Applied Mathematics, Modeling and Computer Simulation
C.-H. Chen et al. (Eds.)

doi:10.3233/ATDE221120

Malaysia Traffic Signs Classification and Recognition Using CNN Method

Yingxin CHEW [a], Zuliani ZULKOFFLI[a,1], Yihing CHU[b], Mohammad Arif ILYAS[b], M.K.A.Ahamed Khan[a], Chun Kit ANG[a] and Limwei HONG[a]

[a] *Department of Mechanical and Mechatronic Engineering, UCSI University Kuala Lumpur, Malaysia*

[b] *Department of Electrical and Electronic Engineering, UCSI University Kuala Lumpur, Malaysia*

Abstract. Automatic Traffic Sign Detection and Recognition (ATDR) system has been expanded and implemented partly in Intelligent Transportation System (ITS) that is actively used today. As traffic congestion increases, the manufacturing industry may have made and installed the ATDR systems in various types of vehicles, including cars, light commercial vehicles, and heavy trucks that act as driver assistance systems. ATDR system not only helps to minimize the number of traffic accidents but also supports road users through legal and compliant guidance and providing all traffic information, so that road users can be more attentive and solve them immediately in a short duration of time. On the point of the safety of the road users and others will be protected throughout a critical situation with identify the driving scene. In this paper, a deep-architectural neural networks, which is the convolutional neural network (CNN/ConvNet) model are chosen to expand in the ATDR system. With its excellent ability to train, research, and organize data, CNN is becoming one of the most widely used machine learning algorithms in key tasks such as prediction and classification. First, various open source of deep learning libraries will be studied. The library has been tested using the Malaysian Traffic Sign (MTS) dataset, which contains 32891 labelled images with real-world signs. Then the combination of both designed detection model and the CNN classification model will be trained under some parameter settings. Finally, the proposed system will detect and recognize the different types of traffic sign images in real-time display and Graphical User Interface (GUI). The achievable test accuracy is exceeded by 98% on a total of 43 different classes of MTS data that is obtained and for all study cases.

Keywords. Intelligent transportation system; Artificial intelligence; Computer vision; Road and traffic sign recognition, Convolutional neural network

1. Introduction

In the automotive business, machine learning algorithms play a very important role in developing Automatic Traffic Sign Detection and Recognition (ATDR) systems to improve driver safety and comfort. Deep learning (DL) is presently regarded by several automotive firms as jointly of the foremost promising new technologies, which might not solely improve the ATDR system, however, conjointly produce opportunities for

[1] Corresponding Author, Zuliani ZULKOFFLI, Department of Mechanical and Mechatronic Engineering, UCSI University Kuala Lumpur, Malaysia; E-mail: MohamedKhan@ucsiuniversity.edu.my.

self-driving smart cars [1], wherever the entire automotive industry has entered a replacement era in the future. In fact, DL has the power to improve and protect people's lives by providing extra diagnostic accuracy, as well as disaster assessment [2], drug discovery [3], and cancer diagnosis [4], in nearly every scientific field with the positive feedback of this technology. The world's leading technology and business-oriented companies are working hard to enhance Deep Learning (DL).

This research focuses on developing a new detection model with the combination of Convolutional Neural Network (CNN) classification model for the traffic sign detection and recognition purposes with a new Malaysia traffic sign dataset that has been prepared. Various advantages and the importance of implementing an ATDR system have been mentioned in the introduction section, as this requires the CNN model to be able to recognize the input images of a given traffic sign, recognize them and classify them into the correct category for real-time display in the test. Deep learning algorithms based on the CNN model are executed using deep learning libraries such as TensorFlow and its advanced API Keras library. Other open-source libraries, such as Pandas and Matplotlib, are also used to build CNN models. Methods such as the data augmentation is used by us to produce a new dataset for the Malaysia traffic signs and use the prepared dataset to train the proposed model that will be discussed in detail later.

However, there are some challenges that are remain for a successful ATDR system to apply in real world. These challenges are classified into two groups of hardware factor and environment factor. The example for hardware factors are lightning condition, motion blur, and vehicle speed. While for the environment factors are similar traffic sign inference as wrong sign, object looks similar to traffic sign, different weather condition (e.g., fog and rain), traffic sign scratch which causes by vandalism. These challenges could affect the model performance [5] [6].

Hence, to maximize the efficiency and accuracy for the detection and recognition of the model, different non-parametric statistical analyses was employed in this research to verify the performance improvement achieved by the proposed CNN model in solving the detection and classification problems. Moreover, the structural representation of the proposed CNN model with the convolutional layers, pooling layers, and the fully connected layers with the type of activation function being used in the model were also discussed. Finally, a test was conducted throughout the real-time detection of the traffic sign by showing its class ID name and the detection probability for each class. Moreover, a traffic sign detection and recognition GUI was also prepared for this project.

2. Related Works

ATDR systems are extremely significant in today's world, especially when it comes to autonomous driving. This section covers some of the vast research that has been done in the topic of traffic sign recognition and classification. On the one hand, a convolutional neural network and support vector machine (CNN-SVM) technique has been presented by Lai et al. [7]. The author feeds the input image into the CNN in Green (Y), Blue (Cb), Red (Cr) called YCbCr colour space to split the colour channels and extract several distinct characteristics, and then uses support vector machine (SVM) for classification with 98.6 % accuracy. Ruta et al. [8] developed a real-time traffic sign recognition system based on high-performance deep convolutional architecture. Teague suggested a color-based segmentation approach based on feature extraction using histogram-oriented gradient (HOG) and classification using SVM [9]. For the display of colours, the model employs CIECAM97, and it was used to segment and extract colour information. FOSTS [10] is another shape representation model that has achieved a 95% confidence accuracy rate. Srinivas et al. [11] implemented features extraction using HOG and local binary patterns (LBP) and then injecting them into a neural network with the maximum learning in classification and recognition system.

In addition, Huang et al. developed a Traffic Sign Recognition (TSR) system that extracts features from the histogram of the Oriented Gradient Variable (HOGv), then trains a single classifier using Extreme Learning Machine [12]. Apart from that, Zhu et al. created a new dataset with 100,000 photos and offered a robust end-to-end technique based on CNN model to detect and classify traffic signs, with an accuracy of 84 % [13]. Pei et al. suggested a multi-scale deconvolution network (MDN) that detects local traffic signs using a technique with a 99.1 % accuracy [14].

A research on the available approaches of applying CNN for TSR systems was provided by Hatolkar et al. [15], they also developed a problem-solving strategy for the time complexity and accuracy difficulties in CNN that have used hidden edge detection to extract the edges of traffic symbols before injecting them into CNN for classification. Moreover, a fuzzy classification approach is used by Pei et al. to help in classification and identification.

The author of [14] have also presented a TSRC system based on CNN scalability. The proposed system comprises of two CNNs where a fully convolutional network (FCM) was also employed. One is for the traffic sign location predictions; secondly is for classifying in each region. With the system achieving a 99.8 % accuracy rate, the goal of accomplish scale-invariant detection is succeed by the authors [16]. In terms of Malaysia Traffic Sign (MTS) datasets, the most well-known accessible dataset is released by Lim et al. [17]. Wali et al. [18] used a dataset from [17] to propose a TSDR system that included color segmentation and pattern matching, as well as a SVM classifier. The research showed an accuracy of 95.7% and a false positive rate of 0.9% when employing receiver operating characteristic (ROC) curve analysis.

3. Methodology

This research study involved of the following two stages: (1) detection and (2) classification and recognition. For the detection part, the research study involved of three folds; first, a new database for the Malaysia Traffic Signs prepared with each of the data image annotation files as for the training purposes for a detection model; a detection model called Nanodet was used to detect the Malaysia traffic signs; third, a real-time testing and images demo will be prepared. While for the classification and recognition part, the research study has involved three folds; first, a new development database for the Malaysia Traffic Signs; second, an architecture of a deep CNN will be designed and developed; third, a real-time testing will be prepared, and testing will be made after the models training is done. The MTS detection and recognition system architecture is depicted in figure 1. For the entire system, there are four stages: image capture, pre-processing, detection, and classification.

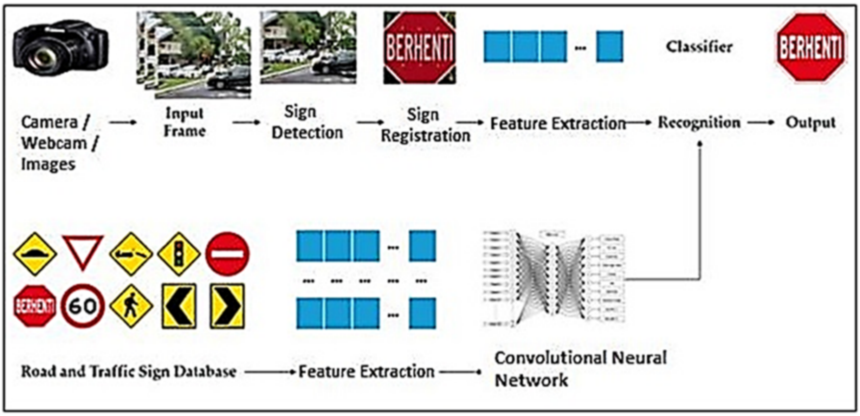

Figure 1. Malaysia Traffic Sign detection and Recognition System Architecture

3.1. Dataset Preparation

The dataset contains a total of 43 classes with 32891 images traffic sign in Malaysia, which acquired from [17]. The data was split into sets of training (60%), validation (20%), and testing (20%). The training sets will be used in the building model, while the validation sets are for evaluating purposes towards the performance of the model. This separation was created for cross-validation, which is a statistical method for evaluating the performance of machine learning (ML) algorithms.

3.1.1. Traffic Sign Detection

The presence or absence of a traffic sign is identified at this stage is using the Nanodet detection model [19], an open-source and ultra-fast lightweight real-time anchor-free target detection model, has debuted on GitHub. In this project, every region on interest (ROI) extracted in the previous step was allocated to one category; “trafficsign”, where the ROI contains the traffic sign, making the traffic sign detection model a unary classification problem rather than a multi-classification problem.

3.1.2. Traffic Sign Recognition

In general, the convolutional layer consists of the linear operation (convolution operation) and nonlinear operation (activation function) to perform the operation of feature extraction. Figure 2 shows block diagram of the entire classification system built for this study.

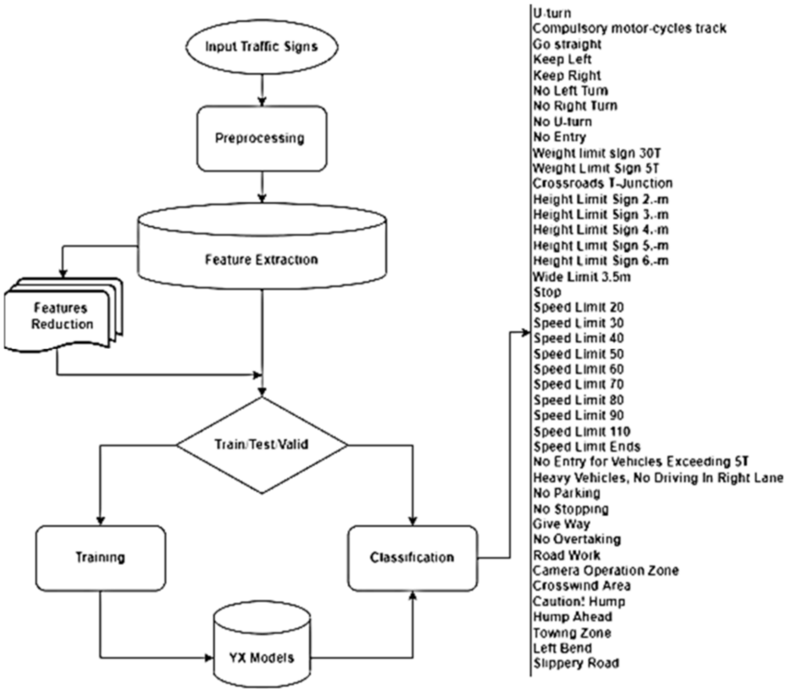

Figure 2. Block diagram of the system built for the classification model

Equation (1) was illustrated the results of the output of convolution for the next layer of one pixel is calculated. The output of the next layer is represented by net(i, j), and x is the input image, and w is the kernel or matrix of the filter and is the convolution operation.

$$\text{net}(\text{i, j}) = (x * w)[i, j] \quad = \sum_m \sum_n x[m, n] w[i - m, j - n] \tag{1}$$

The next step is the down sampling of each feature map within the subsampling layer. Pooling layers are similar to the convolutional layers, it reduces the size of the convolved feature or feature map to lower the use of the computational power through dimension reduction process in the data. The most favored and widely used pooling method is the maximum pooling.

Nonlinear activation functions are basically used after each trainable layer, such as the convolutional layers and the fully connected layers with weights in the CNN model. It will allow the implementation of error back-propagation for training a neural network, as Current Rectified Linear Unit (RELU) has become the most used activation function in the CNN [20] [21] [22] [23] [24].

Next, in a Fully Connected (FC) layer, each neuron within the FC layer will be connected to the output of the previous layer of feature maps. Before these feature maps connect to the next FC layers, a flattened technique is used to transform the dimensions from two-dimensional (2D) or three-dimensional (3D) into a one-dimensional (1D) numerical matrix. The classification scores with different results are generated for all the output classes by using the last layer of activation function [e.g., SoftMax]. Thus, the output of the FC layer is the final output for the CNN model.

The collected traffic sign is classified in the traffic sign recognition step. The image augmentation will be done on each class with the collected images, as until now there is no large Malaysia traffic signs dataset that has been prepared or collected. Then the input material which will serve to the proposed CNN model is the prepared Malaysia traffic sign dataset which contains 32891 images in total with 43 classes. Once it is done in training, the CNN model will be ready to use in the next classification of images with a whole new class of images which was prepared. Table 1 shows the summary of the proposed CNN model named as YX Model. This table shows the Input Layer, Convolutional Layer, Max Pooling, Dropout, Flatten and SoftMax parameter values.

4. Results and Discussions

In this paper, both training and testing were performed on a workstation running on an Intel Core i7-9750H 2.60 GHz processor, 16GB RAM, and a 512GB of Solid-State Drive (SSD). Our proposed model is shown in table 1, the name of the model named as YX Model. The results on comparing of used between the CPU and GPU with training and inferencing time with both models are showed in table 2 and table 3. The prototype was developed in the Visual Studio Code environment, with a GUI prepared. Figure 3 shows some input images of real-time images for classification model.

Table 1. Summary of The Proposed CNN Model (YX Model)

No.	Activation shape	Activation Size	Parameters
Input layer	(32,32,3)	3072	0
Convolutional layer 1 (f=5, s=1)	(28,28,60)	47040	1560
Convolutional layer 2 (f=5, s=1)	(24,24,60)	34560	90060
Max Pooling	(12,12,60)	8640	0
Convolutional layer 3 (f=3, s=1)	(10,10,30)	3000	16230
Convolutional layer 4 (f=3, s=1)	(8,8,30)	1920	8130
Max Pooling	(4,4,30)	480	0
Dropout (50%)	(4,4,30)	480	0
Flatten	(480,1)	480	0
FC layer 1	(500,1)	500	240500
Dropout (50%)	(500,1)	500	0
FC layer 2 (SoftMax) (Output layer)	(43,1)	43	21543

Table 2. Comparison between CPU and GPU used in detection model

Used of Hardware	Batch size	Total Epochs	Training Time	Real time Inference Time (Fps)
CPU	4	300	-	16
GPU	4	300	24 mins	13

Table 3. Result of CPU hardware and GPU hardware used for training and inferencing

	CPU		GPU	
	Training	Inference	Training	Inference
Experiments	1	2	1	2
Number of Test Samples	6571	6571	6571	6571
Number of Batch Size	3	50	3	50
Steps Per Epoch	10	400	10	400
Epochs Validation	5	100	5	100
Classification model Accuracy (%)	4.45	98.63	6.18	98.42
Training Time (mins)	5	160	1	11
Real time Inference Time (Fps)	-	23	-	25

Figure 3. Example of real-time input images for classification model

4.1. Traffic Sign Detection

Figure 4 shows some experimental input frames or the scene images and the corresponding output frames for the detection model. Four scenes were presented here for “BERHENTI” or Stop sign, Road Bumper sign, “60” 60 km/h speed sign and Roundabout within Stopping and Parking Forbidden signs were detected in the tested scene image.

Figure 4. Some experimental input frames and the corresponding output frames for the detection model

4.2. Traffic Sign Recognition and Robustness Testing

After the training which took around 11 minutes with using the GPU is done, the classification model has reached an accuracy of 98%. Robustness testing was implemented in determining a product's weaknesses in the presence of unexpected inputs or in a stressful environment. Some of the robustness testing for this proposed system was done using the real time view images disorientation angle images, faded signs images, natural blocked sign images where, the sign was blocked by trees or other structures, signs with motion blur images and night vision sign images.

4.3. Comparison of State-of the-art Models

There have been several earlier implementations of the traffic sign classification model. Despite this, the proposed CNN YX model has a higher accuracy (98%) than most of the other models proposed in the past, such as ResNet50, NASNetMobile, NASNetLarge, and others [25]. According to the results in table 4, all of the models perform rather well on the validation dataset. However, when compared to the proposed model in this article, the test dataset reveals which models effectively adapt to new images. With a score of 98.3% on the test dataset, the DenseNet169 pre-trained model is the top performer, however the accuracy attained is still lower than the YX Model provided in this study. Additionally, the parameters used by the YX Model is lower compared to all the models in [25], where less computational power and cost required.

Table 4. Comparison of the state-of-the-art models

	Model	**Validation Set (%)**	**Test set (%)**
References [25]	Xception	97.61	91.67
	VGG19	98.78	91.67
	VGG16	99.61	93.33
	ResNet50	20.00	21.67
	NASNetMobile	89.78	83.33
	NASNetLarge	90.67	76.67
	MobileNet	99.61	96.67
	InceptionV3	92.28	76.67
	InceptionResNetV2	96.56	90.00
	DenseNet121	99.28	91.67
	DenseNet169	98.83	98.33
	DenseNet201	99.11	95.00
	Vanilla CNN	100.00	78.33
Proposed Model	**YX model**	**98.10**	**98.41**

5. Conclusion

This paper carried out different non-parametric analyses to maximize the efficiency and accuracy on Malaysia Traffic Sign detection and classification. A model proposed to perform classification task which parameter settings used for the classification model achieved the high accuracy of 98%. It is observed that the proposed detection and classification model has demonstrated outstanding performances with the Malaysia traffic sign (MTS) database with excellent accuracy and overcome the overfitting problem. Finally, the developed model was successfully achieved real time road sign recognition performance, it enables Automatic Traffic Sign Detection and Recognition (ATDR) solution in real world.

References

[1] N. S. Yeshodara, N. S. Nagojappa and N. Kishore. Cloud Base Self Driving Cars. 2014 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp. 1-7, 2014.

[2] S. Nevo, V. Anisimov, G. Elidan, R. E. Yaniv, P. Giencke, Y. Gigi, A. Hassidim, Z. Moshe, M. Schlesinger, G. Shalev, A. Tirumali, A. Wiesel, O. Zlydenko and Y. Matias. ML for Flood Forecasting at Scale. arXiv preprint arXiv:1901.09583, 2019.
[3] H. m. Chen, O. Engkvist, Y. h. Wang, M. Olivecrona and T. Blaschke. The rise of deep learning in drug discovery. Drug discovery today, vol. 23, no. 6, pp. 1241-1250, 2018.
[4] Y. Benhammou, B. Achchab, F. Herrera and S. Tabik. BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. Neurocomputing, vol. 375, pp. 9-24, 2020.
[5] J. Guo, J. Y. Lu, Y. Y. Qu and C. H. Li. Traffic-Sign Spotting in the Wild via Deep Features. 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 120-128, 2018.
[6] T. M. Hoang, N. R. Baek, S. W. Cho, K. W. Kim and K. R. Park. Road Lane Detection Robust to Shadows Based on a Fuzzy System Using a Visible Light Camera Sensor. MDPI open access journal, vol. 17, no. 11, 2017.
[7] Y. Lai, N. X. Wang, Y. S. Yang and L. Lin. Traffic signs recognition and classification based on deep feature learning. 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp. 622-629, 2018.
[8] A. Ruta, Y. Li and X. Liu. Real-time traffic sign recognition from video by class-specific discriminative features. Pattern recognition, ScienceDirect, 2010.
[9] M. R. Teague. Image analysis via the general theory of moments. J. Opt Soc. Am, vol. 70, no. 8, pp. 920-930, 1980.
[10] A. Lee. Comparing Deep Neural Networks and Traditional Vision Algorithms in Mobile Robotics. 2016.
[11] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi and R. Venkatesh Babu. A Taxonomy of Deep Convolutional Neural Nets for Computer Vision. Frontiers in Robotics and AI, vol. 2, no. 36, 2016.
[12] Z. Y. Huang, Y. L. Yu, J. Gu and H. P. Liu. An Efficient Method for Traffic Sign Recognition Based on Extreme Learning Machine. IEEE Transactions on Cybernetics, vol. 47, no. 4, pp. 920-933, 2017.
[13] Z. Zhu, D. Liang, S. H. Zhang, X. L. Huang, B. L. Li and S. M. Hu. Traffic-Sign Detection and Classification in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2110-2118, 2016.
[14] S. W. Pei, F. W. Tang, Y. F. Ji, J. Fan and Z. Ning. Localized Traffic Sign Detection with Multi-scale Deconvolution Networks. arXiv preprint, 2018.
[15] Y. Hatolkar, P. Agrawal and S. Patil. A Survey on Road Traffic Sign Recognition System using Convolution Neural Network. International Journal of Current Engineering and Technology, vol. 9, no. 01, 2018.
[16] Y. C. Yang, S. Liu, W. Ma, Q. Y. Wang and Z. Liu. Efficient Traffic-Sign Recognition with Scale-aware CNN. arXiv preprint arXiv:1805.12289., 2018.
[17] K. H. Lim, K. P. Seng and L.-m. Ang. MIMO Lyapunov Theory-Based RBF Neural Classifier for Traffic Sign Recognition. Applied Computational Intelligence and Soft Computing, p. 7, 2012.
[18] S. B. Wali, M. A. Hannan, A. Hussain and S. A. Samad. An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation, Shape Matching, and SVM. Hindawi, 2015.
[19] Rangi Lyu. RangiLyu/nanodet: NanoDet-Plus Super fast and lightweight anchor-free object detection model. Only 980 KB(int8) / 1.8MB (fp16) and run 97FPS on cellphone. GitHub.
[20] Y. Le Cun, Y. Bengio and G. Hinton. Deep learning. Journal nature, vol. 521, pp. 436--444, 2015.
[21] A. Krizhevsky, I. Sutskever and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst 25, 2012.
[22] V. Nair and G. E. Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on Machine Learning, 2010.
[23] P. Ramachandran, B. Zoph and Q. V. Le. Searching For Activation Functions. arXiv, 2017.
[24] X. Glorot, A. Bordes and Y. Bengio. Deep Sparse Rectifier Neural Networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, vol. 15, pp. 315--323, 2011.
[25] D. N. T. How, Y. C. Hou, K. K. Sahari and O. Gumaan Saleh Basubeit. Recognizing Malaysia Traffic Signs with Pre-Trained Deep Convolutional Neural Networks. 2019 4th International Conference on Control, Robotics and Cybernetics (CRC), 2019.