Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/73/20240391

Enhancing road sign recognition using Convolutional Neural Networks

Yuning Zhuang

Nanjing Normal University Taizhou College, Taizhou, Jiangsu Province, China

chandlerv069@gmail.com

Abstract. The significance of road sign recognition technology has grown in tandem with the development of autonomous driving systems. However, traditional methods have encountered limitations when faced with intricate road scenarios. In contrast, CNNs have the capacity to autonomously glean complex features from image data, leading to a substantial enhancement in the stability and accuracy of road sign recognition. In this research, we employ a road sign dataset obtained from Kaggle and apply image preprocessing techniques, coupled with data augmentation methods. Furthermore, we employ transfer learning by utilizing ResNet34 to construct a robust road sign recognition model. The experimental outcomes vividly demonstrate the efficacy of this model in acquiring intricate features of road signs, resulting in the augmented accuracy and robustness in recognition. Nevertheless, we do observe certain limitations, including recognition instability and shortcomings in generalization. Nonetheless, on the whole, CNNs exhibit significant potential for application within the realm of road sign recognition.

Keywords: Convolutional Neural Network, Road Sign Recognition, Transfer Learning, Data Augmentation, Deep Learning.

1. Introduction

In this era of rapid technological advancement, autonomous driving technology is emerging as a prominent focus of research, leading the automotive industry towards comprehensive transformation. Beyond enhancing traffic safety, this technology holds disruptive potential for building a more efficient and environmentally friendly transportation system. Nevertheless, despite its promising prospects, achieving fully autonomous driving presents significant technical challenges, with road sign recognition standing out as a critical hurdle.

Road sign recognition plays a pivotal role in autonomous driving systems, serving not only as a provider of road information but also as the "eyes" of the driving system, offering essential guidance and decision support for vehicle navigation. However, while the concept of road sign recognition appears straightforward, its practical application is far more complex. Factors such as varying lighting conditions, adverse weather effects, and diverse viewpoints directly impact the clarity and recognizability of road signs. These complexities limit the effectiveness of traditional road sign recognition techniques, which rely on template matching and manual feature design, thereby affecting the reliability of autonomous driving systems.

Addressing these challenges necessitates the development of robust and adaptive road sign recognition techniques capable of overcoming multiple obstacles, including changes in lighting, adverse

© 2024 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: 10.54254/2755-2721/73/20240391

weather conditions, and diverse viewpoints. In this context, deep learning technology demonstrates remarkable potential. Through the construction of deep neural networks, these technologies can learn and extract intricate features of road signs from extensive data, facilitating accurate and stable road sign recognition in various complex environments. Therefore, the development of robust road sign recognition technology becomes imperative.

Deep learning, as a pivotal branch of artificial intelligence, has already exhibited substantial application potential across various domains, including road sign recognition. The distinguishing feature of deep learning lies in its exceptional data processing and feature extraction capabilities, offering significant advantages in addressing the challenges of road sign recognition. Deep learning models, particularly Convolutional Neural Networks (CNNs), possess the ability to automatically learn complex features from data through extensive training. This enables them to capture information pertaining to the shape, color, texture, and other aspects of road signs, resulting in more precise recognition. In comparison to traditional methods reliant on rules and manual feature design, deep learning excels in autonomously learning higher-level and more abstract feature representations, thereby enhancing recognition stability and accuracy.

2. Previous work

In the realm of road sign recognition technology, extensive prior research and various methodologies have been diligently explored. Preceding the widespread adoption of deep learning, researchers predominantly concentrated on traditional computer vision techniques, such as template matching and manually crafted feature extraction methods.

Template matching, a quintessential traditional computer vision approach, involves the prior preparation of a series of templates. These templates are then systematically compared with distinct regions of an image to identify the region that bears the closest resemblance to the template, thereby determining the road sign's position. However, template matching is plagued by several inherent limitations. It exhibits sensitivity to factors like scale and rotation variations of the target object, intricate backgrounds, and image noise. Consequently, template matching struggles to cope effectively with the intricacies and diversity of real-world scenes in practical applications.

Conversely, manually designed feature extraction methods were widely employed to distill features that best encapsulate relevant information from images. These features could encompass attributes such as color, edges, and shapes, aiding computers in discerning different road signs. Nonetheless, this approach suffered from a significant drawback: the selection and design of features relied on human expertise, rendering it a time-consuming and intricate process, particularly for capturing complex and abstract features. These challenges could potentially hamper the effectiveness of the method in practical applications. Furthermore, both template matching and manually designed feature extraction methods demonstrated a notable sensitivity to variations in lighting conditions, implying their potential to yield inaccurate results and fail to capture precise information in diverse environments.

With the advancement of machine learning techniques, several groundbreaking methods emerged, including Support Vector Machines (SVMs) and decision trees, among other machine learning algorithms. SVM, as a supervised learning algorithm employed for classification and regression tasks, operates by mapping low-dimensional data into a high-dimensional feature space and subsequently identifying an optimal hyperplane for classification within this space [2]. Decision trees, on the other hand, rely on tree-like structures and find application in classification and regression tasks. They partition data into various subsets through a series of decision rules, ultimately achieving data classification or prediction. While these methods achieved success in specific scenarios, they were still constrained by the feature design challenge, rendering them unsuitable for handling high-dimensional and intricate data.

Consequently, the introduction of Convolutional Neural Networks (CNNs) heralded a transformative era in road sign recognition technology. As specialized models tailored for processing images and visual data, CNNs autonomously acquire hierarchical features from raw image data through operations like layer stacking, convolution, pooling, and activation functions like ReLU [3]. This abstraction process

DOI: 10.54254/2755-2721/73/20240391

of hierarchical features endows CNNs with exceptional proficiency in image recognition tasks, thereby significantly enhancing the stability and accuracy of road sign recognition.

3. Dataset and Preprocessing

In this study, the road sign dataset obtained from Kaggle

(https://www.kaggle.com/datasets/andrewmvd/road-sign-detection) served as the foundational dataset. This dataset encompasses a total of 877 images spanning various categories, including traffic signals, stop signs, speed limit signs, and pedestrian signs. In the context of road sign recognition tasks, data preprocessing assumes a pivotal role, aiming to enhance the quality of raw data and thereby bolster the robustness and accuracy of models during both the training and evaluation phases. The preprocessing procedure encompasses several key steps, outlined as follows:

(1) Resizing: To ensure uniform input image dimensions that align with the model's requirements, all images in the dataset undergo resizing. Typically, images are resized to square dimensions or adjusted according to the model's specified input dimensions.

(2) Normalization: The pixel values within the images are subjected to normalization, a process that scales pixel values to a predefined range, often [0, 1]. This normalization not only expedites the model's convergence during training but also ensures that pixel values across different images maintain consistent scale ranges.

(3) Image Augmentation: Multiple image augmentation techniques are judiciously applied to diversify the dataset. These techniques encompass random rotation, horizontal flipping, and adjustments to brightness. By incorporating these augmentation strategies, the dataset is enriched with a broader spectrum of training samples. Consequently, the model's generalization capabilities are enhanced, enabling it to adapt more effectively to road sign images captured under varying scenarios.

Through the meticulous implementation of the aforementioned preprocessing steps, the dataset is thoughtfully prepared. This preparation process yields more resilient inputs for model training and evaluation, ultimately culminating in more dependable results within the realm of road sign recognition tasks.

4. Model

Transfer learning and data augmentation are fundamental strategies widely embraced in the realm of deep learning, particularly when confronted with the challenge of limited data. These strategies play pivotal roles in elevating the performance of models. Currently, many successful deep learning models heavily rely on vast quantities of labeled data. However, procuring an ample supply of labeled data for numerous learning tasks poses substantial difficulties, and the computational cost associated with training models from scratch can be prohibitively high. In scenarios characterized by such constraints, transfer learning emerges as an efficacious solution, enabling the acquisition of knowledge from simulated or related domains and its application in real-world contexts [4]. It is worth noting that when implementing transfer learning, a common challenge surfaces: the data distribution in the source domain may not align perfectly with that in the target domain. Despite sharing the same feature space, differences in marginal probability distributions exist. For example, training autonomous vehicle models using simulations is relevant, yet simulations inevitably fall short of replicating the intricacies of the real world. In such cases, transfer learning becomes particularly valuable, especially when confronted with data scarcity.

Moreover, data augmentation assumes a pivotal role in addressing a pervasive issue in deep learning, that of overfitting, while concurrently enhancing model robustness [5]. Overfitting is a prevalent problem in deep learning, often stemming from prolonged training durations driven by the scarcity of data. By employing data augmentation techniques—such as image flipping, rotation, or the introduction of noise into images—we can effectively alleviate the adverse effects of overfitting.

DOI: 10.54254/2755-2721/73/20240391

In the context of this model, both transfer learning and data augmentation strategies were judiciously employed. In summary, this study harnessed the power of the pre-trained ResNet34 model [6] and complemented it with data augmentation techniques applied to the input data. A comprehensive breakdown of the specific model structure is meticulously detailed in Table 1.

Table 1. Model core code

```
class model(nn.Module):
    def__init__(self):
        super(model, self).__init__()
        resnet = models.resnet34(pretrained=True)
        layers = list(resnet.children())[:8]
        self.features1 = nn.Sequential(*layers[:6])
        self.features2 = nn.Sequential(*layers[6:])
        self.classifier = nn.Sequential(nn.BatchNorm1d(512), nn.Linear(512, 4))
        self.bb = nn.Sequential(nn.BatchNorm1d(512), nn.Linear(512, 4))

    def forward(self, x):
        x = self.features1(x)
        x = self.features2(x)
        x = F.relu(x)
        x = nn.AdaptiveAvgPool2d((1,1))(x)
        x = x.view(x.shape[0], -1)
        return self.classifier(x), self.bb(x)
```

5. Results

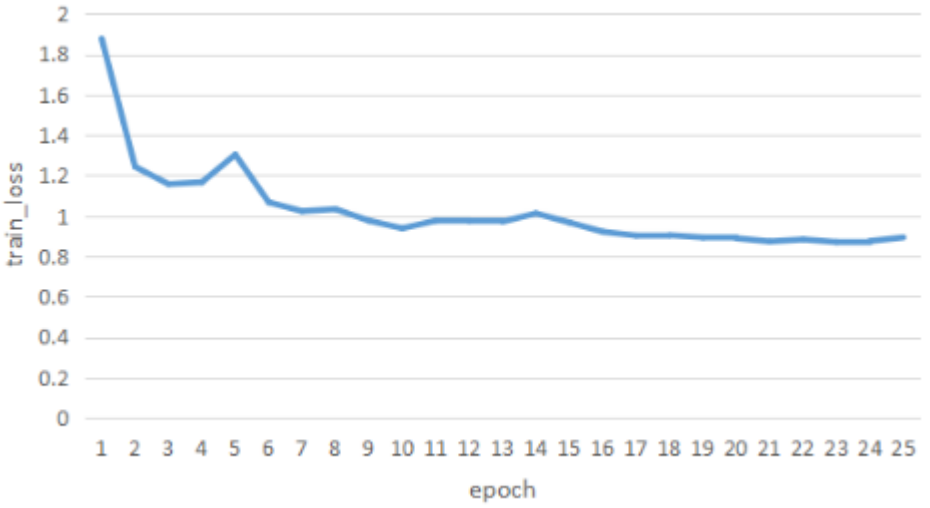

Figure 1. Loss value of CNN model on training set

DOI: 10.54254/2755-2721/73/20240391

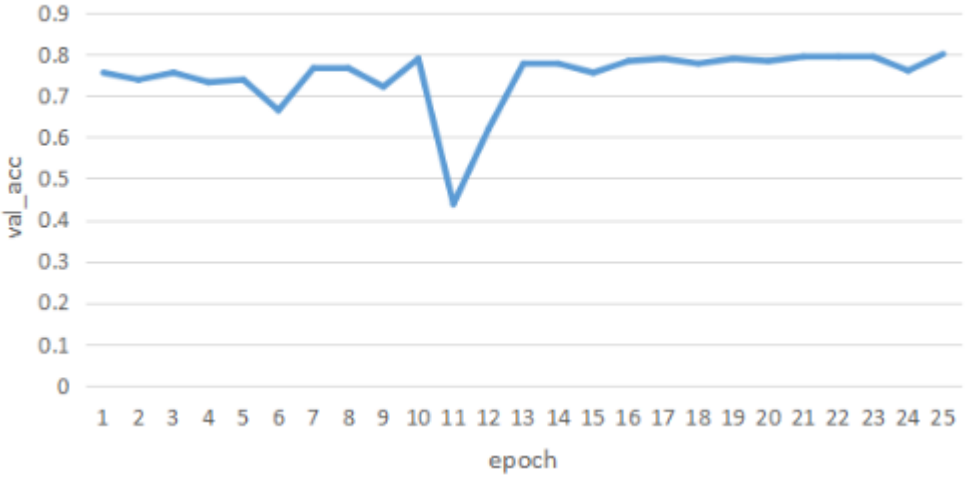

Figure 2. Accuracy of CNN model on validation set

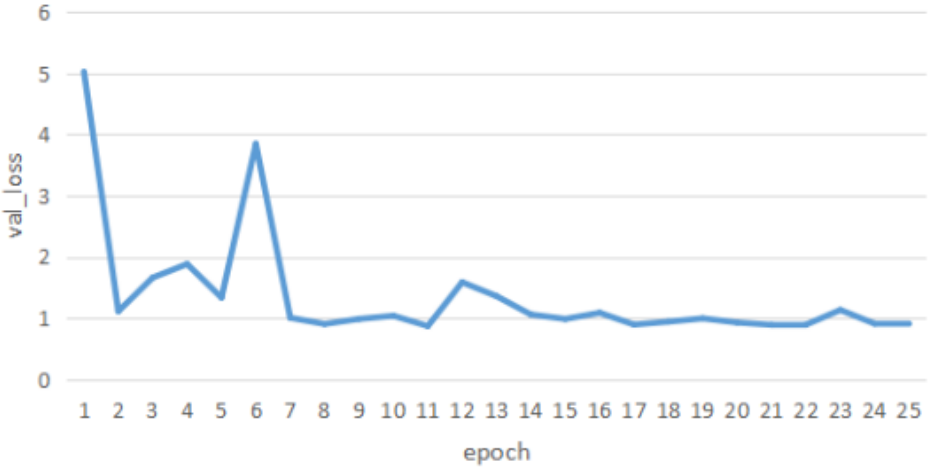

Figure 3. Loss value of CNN model on validation set



Figure 4. Road sign recognition application effect diagram

DOI: 10.54254/2755-2721/73/20240391

In summary, this model boasts a multitude of advantages. Firstly, within the domain of road sign recognition, it showcases an exceptional capacity to capture intricate features, enabling the effective differentiation of various road sign types, as exemplified in Figure 4. Furthermore, the model's support for transfer learning facilitates rapid model development and enhancement by allowing training on one task and applying acquired knowledge to related tasks. Moreover, the model's adaptability to diverse environmental and lighting conditions empowers vehicles to accurately identify road signs across a spectrum of scenarios, including daytime, rainy, nighttime, and snowy conditions. This adaptability significantly bolsters its robustness in practical applications. The strides made in road sign recognition through this model ultimately contribute to heightened driver safety, the reduction of traffic accidents, and the overall enhancement of road traffic safety.

Nevertheless, it is important to acknowledge that deep learning-based road sign recognition systems, while making substantial strides in augmenting recognition accuracy and robustness, do exhibit certain limitations. During the recognition process, the model may generate multiple recognition boxes, as illustrated in Figure 5

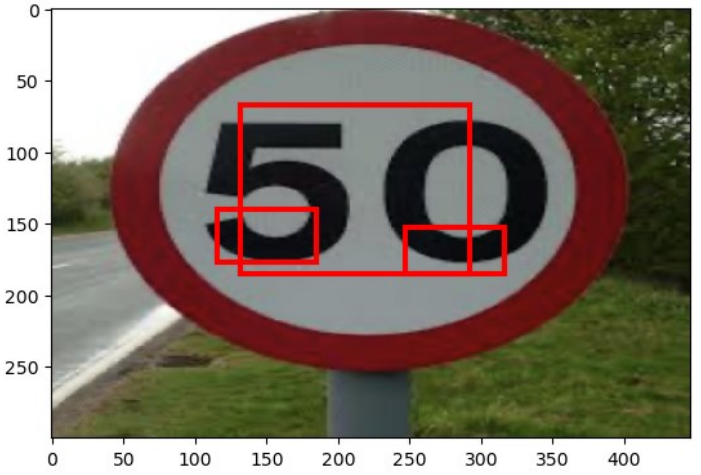

Figure 5. Effect of error recognition

This model typically demands a substantial volume of data for effective training. However, the limited availability of datasets for collecting and annotating extensive road images, along with the associated road sign information, poses constraints on acquiring a large-scale dataset. Simultaneously, there may be occurrences of missing road signs in specific regions. During the training process, the model can become biased by particular scenarios, resulting in limited generalization capabilities when confronted with diverse weather or road conditions. In the practical realm of autonomous driving, road sign recognition necessitates swift execution within milliseconds. Consequently, the algorithmic complexity of this model may impede its real-time performance. It is a well-known characteristic of deep learning models that their inference speed tends to be sluggish, thereby necessitating supplementary optimizations to meet real-time requirements.

6. Conclusion

Despite the significant strides made in the realm of deep learning-based road sign recognition systems, there persist several notable challenges that warrant both further exploration and comprehensive investigation. Among these challenges, recognition stability emerges as a primary concern. The constrained scale of available datasets poses a substantial hurdle, often hindering the accumulation of adequate training data. Consequently, models become excessively reliant on specific scenarios, leading to a notable deficiency in generalization capabilities. To fortify the model's robustness, it is imperative that forthcoming research endeavors prioritize the augmentation of dataset sizes and the diversification of scenarios, ensuring that the model can adapt effectively to a multitude of real-world conditions.

DOI: 10.54254/2755-2721/73/20240391

Another pressing issue revolves around real-time performance. Deep learning models inherently possess a high computational complexity, rendering the fulfillment of millisecondlevel real-time requirements in autonomous driving a formidable undertaking. Consequently, further algorithmic optimizations stand as an imperative necessity to strike a harmonious balance between recognition accuracy and the stringent constraints of real-time operation.

Moreover, the precise comprehension of intricate and multifaceted scenarios poses an additional challenge for deep learning models. Instances involving multiple road signs, for example, may lead to erroneous identifications by the model. To fortify the model's scene understanding capabilities, future research endeavors should delve into methodologies incorporating multi-source data inputs and leveraging prior knowledge. This holistic approach enables the establishment of a profound understanding of scene semantics, enabling the model to navigate complex scenarios with greater precision and reliability.

In summary, while deep learning has undoubtedly revolutionized road sign recognition, the construction of genuinely reliable systems necessitates the proactive addressal of key issues. Recognition stability, real-time performance, and scene understanding stand at the forefront of these challenges. Achieving this paramount objective mandates a continuous process of technological refinement and close collaboration across diverse disciplinary domains. The relentless pursuit of these advancements will undoubtedly pave the way for safer and more efficient autonomous driving systems, benefiting society as a whole.

References

[1] C.-Y. Fang, S.-W. Chen, and C.-S. Fuh, “Road-sign detection and tracking,” IEEE Trans. Veh. Technol., vol. 52, no. 5, pp. 1329–1341, Sep. 2003, doi: 10.1109/TVT.2003.810999

[2] V. K. Chauhan, K. Dahiya, and A. Sharma, “Problem formulations and solvers in linear SVM: a review,” *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 803–855, Aug. 2019, doi: 10.1007/s10462-018-9614-6

[3] D. Dai, “An Introduction of CNN: Models and Training on Neural Network Models,” in *2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR)*, Nov. 2021, pp. 135–138. doi: 10.1109/ICBAR55169.2021.00037

[4] C. Narvekar and M. Rao, “Flower classification using CNN and transfer learning in CNN-Agriculture Perspective,” in 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Dec. 2020, pp. 660–664. doi: 10.1109/ICISS49785.2020.9316030

[5] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, “A review of medical image data augmentation techniques for deep learning applications,” *J. Med. Imaging Radiat. Oncol.*, vol. 65, no. 5, pp. 545–563, 2021, doi: 10.1111/1754-9485.13261

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015, Accessed: Jun. 29, 2021. [Online]. Available: https://arxiv.org/abs/1512.03385v1