misclassifications, especially in the prohibitory category which was sometimes mistakenly detected as another category or background. Overall, the model performed well in detecting traffic signs in various environmental conditions, but still needs improvement to increase accuracy in certain cases.

KEYWORDS

Traffic Sign Detection, YOLOv9, Computer Vision, Artificial Intelligence

INTRODUCTION

Modern traffic is one of the biggest challenges for urban infrastructure development and transportation management. The increase in the number of vehicles on the road has led to an increased risk of traffic accidents and congestion. One of the factors causing accidents is the inability of drivers to recognize traffic signs quickly and accurately. Traffic signs play an important role in regulating driver behavior and maintaining road safety (Zhou et al., 2019). However, under certain conditions such as bad weather, limited visibility, or driver ignorance, these signs often go unnoticed or even ignored (Duan et al., 2020).

Conventional technologies used in traffic sign recognition often rely on human vision which is prone to errors and limitations, such as fatigue or lack of focus (Sermanet & LeCun, 2011). In addition, manual detection of traffic signs by humans is time-consuming and inefficient, especially in large cities with high traffic density (Larsson et al., 2019). Therefore, there is a need for technological solutions that are able to detect and recognize traffic signs automatically, quickly, and accurately to help reduce the risk of accidents and support a safer and smarter transportation system.

The use of artificial intelligence (AI) and computer vision-based technologies has opened up new opportunities in the development of automated traffic sign detection systems (Khalid et al., 2020). Previous research on traffic sign detection using computer vision technology has been conducted, with various approaches and algorithms applied. For example, feature-based detection methods such as SIFT (Scale-Invariant Feature Transform) and HOG (Histogram of Oriented Gradients) have been used to detect traffic signs, but these methods often lack efficiency in terms of speed and accuracy, especially in real-time conditions. Additionally, CNN (Convolutional Neural Network) based algorithms such as Faster R-CNN and YOLO (You Only Look Once), have been shown to provide accurate results, but have drawbacks in terms of processing speed (Redmon et al., 2016).

With the advent of the YOLO algorithm, particularly earlier versions such as YOLOv3 and YOLOv5, many studies showed significant improvements in terms of detection speed without compromising accuracy (Bochkovskiy et al., 2020). However, YOLOv9 offers further improvements in real-time detection, with better performance under complex conditions such as viewpoint and lighting variations, which is one of the focuses in this study.

RESEARCH METHOD

This research uses the python programming language, with Tesla T4 15360MiB GPU processing device specifications, and CUDA Version: 12.2.

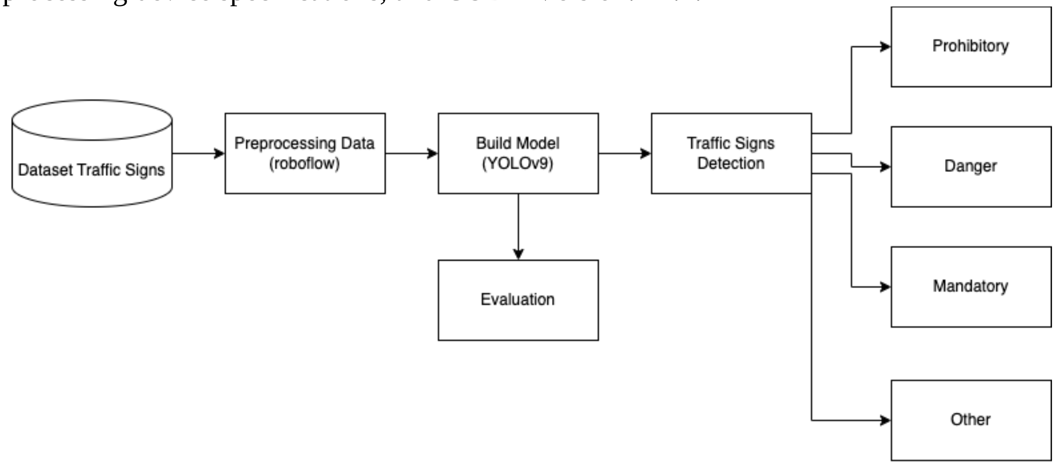

Figure 1. Proposed Methodology

1. **Dataset Traffic Signs**
The process starts by collecting or selecting a dataset of traffic signs. This dataset contains images of various traffic signs that are used to train the model to detect and classify them into different categories. The dataset used in this research is from kaggle (Sichkar, 2020). Total dataset used in this research is 740 images, of which 1211 images have been annotated. The distribution of the dataset can be visualized in the following Figure 2.

Figure 2. Dataset Distribution

2. **Preprocessing Data**
Before the dataset is input into the model, data preprocessing is performed. Roboflow is mentioned here as a tool for preparing data, which typically includes tasks such as image resizing, augmentation (increasing data variety), and data normalization for input consistency. This resulted in the final data distribution used in this study of 1924 images presented in Figure 3. Below.

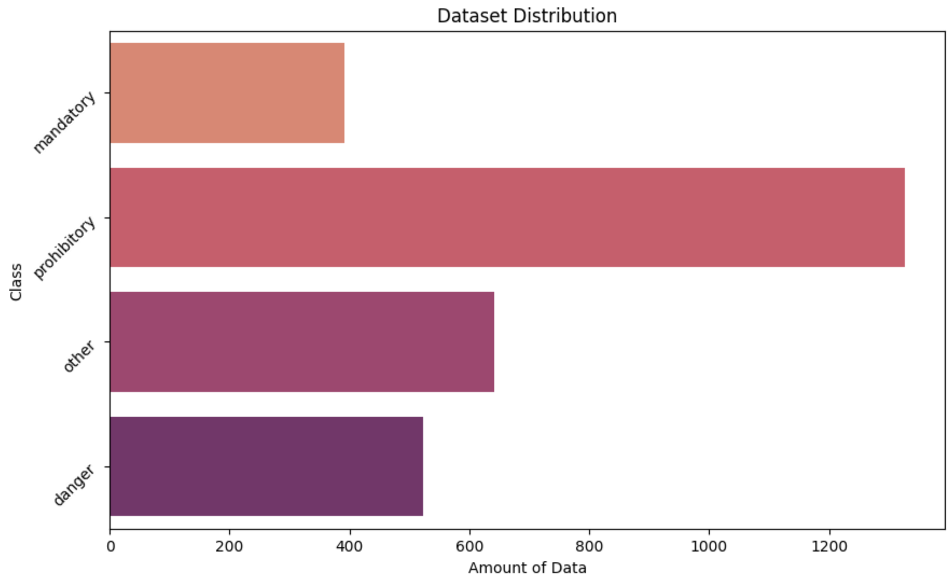

Figure 3. Final Dataset Distribution

3. **Build Model**
The latest version of the "You Only Look Once" model, used to build a traffic sign detection system. YOLOv9 is designed for object detection and can process images in real-time, detecting and classifying objects (in this case, traffic signs) in the image. The following is the architecture used which is presented in Figure 4.

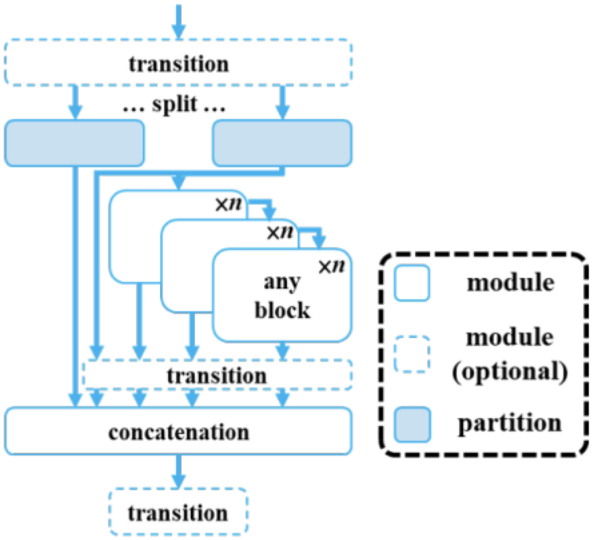

Figure 4. GELAN Architecture

The network architecture used is Generalized ELAN (GELAN), which combines the advantages of CSPNet and ELAN to create a network that is efficient in terms of size, inference speed, and accuracy (Wang et al., 2024). By extending the capabilities of ELAN, which previously only used convolution layers, GELAN is able to utilize different types of computing blocks, making it more flexible and optimal for various applications. The architecture is designed to maximize performance with efficiency and speed in mind without sacrificing accuracy. This architecture produces 621 layers, 25440156 parameters, 25440140 gradients, 103.2 GFLOPs. In the process of making this model using SGD optimizer (lr=0.01) with parameter groups 154 weight (decay=0.0), 161 weight (decay=0.0005), 160 bias. Then the albumentations used Blur(p=0.01, blur_limit=(3, 7)), MedianBlur(p=0.01, blur_limit=(3, 7)), ToGray(p=0.01, num_output_channels=3, method='weighted_average'), CLAHE(p=0.01, clip_limit=(1, 4.0), tile_grid_size=(8, 8).

4. **Evaluation**
   After detection, the performance of the model is evaluated. This step usually involves testing the model on validation or test datasets and calculating metrics such as Precission, Recall, mAP50, and mAP50-95, as well as to measure how well the model detects and classifies traffic signs using a confusion matrix.
5. **Traffic Signs Detection**
   Once the model is trained, it is applied to detect traffic signs in images. The model analyzes the input images to find traffic signs and classifies them into the appropriate categories.
6. **Classification Category**
   The detected traffic signs are classified into four main categories :
   a. **Prohibitory**
      Signs indicating actions that are not allowed (e.g., “No entry,” “No parking”)

b. **Danger**
Signs that warn about potential hazards or risks (e.g., “Slippery road,” “Sharp bend”)

c. **Mandatory**
Signs that indicate the action to be taken (e.g., “Turn right,” “Pedestrian crossing”)

d. **Other**
Other types of signs that do not fall into the above categories.

RESULT AND DISCUSSION

1. Model training results

This model was trained for 15 epochs, resulting in 467 layers, 25414044 parameters, 0 gradients, 102.5 GFLOPs. The results are presented in Table 1.

Table 1. Model training results

Class	Images	Instances	Precision	Recall	mAP50	mAP50-95
all	148	240	0.974	0.921	0.967	0.768
Danger	148	43	0.975	0.912	0.975	0.79
mandatory	148	31	0.933	0.903	0.935	0.763
prohibitory	148	57	0.998	0.895	0.962	0.722
other	148	109	0.991	0.975	0.994	0.797

In addition to Table 1. The training results of the model are presented in Figure 5. Below.

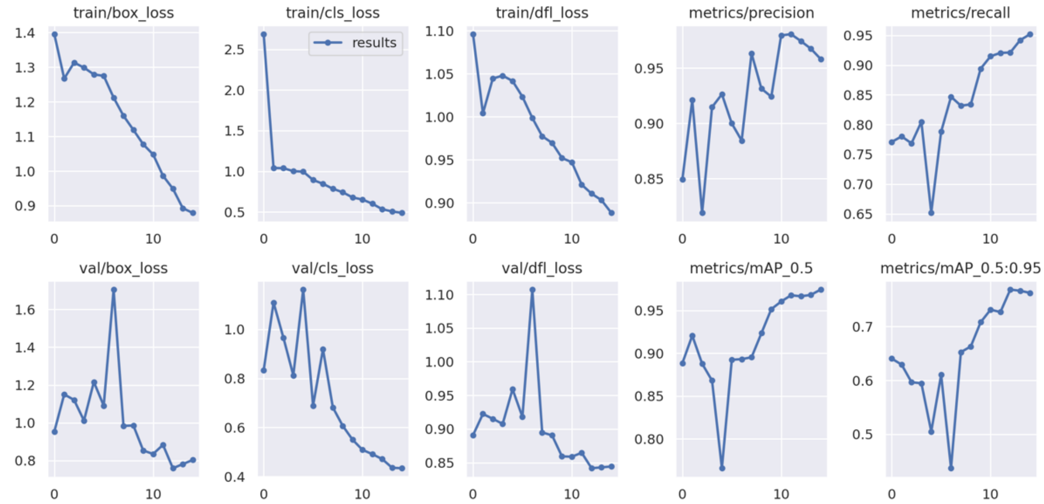

Figure 5. Model training results

From the graph of traffic sign detection training results using YOLOv9, the following conclusions can be drawn:

a. **Decrease in Loss**
In the train/box_loss, train/cls_loss, and train/dfl_loss sections, there is a significant downward trend. This indicates that as the epochs go by, the

model learns to predict the bounding box, classification, and distribution more accurately. For val/box_loss, val/cls_loss, and val/dfl_loss (loss on validation data), although there are initial fluctuations, eventually the loss also shows a significant downward trend. This indicates that the model is getting better at predicting on validation data, which is not used for direct training.

b. Improved Precision and Recall

In the metrics/precision and metrics/recall graphs, precision and recall increase as epochs increase. This shows that the model is getting more accurate in detecting traffic signs (precision) and is able to detect more correct signs from the total available signs (recall).

c. Improved mAP (Mean Average Precision)

Metrics/mAP_0.5 and metrics/mAP_0.5:0.95 both show an increasing trend, although they fluctuate at the beginning of the epoch. This indicates that the model as a whole is getting better at detecting and classifying traffic signs at various Intersection over Union (IoU) thresholds. mAP_0.5 increased close to 0.95, indicating excellent detection performance at an IoU threshold of 0.5. mAP_0.5:0.95, which measures performance at various IoU threshold values, also shows an increasing trend, albeit slower, with values approaching 0.75. The graph shows that the YOLOv9 model is getting better at detecting traffic signs after 15 epochs. The decreasing loss, increasing precision and recall, and increasing mAP show that the model is effective and ready to be used in real-life traffic sign detection applications.

2. Traffic sign detection results

The resulting model was tested in the real world environment presented in Figure 6.

Figure 6. Traffic Sign Detection Result

Based on Figure 6. the results of testing traffic sign detection using the model that has been made, it can be concluded as follows:

a. **Detection Result**
The YOLO v9 algorithm is able to detect several types of traffic signs with varying degrees of accuracy. Each sign detection is labeled with the sign type (e.g. prohibitory, danger, mandatory, other) and a confidence score ranging from 0.4 to 0.9.

b. **Categories of Detected Signs**
Prohibitory: Prohibitory signs (with confidence between 0.5 - 0.9) were detected in many images, demonstrating YOLO's ability to recognize prohibitory signs such as no entry or other prohibitions. Danger: Signs indicating danger (with confidence up to 0.9) were detected in several areas, demonstrating strong detection of situations indicating potential risk. Mandatory: Mandatory signs (with confidence around 0.4 to 0.9) were detected, although some detections showed lower confidence values. Other: The “other” category appeared with varying confidence, indicating other objects or signs that were not identified as standard signs, but were still recognized by the algorithm.

c. **Variations in Road Conditions**
The captured images show the detection of signs on diverse road conditions, ranging from urban areas, highways, to small roads. This shows the model's flexibility in detecting signs in various environments and lighting.

d. **Detection Accuracy**
While some detections show high confidence (up to 0.9), there are also some detections with low confidence (around 0.4 to 0.5). This suggests that in some cases, the model may need more data or training to improve accuracy on specific signs or different lighting conditions.

In conclusion, the results of the trained model show a fairly good performance in detecting traffic signs in various environments and conditions, but there are some detections that can be improved, especially in terms of confidence in some sign categories.

3. **Evaluation**
This model is evaluated with the Confusion Matrix presented in Figure 6 below.

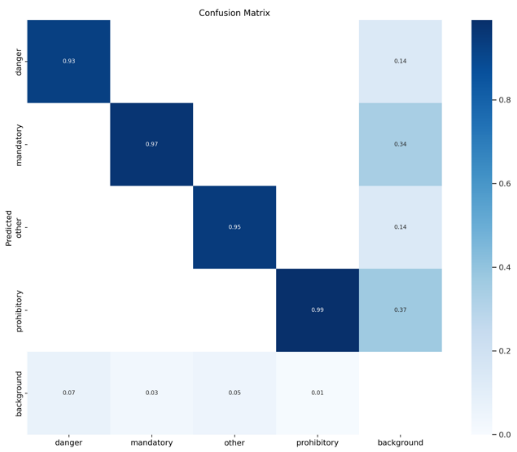

Figure 6. Confusion Matrix

Based on the confusion matrix image, the following conclusions can be drawn:

a. **Category Classification Accuracy**
   Danger: This category has a fairly good accuracy with a correct prediction value of 0.93, but there is a slight error with some data misclassified as background (0.07) and other categories. Mandatory: This category also has a high accuracy, with a correct prediction of 0.97, but there are a small number that are misclassified as prohibitory (0.34) and a few to background (0.03). Other: The other category had a correct prediction of 0.95, but there were a small number of mispredictions to prohibitory (0.14) and to background (0.05). Prohibitory: This is the category with the highest accuracy at 0.99, but there are still slight mispredictions to background (0.01) and some to other (0.37).
b. **Background category**
   Data that should have been background was misclassified as a sign category several times, although the percentage of errors was relatively small. For example, there were mispredictions from background to danger (0.07), mandatory (0.03), other (0.05), and prohibitory (0.01) categories.
c. **Classification Error**
   Prohibitory tends to experience errors more often than other categories. This can be seen from the wrong predictions to other categories such as other (0.37) and mandatory (0.34). Danger and other are more consistent in classification with minimal error.

Overall, the YOLO detection model performed quite well with high accuracy in classifying the categories of traffic signs. However, there are some significant errors especially in distinguishing between prohibitory signs and other categories, such as mandatory and other. Background categories are often incorrectly detected as signs, although the percentage is low. Further improvements to the model can be made by correcting the incorrect classifications, especially for the prohibitory and background cases.

CONCLUSION

The trained model successfully detected different types of traffic signs with a fairly high accuracy rate, especially for the prohibitory (0.99), mandatory (0.97), and danger (0.93) categories, demonstrating the effectiveness of the model in different environments. Although most of the signs were detected correctly, there were significant errors in distinguishing the prohibitory category from other categories, especially mandatory and other, which led to several misclassifications. In the test scenarios conducted in various road and environmental conditions, and the model showed consistent performance in recognizing signs in both highway and urban environments, proving the flexibility of the model. Although the overall results are good, improvements in the detection of specific categories such as prohibitory and reduction of misclassification to background would further improve the model's performance.

ACKNOWLEDGMENT

We would like to extend our heartfelt gratitude to the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, particularly the Directorate General of Vocational Education, for their support and funding of this research.

REFERENCES

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint*, arXiv:2004.10934. https://arxiv.org/abs/2004.10934

Duan, L., Zhao, Y., Liu, J., & Zhu, Z. (2020). Robust traffic sign recognition using multi-feature fusion and a CNN. *Journal of Visual Communication and Image Representation*, 73, 102956. https://doi.org/10.1016/j.jvcir.2020.102956

Khalid, S., Nawaz, S., & Ullah, Z. (2020). Automated traffic sign detection and classification system using deep learning. *Multimedia Tools and Applications*, 79(25), 17735-17750. https://doi.org/10.1007/s11042-020-08991-1

Larsson, E., Åström, K., & Heyden, A. (2019). Real-time traffic sign detection using an extended AdaBoost algorithm. *Computer Vision and Image Understanding*, 184, 58-67. https://doi.org/10.1016/j.cviu.2019.05.002

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788). https://doi.org/10.1109/CVPR.2016.91

Sermanet, P., & LeCun, Y. (2011). Traffic sign recognition with multi-scale convolutional networks. In *Proceedings of the 2011 International Joint Conference on Neural Networks* (pp. 2809-2813). IEEE. https://doi.org/10.1109/IJCNN.2011.6033589

Sichkar, V. (2020). Traffic signs dataset in YOLO format [Dataset]. *Benchmark*. http://benchmark.ini.rub.de/?section=gtsdb&subsection=dataset

Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv preprint*, arXiv:2402.13616. https://arxiv.org/abs/2402.13616

Zhou, W., Zhang, Y., Yang, F., & Wang, Q. (2019). Traffic sign recognition in the wild: A new benchmark. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), 1900-1912. https://doi.org/10.1109/TITS.2019.2928530