CNN-Based Traffic Sign Recognition

Shin Wee Fiona Liou, Hau-Lee Tong(✉), Kok-Why Ng, and Hu Ng

Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia
hltong@mmu.edu.my

Abstract. Traffic signs are a crucial part of maintaining driver and pedestrian safety on the road since they are being designed to provide essential information and alerts of potential hazards. With the rapid development of Advanced Driver Assistance Systems (ADAS), traffic sign recognition is also becoming much of a concern. However, due to real-world variations such as lighting conditions, occlusion, weather factors, motion blur and colour fading, there are still some failures in traffic sign recognition that cannot be perfectly resolved. Therefore, we implement image enhancement techniques and a pre-trained convolutional neural network for traffic sign recognition in this paper. Our proposed model uses the pre-trained VGG19 model as the baseline model and changes the fully connected layer and classifier of the VGG19 model. The experimental results demonstrate the effectiveness of applying image enhancement. Our proposed model was able to outperform the traditional machine learning method but did not surpass other deep learning methods.

Keywords: CNN · Traffic sign · Deep learning · Transfer learning · Image enhancement

1 Introduction

Traffic signs are crucial for every road user as they can be used to warn and guide road users. Moreover, traffic signs can help minimize road accidents and provide a safe traffic environment for road users. Traffic signs are usually designed to be easily noticeable which have a brighter colour and simple shapes. The first Traffic Sign Recognition System (TSRS) can be traced back to 1987, during that time, Akatsuka and Imai made an early TSRS [1]. As the development and interest in Advanced Driver Assistance Systems (ADAS) increased, the interest in traffic sign recognition also increased. Traffic sign recognition has been used in a variety of real-world applications, including traffic monitoring, transport system management, driver assistance, and traffic scene analysis. Even though traffic sign recognition has been used in many real-world applications, the development of powerful real-time TSRS still faces many challenges because of the real-world variability, including lighting conditions, scale variants, bad viewpoints, weather factors, motion-blur, and faded colour. Traffic sign recognition, in general, has included detection and recognition, but in this paper, we will only focus on traffic sign

S.-C. Haw and K. Sonai Muthu (Eds.): CITIC 2022, 10, pp. 195–204, 2022.
https://doi.org/10.2991/978-94-6463-094-7_16

classification. Detection is the process of extracting targets, whereas recognition is the process of classifying and identifying targets.

Recently, most of the image recognition research work has been done by using deep learning methods such as convolutional neural networks (CNN). The huge progress made in image classification over the past decade has inspired us to take a deeper look at the performance of CNN's pre-trained model for traffic sign recognition.

Classifying traffic signs accurately is important in Advanced Driver Assistance Systems (ADAS) because it helps to provide essential information, for instance, the route direction, traffic rules and the road condition. Therefore, we implemented the image processing and a CNN's pre-trained model to classify the German Traffic Sign Recognition Benchmark (GTSRB) dataset. In addition, we assessed the model's performance in traffic sign classification.

Next section will review the traffic sign dataset, existing traffic sign recognition techniques and transfer learning techniques. Section 3 will discuss our proposed method. Section 4 is our experimental results. Last section will conclude our work.

2 Literature Review

As the first suggestion for traffic sign recognition has passed more than 20 years, tremendous progress has been made in the traffic sign recognition field. It is difficult to evaluate different classification models because many researchers used different datasets and methods to recognise traffic signs. The detection and classification of traffic signs are the two main components of traffic sign recognition. Some of the researchers have combined detection and classification in their research, while others have only done classification. A real-world TSRS typically requires both detection and classification functions to recognise traffic signs, however, for this project, we will only be concerned with classifying the traffic sign images.

2.1 Traffic Sign Dataset

According to [2], although many studies on traffic sign recognition have been published, all of them are based on private data that is not available to the public. As a result, the author has introduced an open-source dataset, the GTSRB dataset, to address the lack of large and publicly available datasets that can be used for traffic sign recognition. After the presence of the GTSRB dataset, an increasing number of publicly available datasets which could be used for traffic sign recognition have been created. As shown in Fig. 1, in addition to the GTSRB dataset, there are some other publicly available traffic sign datasets, such as STS, KUL, RUG, Stereopolis, and LISA, which are all collected from different countries. Among all of these datasets, the GTSRB dataset is the most well-known because it was used for competition at IJCNN 2011, and it is the largest dataset available. The STS and KUL datasets are also commonly used for traffic sign recognition, as they both contain full images that can be used for detection. Apart from this, only the KUL and LISA datasets have included videos in their dataset.

	GTSRB	STS Dataset	KUL Dataset	RUG Dataset	Stereopolis	LISA Dataset
Number of classes:	43	7	100+	3	10	49
Number of annotations:	50000+	3488	13444	0	251	7855
Number of images:	50000+	20000	9006	48	847	6610
Annotated images:	All images	4000 images	All images	0	All images	All images
Sign sizes:	15x15 to 250x250 px	3x5 to 263x248 px	100x100 to 1628x1236 px	N/A	25x25 to 204x159 px	6x6 to 167x168 px
Image sizes:	15x15 to 250x250 px	1280x960 px	1628x1236 px	360x270 px	1920x1080 px	640x480 to 1024x522 px
Includes videos:	No	No	Yes, 4 tracks	No	No	Yes, for all annotations
Country of origin:	Germany	Sweden	Belgium	The Netherlands	France	United States
Extra info:	Images come in tracks with 30 different images of the same physical sign.	Signs marked visi-ble/blurred/occluded and whether they belong to the current road or a side road.	Includes traffic sign annotations, camera calibrations and poses.	Does not include any annotations, only raw pictures.		Images from various camera types.

Fig. 1. Traffic sign dataset that is available to the public [3]

2.2 Traffic Sign Recognition

Previous research has proposed various classifiers and algorithms to recognise traffic signs. The classifier is further categorized into two types which are traditional machine learning methods and deep learning methods. Traditional recognition methods include (but not limited to) support vector machines (SVM) [4, 5], random forest [5] and KD-tree [5] whereas deep learning methods include CNN [6–10, 16, 17], Capsule Network [11] and Extreme Learning Machine (ELM) [12].

Romdhane et al. [4] and Zaklouta and Stanciulescu [5], both used traditional method to perform traffic sign recognition and achieve high accuracy rate. Although traditional machine learning methods also can perform well in traffic sign recognition, as stated in the Qian et al. [14] manually engineered features take a long time and it is an error-prone process. In contrast, deep learning methods can learn features automatically from data, making a notable contrast to hand-crafted features. Thus, a variety of studies for traffic sign recognition and classification have been carried out by using deep learning.

Ciresan et al. [6] has proposed a committee of CNN and MLP methods to recognise GTSRB dataset and the accuracy score can be achieved to 99.15%. Besides that, Zeng et al. [7] has proposed the combined method to classify the traffic signs. In this paper, CNN was used to perform feature learning, and the Extreme Learning Machine (ELM) was used as a classifier. The result of the proposed method shows that it can correctly classify the GTSRB dataset with an accuracy of 99.4%. Aziz et al. [12] proposed a feature set combination method with an ELM classifier to recognise traffic sign images.The proposed method can provide 99.10% of recognition rate on the GTSRB dataset, while on the Belgium Traffic Sign Classification (BTSC) dataset it achieves a recognition rate of 98.33%.

In the paper of Kumar [11], the author proposes a capsule network to recognize the GTSRB dataset. The author claimed that CNN failed to obtain the position, orientation, and perspective of the images due to the limitations of the max-pooling layer. Therefore, they decided to use a capsule network to recognize images. The result of Caps Net shows that it can classify the traffic sign with a 97.62% accuracy rate. In addition, in Cao et al. [8], the traffic signs are recognised using an improved LeNet-5 model. The improved LeNet-5 model can achieve 99.75% accuracy in classifying the GTSRB dataset. Xu and Srivastava [9] have implemented a Histogram Equalization (HE) method to reprocess the traffic sign images, which can improve the information and brightness of the images. The convolutional neural network (CNN) was then used to automatically detect the traffic sign, and the complex structural information contained in the traffic sign image was obtained by using the hierarchical significance detection approach. Contrast limited adaptive histogram equalization (CLAHE) [15] is commonly used in image enhancement. It can prevent the over-amplification of the contrast that occurs in adaptive histogram equalization (AHE). In CLAHE, it takes the small area of the image to operate the contrast amplification, rather than using the whole image. After each title has done contrast amplification, they will be blended with the neighbourhood using bilinear interpolation so that the arbitrary boundaries can be removed. CLAHE image enhancement technique is employed in this paper to enhance the images quality.

2.3 Transfer Learning

Transfer learning can be referred to as a method of reusing previously trained models for a new but related task. The transfer learning method is extremely useful to solve real-world problems because real-world data usually does not have enough labelled data to train complex models. Besides that, transfer learning is also useful to speed up the training time because it will use the knowledge gained from a previous task to train the model. Some research works have proven that applying transfer learning to the CNN model can achieve high accuracy.

Zhou et al. [10] have developed an Improved VGG (IVGG) model to recognize traffic sign images. The main dataset taken for this research is the GTSRB dataset. Transfer learning and data augmentation have been applied by the authors to optimize the recognition rate of traffic sign images. The accuracy rate for the IVGG model is 99%. Even though the recognition rate for the IVGG model is up to 99%, but the authors have stated the inadequacy of the IVGG model. The IVGG model cannot perform well with the traffic sign images that are under darkness and motion blur background. This paper helps us understand the problem of overfitting as well as how to overcome it by using transfer learning and data augmentation.

Moreover, transfer learning method can also be seen in [13]. The transfer learning of the CNN models has been proposed to diagnose COVID-19 from Chest X-ray datasets. As stated by the author, the most difficult challenge is the classification of medical images because it is limited by the availability of annotated medical images. However, the transfer learning method can be used to solve this problem. In this paper, the classification work done by using the DeTrac and the VGG19 model can achieve the highest accuracy rate. We learned from this paper that even if we have a limited dataset, we can train and test our model using a deep learning-based transfer learning approach to get a good result.

In conclusion, the analysis of these previous works shows that the pre-trained convolutional neural network model can give reliable performance in both the medical and traffic sign recognition fields. Therefore, the CNN pre-trained model is implemented in our project.

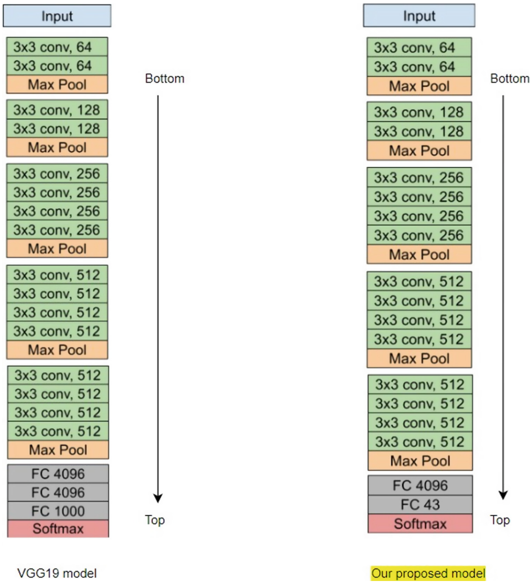

Fig. 2. VGG19 model and our proposed model.

3 Method

3.1 Model Structure

In this project, we implement transfer learning and modify the VGG19 model's architecture. There are 16 convolution layers, 5 max pooling layers, 3 fully connected layers, and 1 softmax classifier in the VGG19 model. However, instead of using two fully connected layers, in this project, we only use one fully connected layer and one softmax classifier in our output layer. Therefore, our proposed model in this project has 16 convolution layers, 5 max pooling layers, 1 fully connected layer, and 1 softmax classifier. The purpose of changing the fully connected layer is that we want the model to be lighter than its original so that it has faster convergence and reduced overfitting issue. The kernel size and activation function that is used in our proposed model are the same as those used in the VGG19 model, which is a 3×3 kernel size and ReLU activation function. Besides that, we will not make any changes to the convolution layer because this will result in the loss of the pre-trained weights learned in ImageNet. Figure 2 shows the VGG19 model and our proposed model.

Transfer learning has been applied by loading the pre- trained weight from ImageNet on the proposed model. Instead of training the whole model, transfer learning allows us to reuse the features that have been learned from the ImageNet. Therefore, we will only train the top layer of the model which includes the output layer and top two blocks of the convolution layers. The training process is shown in Fig. 3.

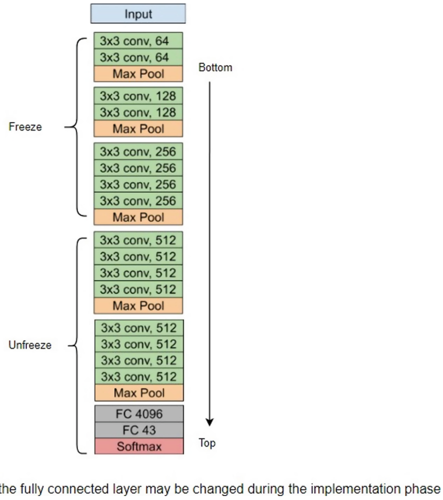

Fig. 3. The training process of the proposed model.

3.2 Loss Function

Since our task in this project is a multi-class classification task, multi-class entropy loss was selected as the loss function in this project. The equation of multi-class cross-entropy is shown as below:

$$\text{Multi} - \text{Class Entropy Loss} = \frac{1}{n}\sum_{k=1}^{n} y_k \log \hat{y}_k \tag{1}$$

in which n is the number of classes, $y\hat{}_k$ is the predicted value and y is the actual value.

4 Experimental Result

4.1 Dataset

The dataset that used in this project was the GTSRB dataset. The GTSRB dataset was created by Stallkamp et al. [2] and it is publicly available. Hence, we downloaded the GTSRB dataset from Kaggle. GTSRB dataset contains over 50 thousand images with 43 classes. The 43 classes can be divided into six subsets which are speed limit, prohibitory, derestriction, mandatory, danger, and unique.

4.2 Image Pre-processing and Enhancement

As the size of the traffic sign images in the GTSRB dataset varies, we resized all of them into 64 × 64 pixels. After resizing all the images into the same size, we implemented the image enhancement techniques, CLAHE to the GTSRB dataset with a clip limit of 0.1.

Table 1. Model accuracy on testing data

Model	Use Enhanced Images	Test Accuracy	Test Loss	Number of Error Predictions
1	No	0.973	0.253	345
2	Yes	0.975	0.126	319

4.3 Implementation Details

All of the models were trained on Adam optimizer and the batch size was 48. The models presented in this paper were implemented using Keras and ran on Jupyter Notebook.

4.4 Result of Testing

To observe the difference between the dataset that applied CLAHE and without CLAHE, we have trained two models which use different datasets. Model 1 was trained with the original images from the GTSRB dataset that did not apply CLAHE, whereas Model 2 was still trained with the GTSRB dataset but with the enhanced images that apply CLAHE. Table 1 shows the model accuracy on testing data. Based on Table 1, we can see that the test accuracy in Model 2 was slightly higher than in Model 1. Furthermore, the test loss and the number of false predictions of Model 2 were both significantly lower than in Model 1. It means that Model 2 can perform better than Model 1. Hence, we can conclude that using the enhanced images to train the model can improve the accuracy of the model on unseen data.

4.5 Evaluation Result

We compared our proposed model with other existing methods. There were eight existing methods, which we divided into three categories: traditional methods, deep learning methods, and transfer learning methods. The quantitative results of our proposed method and other existing methods were shown in Table 2. From Table 2, we can notice that our proposed model outperformed all traditional methods. However, when compared to other deep learning and transfer learning methods, our proposed model did not perform very well despite the difference in accuracy being small.

4.6 Result of Hyperparameter Tuning

To see if there was any hyperparameter that can improve the model accuracy, we performed hyperparameter tuning. The configuration and the results of hyperparameter tuning were shown in Table 3. The result of hyperparameter tuning from Table 3 had shown that the model accuracy did not increase when we tuned the hyperparameter. Therefore, we choose Model 2 as our final model.

Table 2. Quantitative comparison on GTSRB dataset

Method	Accuracy
Traditional Method	
KD-tree [5]	88.73%
Support Vector Machine (SVM) [5]	95.04%
Random Forest [5]	97.2%
Deep Learning Method	
CNN-MLP [6]	99.15%
CNN-ELM [7]	99.4%
Capsule Network [11]	99.1%
Improved LeNet-5 model [8]	**99.75%**
Transfer Learning Method	
Improved VGG (IVGG) model [10]	99%
Our proposed model	97.5%

Table 3. Hyperparameter tuning configuration

Run	Use Enhanced Images	Dropout Rate	Learning Rate	Epoch	Test Accuracy
1	Yes	0.3	1×10^5	30	0.970
2	Yes	0.3	1×10^5	40	0.964
3	Yes	0.3	5×10^6	30	0.956
4	Yes	0.3	5×10^6	40	0.961
5	Yes	0.5	1×10^5	30	0.966
6	Yes	0.5	1×10^5	40	0.967
7	Yes	0.5	5×10^6	30	0.955
8	Yes	0.5	5×10^6	40	0.962

5 Conclusion

In conclusion, we applied image enhancement techniques to the GTSRB dataset as well as transfer learning theory to our proposed model. Experimental result shows that our proposed model outperformed the traditional method such as Support Vector Machine, Random Forest and KD-tree but did not surpass other deep learning methods.

Authors' Contributions. Shin Wee Fiona Liou: Original draft preparation, Conceptualization.
Hau-Lee Tong: Supervision, Reviewing and Editing.
Kok-Why Ng: Supervision, Reviewing and Editing.
Hu Ng: Supervision, Reviewing and Editing.

References

1. H. Akatsuka and S. Imai, “Road signposts recognition system,” SAE transactions, 96(1), pp. 936-943, 1987.
2. J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “The german traffic sign recognition benchmark: a multi-class classification competition,” in The 2011 international joint conference on neural networks. IEEE, 2011, pp. 1453–1460.
3. A. Møgelmose, D. Liu, and M. M. Trivedi, “Detection of us traffic signs,” IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3116–3125, 2015.
4. N. B. Romdhane, H. Mliki, and M. Hammami, “An improved traffic signs recognition and tracking method for driver assistance system,” in 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). IEEE, 2016, pp. 1–6.
5. F. Zaklouta and B. Stanciulescu, “Real-time traffic-sign recognition using tree classifiers,” IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 4, pp. 1507–1514, 2012.
6. D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, “A committee of neural networks for traffic sign classification,” in The 2011 international joint conference on neural networks. IEEE, 2011, pp. 1918–1921.
7. Y. Zeng, X. Xu, Y. Fang, and K. Zhao, “Traffic sign recognition using deep convolutional networks and extreme learning machine,” in Inter- national Conference on Intelligent Science and Big Data Engineering. Springer, 2015, pp. 272–280.
8. J. Cao, C. Song, S. Peng, F. Xiao, and S. Song, “Improved traffic sign detection and recognition algorithm for intelligent vehicles,” Sensors, vol. 19, no. 18, p. 4021, 2019.
9. H. Xu and G. Srivastava, “Automatic recognition algorithm of traffic signs based on convolution neural network,” Multimedia Tools and Applications, vol. 79, no. 17, pp. 11 551–11 565, 2020.
10. S. Zhou, W. Liang, J. Li, and J.-U. Kim, “Improved vgg model for road traffic sign recognition,” Comput., Mater. Continua, vol. 57, no. 1, pp. 11–24, 2018.
11. A. D. Kumar, “Novel deep learning model for traffic sign detection using capsule networks,” arXiv preprint arXiv:1805.04424, 2018.
12. S. Aziz, F. Youssef et al., “Traffic sign recognition based on multi- feature fusion and elm classifier,” Procedia Computer Science, vol. 127, pp. 146–153, 2018.
13. A. Abbas, M. M. Abdelsamea, and M. M. Gaber, “Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network,” Applied Intelligence, vol. 51, no. 2, pp. 854–864, 2021.
14. R. Qian, B. Zhang, Y. Yue, Z. Wang and F. Coenen, “Robust chinese traffic sign detection and recognition with deep convolutional neural network,” In 2015 11th international conference on natural computation (icnc) pp. 791– 796, 2015.
15. S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, ..., K. Zuiderveld, “Adaptive histogram equalization and its variations,” Computer vision, graphics, and image processing, 39(3), pp. 355– 368, 1987.
16. J.S. Ang, K.W. Ng and F.F. Chua, “Modeling Time Series Data with Deep Learning: A Review, Analysis, Evaluation and Future Trend”, In 2020 8th International Conference on Information Technology and Multimedia (ICIMU) pp. 32–37. IEEE, 2020.
17. K. Ong, S.C. Haw, K.W. Ng, “Deep Learning Based-Recommendation System: An Overview on Models, Datasets, Evaluation Metrics, and Future Trends”, In Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems pp. 6–11, 2019.