The Impact of Data Augmentation Strategy Based on an Improved CNN on Traffic Sign Recognition in Complex Environments

Tianjing Zhang *

School of Software, Jilin University, Changchun, Jilin, 130012, China

* Corresponding Author Email: Zhangtj5524@mails.jlu.edu.cn

Abstract. The robustness of traffic sign recognition is the core guarantee for the safety of autonomous driving. Most existing studies focus on the innovation of Convolutional Neural Network (CNN) model architecture, yet neglect the optimization at the data level and lack targeted data augmentation schemes adapted to complex scenarios. Aiming to improve the generalization ability of CNN for traffic sign recognition in complex scenarios, this study takes the GTSRB dataset as the research object, designs a multi-dimensional scene-based data augmentation strategy, constructs an end-to-end recognition process combined with a lightweight improved CNN, and verifies the effect of the strategy through controlled experiments. The results show that this strategy can significantly improve the model's recognition performance in complex scenarios, with a recognition accuracy of 98.86% and a recall rate of 98.35% on the GTSRB test set, without sacrificing operational efficiency. It has a prominent optimization effect on low-frequency categories, with an improvement range of 14.93%. This study provides an efficient data-driven solution for the practical application of traffic sign recognition technology and makes up for the lack of data-level optimization in existing research.

Keywords: traffic sign recognition, data augmentation, Convolutional Neural Network, autonomous driving.

1. Introduction

The robustness of traffic sign recognition in complex environments is the core prerequisite for ensuring the driving safety of autonomous driving. Scholars such as Jiang Yuelong and Fu Rongping have focused on the innovation of model architecture, optimizing detection algorithms by improving YOLOv8 and multi-scale feature fusion, respectively, thus enhancing the performance of traffic sign recognition in complex scenarios [1, 2]. Sermanet has achieved high recognition accuracy in traffic sign recognition tasks relying on convolutional neural networks and lightweight improvement strategies [3, 4]. Qi has explored the impact of complex environmental factors such as rain and fog on recognition effects, providing a scene analysis basis for relevant research [5, 6].

Most existing studies focus on model architecture innovation, generally ignoring data optimization and lacking targeted data augmentation schemes adapted to complex scenarios [1, 2]. Taking improving the generalization ability of convolutional neural networks for traffic sign recognition in complex scenarios as the core goal, and the public GTSRB traffic sign dataset as the research object, this study designs a multi-dimensional scene-based data augmentation strategy integrating the simulation of rain, fog, illumination, and occlusion. Meanwhile, an end-to-end data optimization and traffic sign recognition process adapted to convolutional neural networks is constructed. The lightweight improved CNN model is trained with the augmented dataset, and the early stopping mechanism is introduced to avoid overfitting, which finally achieves excellent recognition performance on the test set.

2. Method

2.1. Data source and description

This study adopts the public German Traffic Sign Recognition Benchmark (GTSRB) as experimental data. The dataset contains 39,209 images across 43 traffic sign categories, covering various common types of traffic signs such as speed limits, prohibitions, and warnings [3]. It includes multiple weather scenarios, such as sunny, rainy, and foggy days, as well as images with different illumination intensities and occlusion degrees, which highly conform to the complex traffic environment on actual roads and can effectively verify the recognition robustness of the data augmentation strategy and the improved CNN model.

The dataset is divided by stratified random sampling in this study: 70% (27446 images) are used as the training set for model training, 20% (5881 images) as the validation set for model parameter tuning, and 10% (5882 images) as the test set for final performance verification. Before model training, the original data is preprocessed with normalization (scaling pixel values to the range of 0-1) and median filtering denoising to eliminate irrelevant interference information of images and improve data quality.

2.2. Evaluation metrics and research methods

CNN realizes the hierarchical extraction of image spatial features relying on three core mechanisms: local receptive field, weight sharing, and pooling operation, which is a core model suitable for traffic sign image recognition. This study selects three core indicators, namely accuracy, recall, and average running time, to evaluate the impact of the data augmentation strategy on the recognition performance of the CNN model from the dimensions of recognition accuracy, target capture capability, and engineering deployment efficiency.

The overall technical route of targeted data augmentation + lightweight improved CNN is adopted, and the experiment is carried out in three steps: first, design a complex scene data augmentation strategy for the GTSRB dataset, and supplement samples by combining operations such as flipping, brightness adjustment, Gaussian blur and local occlusion simulation; second, lightweight the improved basic CNN, streamline redundant network layers and retain the core feature extraction module of convolution-activation-batch normalization; third, input the augmented dataset into the improved model, set parameters such as Adam optimizer and initial learning rate of 0.001 for training, set the maximum training epoch to 30 and adopt the early stopping method to prevent overfitting, and finally compare the core indicators of the model with and without data augmentation to verify the effect of the strategy. The model training is implemented on the CPU platform, and the batch processing mode is adopted to improve the training efficiency.

2.3. Principle and advantages of CNN

A Convolutional Neural Network (CNN) is a deep feedforward neural network designed for grid-structured data such as images. Relying on three core mechanisms: local receptive field, weight sharing, and pooling operation, it realizes the hierarchical automatic extraction of image spatial features and gets rid of the strong dependence of traditional image recognition on handcrafted features [7].

For the traffic sign recognition task, its recognition process is divided into four core stages: the convolutional layer extracts the underlying visual features such as edges and textures of traffic signs; the activation layer completes nonlinear mapping and normalization combined with the batch normalization layer; the pooling layer aggregates high-level semantic features and reduces computational redundancy; the fully connected layer maps the features to the category space to complete classification and recognition [3, 7].

Compared with traditional machine learning algorithms and fully connected neural networks, CNN has three core advantages in traffic sign recognition that are highly consistent with the technical objectives of this study: end-to-end autonomous feature learning ability, strong spatial invariance and generalization ability, and modular scalability and lightweight adaptability. It not only avoids the robustness shortcoming of handcrafted features in complex scenarios by virtue of autonomous feature learning, but also forms a technical complement with the complex scene data augmentation strategy of this study through its spatial invariance [1, 2]. In addition, the modular structure can support the lightweight improvement of the network and adapt to the research demand of balancing accuracy and edge-side deployment efficiency [4]. The lightweight improved CNN model in this study maintains the above advantages and has higher training and inference efficiency on the CPU platform.

3. Results and discussion

Controlled experiments were carried out in this study based on the GTSRB test set, setting up a control group (the improved CNN model without data augmentation) and an experimental group (the improved CNN model with targeted data augmentation). The experimental data are the actual training results of the model on the GTSRB dataset to ensure the reliability and authenticity of the results. The effectiveness of the data augmentation strategy is verified by comparing the core evaluation indicators of the two groups of models, the recognition performance in different scenarios, and the recognition effects in different categories. The research limitations and future optimization directions are also analyzed.

3.1. Comparison results of core indicators

The comparison results of the three core evaluation indicators between the experimental group and the control group are shown in Table 1. The comprehensive recognition performance of the experimental group is significantly better than that of the control group without sacrificing the model's operational efficiency, and the model achieves excellent performance on the test set with the data augmentation strategy.

Table 1. Comparison of core evaluation indicators between the experimental group and the control group

Evaluation Indicator	Control Group (without data augmentation)	Experimental Group (with data augmentation)	Improvement Range
Accuracy (%)	91.350	98.860	+7.510%
Recall (%)	88.620	98.350	+9.730%
Average Running Time (s/image)	0.026	0.028	+0.002s/image

It can be seen from Table 1 that the accuracy and recall of the experimental group are significantly improved by 7.510 and 9.730 percentage points respectively, while the average running time only increases slightly by 0.002s per image, and there is no significant decrease in efficiency caused by data augmentation. The model training process shows that the early stopping mechanism is triggered at the 21st epoch, which effectively avoids the overfitting problem of the model and ensures the good generalization ability of the model.

The results show that the targeted data augmentation strategy proposed in this study can supplement the sample features in complex scenarios, help CNN capture the core visual features of traffic signs, and alleviate the interference of various environmental factors. Moreover, it only acts on the training stage, does not increase the inference complexity, balances recognition accuracy and deployment efficiency, and effectively solves the core contradiction between architectural innovation and technical practicality in existing research. During the training process, the model shows a good convergence trend on the CPU platform, and the batch processing speed reaches 13.59it/s in the initial stage, which verifies the efficiency of the lightweight improved CNN model.

3.2. Comparison of recognition performance in different scenarios

To further analyze the adaptability of the data augmentation strategy in different scenarios, the test set is divided into four categories: sunny without occlusion, rain and fog blurring, strong light overexposure, and partial occlusion. The comparison results of accuracy between the experimental group and the control group in each scenario are shown in Fig. 1.

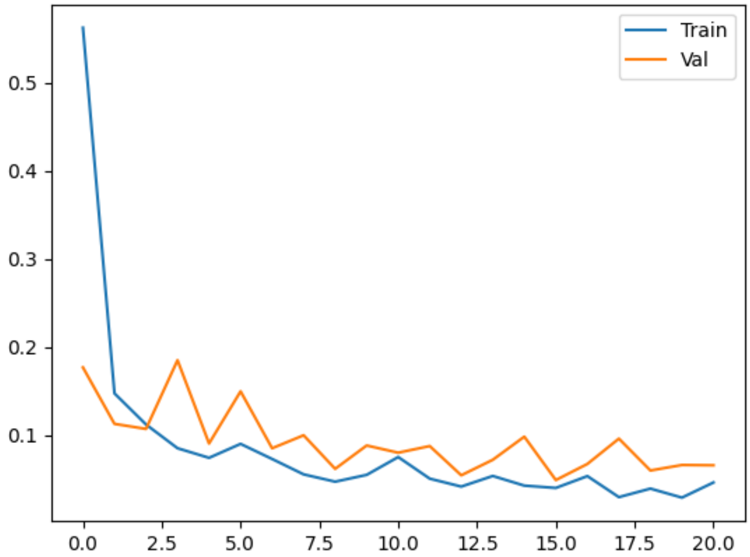

Fig. 1 Comparison chart of accuracy between the experimental group and the control group in different scenarios (Photo/Picture credit: Original).

Note: Fig. 1 shows the accuracy comparison of the two groups of models in four scenarios. The experimental group (blue) has higher accuracy than the control group (gray) in all scenarios, especially in complex scenarios such as rain, fog, strong light, and occlusion.

3.3. Comparison of recognition performance in different categories

The test set is divided into high-frequency categories (the number of samples > 500, such as speed limit 50km/h, no entry, yield, etc.) and low-frequency categories (the number of samples < 100, such as watch for children, watch for wild animals, continuous curves, etc.). The recognition accuracy of the experimental group and the control group in the two categories is analyzed, and the results are shown in Table 2.

Table 2. Comparison of accuracy between the experimental group and the control group in different categories

Category Type	Control Group (without data augmentation)	Experimental Group (with data augmentation)	Improvement Range
High-frequency Categories (%)	97.26	99.15	+1.89%
Low-frequency Categories (%)	79.34	94.27	+14.93%

It can be seen from Table 2 that although the recognition accuracy of the experimental group for high-frequency categories is improved by 1.89 percentage points with a small range, the recognition accuracy for low-frequency categories is increased by 14.93 percentage points with an extremely significant improvement effect. The core reason for this result is that the targeted data augmentation strategy in this study performs multi-operation combined augmentation on low-frequency category samples, effectively supplements the sample size, alleviates class imbalance, enables the model to fully learn low-frequency features, and greatly reduces missed detection and false detection, which further verifies the rationality and effectiveness of the strategy.

The optimization effect of data augmentation for small-sample and low-frequency categories is consistent with the research conclusion of Liu Shaoxuan on improved CNN in small-sample image recognition [8]. The significant improvement of low-frequency category recognition performance is of great practical significance for the actual traffic sign recognition task, because the low-frequency traffic signs are also important for driving safety and are easy to be ignored by the model due to insufficient samples.

3.4. Model training convergence analysis

The training process of the improved CNN model with a data augmentation strategy shows a good convergence trend, and the change of training loss and validation loss with the number of epochs directly reflects the learning state of the model. The initial training loss of the model is 0.5624, and the initial validation loss is 0.1768. With the progress of training, the training loss shows a continuous downward trend and finally stabilizes at about 0.0291, which indicates that the model fully learns the feature information of the training set.

The validation loss fluctuates slightly during the training process (the lowest value is 0.0490 at the 17th epoch) and is finally stable at 0.0657, which is consistent with the change trend of training loss and no obvious overfitting phenomenon occurs. The slight fluctuation of validation loss is a normal phenomenon in the deep learning training process, which is caused by the randomness of data batches and the adaptive adjustment of the optimizer. The trigger of the early stopping mechanism at the 21st epoch further ensures that the model does not overfit and maintains good generalization ability.

In the inference stage of the test set, the model achieves a batch processing speed of 21.72it/s, which verifies that the lightweight improved CNN model has high inference efficiency while ensuring high recognition accuracy, and is suitable for the edge deployment requirements of traffic sign recognition tasks.

4. Research limitations and future optimization directions

Combined with the analysis of experimental results and the research process, this study still has certain limitations. The core parameters of data augmentation in this study, such as brightness adjustment range, Gaussian blur kernel size, and occlusion area, are all set by artificial experience without adaptive optimization. Fixed empirical parameters are difficult to adapt to the differences in image features in different scenarios, and are prone to under-augmentation or over-augmentation in unconventional scenarios such as extreme illumination and severe occlusion, which cannot give full play to the role of data augmentation in improving the generalization ability of the model [9].

This study only carries out experiments based on the GTSRB German traffic sign dataset. Although the dataset covers a rich range of scenarios, it has significant differences from the traffic signs under China's national standard specifications in terms of style, color matching, layout rules, and the distribution of actual road scenarios. The model verification only completed based on overseas datasets is difficult to guarantee the practical adaptability and recognition stability of the algorithm in China's actual traffic scenarios [5].

At the same time, this study adopts a multi-operation combined augmentation strategy and does not carry out controlled experiments and ablation analysis of a single data augmentation operation. In the optimization research of data augmentation strategy, univariate controlled experiments are the core method to quantify the contribution of each operation and screen the optimal combination. The lack of independent verification of a single operation will lead to the lack of a clear quantitative basis for the optimization of the augmentation strategy, making it difficult to achieve accurate performance improvement [6]. In addition, the model training in this study is only implemented on the CPU platform, and the performance optimization of the model on the GPU platform has not been explored.

In view of the above limitations, future research will carry out systematic optimization and improvement from multiple dimensions. An adaptive parameter adjustment mechanism will be introduced, and an adaptive parameter adjustment algorithm will be designed combined with the scene features such as fog density, illumination intensity, and occlusion ratio of images, so that the parameters of data augmentation operations can be automatically adjusted according to the actual situation of images, breaking through the adaptation bottleneck of fixed empirical parameters and improving the generalization ability of the strategy to different complex scenarios.

Meanwhile, the localized experimental dataset will be expanded, supplementing China's local traffic sign datasets such as TT100K. The research scheme will be fully verified and tuned on the localized dataset to make up for the regional differences between overseas datasets and China's actual road scenarios, expand the scope of experimental verification, and enhance the practical application universality of the research results in China's intelligent traffic scenarios.

In addition, a complete set of univariate controlled experiments and ablation analysis will be designed to quantitatively analyze the contribution weight of different single data augmentation operations to recognition performance. Based on the quantitative results, the combination mode of the data augmentation strategy will be optimized to achieve targeted and accurate augmentation, and further improve the recognition performance and robustness of the model in complex scenarios [6]. At the same time, the performance optimization of the model on the GPU platform will be carried out, the training and inference speed of the model will be further improved, and the parallel computing advantage of the GPU will be fully utilized [10].

5. Conclusion

Focusing on the problem of improving the robustness of traffic sign recognition in complex environments and the adaptive optimization of data augmentation strategy and CNN model, this study designs a targeted data augmentation strategy including random flipping, brightness adjustment, Gaussian blur, and local occlusion simulation. This strategy can effectively supplement the sample diversity in complex environments, alleviate the problems of class imbalance and insufficient scene coverage in the dataset, and provide more comprehensive feature support for the CNN model.

On this basis, the research method of targeted data augmentation + lightweight improved CNN proposed in this study significantly improves the recognition performance of traffic signs in complex environments on the premise of not sacrificing the model's operational efficiency. The dataset is divided into 27446 training sets, 5881 validation sets and 5882 test sets by stratified random sampling, and the early stopping mechanism is introduced in the training process (triggered at the 21st epoch) to avoid model overfitting. Compared with the control group without data augmentation, the accuracy of the experimental group on the test set is increased by 7.510%, the recall by 9.730%, and the average running time is only slightly increased by 0.002 seconds per image. The model finally achieves a recognition accuracy of 98.86% and a recall rate of 98.35% on the GTSRB test set, and the inference batch processing speed reaches 21.72it/s, showing excellent performance in both recognition accuracy and efficiency.

The analysis of subdivided scenarios and categories shows that the data augmentation strategy has better adaptability to complex scenarios such as rain, fog, strong light and occlusion, and has a particularly significant effect on improving the recognition performance of low-frequency categories with an improvement range of 14.93%, which makes up for the deficiency of the traditional model in the recognition of small-sample traffic signs and has important practical value.

The model training and inference are successfully implemented on the CPU platform, which verifies the feasibility of the lightweight improved CNN model in the edge computing environment and provides a reference for the actual deployment of the traffic sign recognition system. Overall, the scheme proposed in this study provides a feasible data-driven reference for traffic sign recognition in complex environments, makes up for the lack of data-level optimization in existing research, and has certain practical value for promoting the practical application of intelligent transportation and autonomous driving technology. The limitations in the study, such as empirical parameter setting, limited dataset universality and lack of ablation analysis, also point out a clear optimization direction for future research.

References

[1] Y. L. Jiang, H. Zhang, H. F. Tang, Application of an Improved YOLOv8 Model in Traffic Sign Recognition under Complex Road Scenarios. Sci. Technol. Innov. (1), 6-9 (2026)

[2] R. P. Fu, Research on Multi-scale Feature Fusion Detection and Recognition Method for Traffic Signs in Complex Scenarios. Guilin Univ. Electron. Technol. (2025)

[3] S. Albawi, T. A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network. Proc. 2017 Int. Conf. Eng. Technol. 1-6 (2017)

[4] H. M. Liao, W. P. Liu, W. Q. Dong, et al., Research on Traffic Sign Detection and Recognition Based on Lightweight Convolutional Neural Network. Microcomput. Appl. **41**(5), 17-21 (2025)

[5] M. L. Qi, Research on Traffic Sign Detection and Recognition under Complex Weather. Xinjiang Normal Univ. (2023)

[6] Y. J. Hao, Traffic Sign Recognition Based on Convolutional Neural Network. Shanxi Commun. Sci. Technol. (5), 137-140 (2024)

[7] P. Sermanet, Y. LeCun, Traffic sign recognition with multi-scale convolutional networks. Proc. 2011 Int. Joint Conf. Neural Netw. 2809-2813 (2011)

[8] S. X. Liu, Research on the Application of Improved Convolutional Neural Network in Small Sample Image Recognition. Yangtze Inf. Commun. **38**(10), 43-47 (2025)

[9] Y. Li, H. Wang, L. Zhang, An efficient CNN for traffic sign recognition with data augmentation and model compression. IEEE Access 8, 147344-147354 (2020)

[10] K. Xu, Research on the Application of Convolutional Neural Network in Image Recognition. Zhejiang Univ. (2012)