Recognizing Malaysia Traffic Signs with Pre-Trained Deep Convolutional Neural Networks

Dickson Neoh Tze How
Department of Electrical & Electronics Engineering
College of Engineering
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: dickson@uniten.edu.my

Khairul Salleh Mohamed Sahari
Institute of Informatics and Computing in Energy
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: khairuls@uniten.edu.my

Yew Cheong Hou
Institute of Informatics and Computing in Energy
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: ychou@uniten.edu.my

Omar Gumaan Saleh Basubeit
Department of Electrical & Electronics Engineering
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: omarbasubeit@gmail.com

***Abstract*—An essential component in the race towards the self-driving car is automatic traffic sign recognition. The capability to automatically recognize road signs allow self-driving cars to make prompt decisions such as adhering to speed limits, stopping at traffic junctions and so forth. Traditionally, feature-based computer vision techniques were employed to recognize traffic signs. However, recent advancements in deep learning techniques have shown to outperform traditional color and shape based detection methods. Deep convolutional neural network (DCNN) is a class of deep learning method that is most commonly applied to vision-related tasks such as traffic sign recognition. For DCNN to work well, it is imperative that the algorithm is given a vast amount of training data. However, due to the scarcity of a curated dataset of the Malaysian traffic signs, training DCNN to perform well can be very challenging. In this demonstrate that DCNN can be trained with little training data with excellent accuracy by using transfer learning. We retrain various pre-trained DCNN from other image recognition tasks by fine-tuning only the top layers on our dataset. Experiment results confirm that by using as little as 100 image samples for 5 different classes, we are able to classify hitherto traffic signs with above 90% accuracy for most pre-trained models and 98.33% for the DenseNet169 pre-trained model.**

Keywords–deep learning; convolutional neural network; object recognition; Malaysia traffic sign

I. INTRODUCTION

In the pursuit of an intelligent transportation system, traffic sign detection and recognition plays an important role to ensure the safety of the pedestrians and vehicles on the road [1]. Automated traffic sign detection and recognition enable vehicles to operate abiding existing traffic rules and regulation with minimal human intervention. Since traffic signs are designed to be visible for human eyes, they are placed at obvious and strategic locations along driveways. It is not surprising that many intelligent vehicle systems are designed to detect and recognize these traffic signs from image streams [2]. Nevertheless, detecting and recognizing the traffic signs remains a challenging task due to various complexities such as languages and environments since the learning-based methods usually applied to the predefined set of traffic signs. The emergence of the deep learning field brings forth many applications in object detection and recognition. The contributing factor to the influence of deep learning in many fields can be attributed to two main factors i.e. the drastic reduction in computational cost and the availability of massive data [3]. As a result, many state-of-art algorithms for object recognition [4], speech translation [5], image captioning [6], natural language processing [7] and so forth involve the extensive use of deep learning.

Despite the massive successes of deep learning in other domains, training a deep learning algorithm to recognize Malaysia traffics sign can prove to be challenging. The main setback can be attributed to the scarcity of a curated dataset consisting the Malaysia traffic sign. The Malaysia traffic signs are similar to that is used in Europe with a few distinctions as shown in Figure 1. For example, some traffic signs incorporate Malay words as part of the road sign. This feature makes Malaysia traffic signs unique to the country and is not used elsewhere. At the moment, rich and properly curated datasets on the Malaysian traffic signs are still scarce. At the point of this writing, available Malaysian traffic signs dataset includes only small patches (32×32 pixels) of traffic sign images which may be of limited use especially in training very deep neural networks [8]. While small image patches can be useful in speeding up the training of deep learning networks, the down-sampled images may have omitted some features visible only in higher resolution images. Furthermore, it is reported by Wu et al. that a larger

978-1-7281-4620-1/19/$31.00 ©2019 IEEE
DOI 10.1109/CRC.2019.00030

resolution image improves recognition accuracy [9]. Also, since each subsequent convolution operation reduces the image size, to train a deeper network for improved performance, a larger resolution image is typically required.

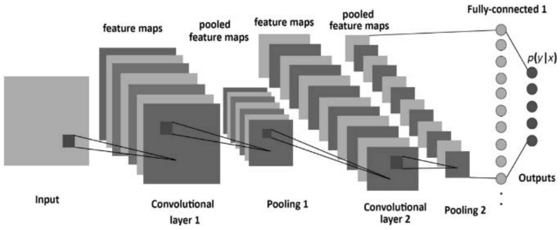


Figure 1. Convolutional Neural Network architecture [23].

In this study, we demonstrate that by leveraging on the transfer learning training scheme, we are able to train deep convolutional neural networks (CNN) with relatively little data. The contributions of this study are as follows:

- A dataset of Malaysia traffic signs consisting of 5 classes with 100 samples per class with a minimum image resolution of 200×200 pixels.
- Evaluation of classification accuracy on various pre-trained models finetuned on our dataset.

II. Related Work

This section highlights some of the related works pertaining to traffic sign recognition systems. A bulk of recent works on traffic sign recognition involve deep learning. Quan et al. [10] proposed a real-time traffic sign recognition using efficient deep convolution-based architecture. The authors reported 98.6% accuracy rate on the German traffic sign dataset. Jain et al. [11] utilized a genetically optimized CNN to detect traffic signs by training on the Belgium and Chinese traffic sign dataset. The authors claimed to have outperformed all existing approaches and provided a new benchmark of 99.16% (Belgium traffic sign dataset) and 96.28% (Chinese traffic sign dataset).

With regards to the Malaysian traffic sign dataset, the most prominent available dataset is published by Lim et al. [8]. Apart from that, Madani et al. [12] [13] proposed an alternative Malaysian traffic sign dataset that includes a variety of scenes in order to solve the gap for the dataset used in Malaysia traffic sign recognition systems. Based on these datasets, a number of works have been carried out. Lau et al. [14] discussed a comparative study of the Malaysian traffic sign recognition rate between deep CNNs and radial basis function neural network (RBFNN) trained on the mean squared error loss function. The authors showed that deep architecture of CNN outperforms its shallow counterparts. The authors have utilized dataset from [8] Wali et al. [15] proposed a traffic sign detection and recognition system with dataset from [8] by using color segmentation and shape matching in conjunction with a support vector machine (SVM) classifier. The study reported accuracy of 95.71% and false positive rate 0.9% by using receiver operating characteristic (ROC) curve analysis. Islam et al. [16] [17] proposed an artificial neural network (ANN) based classifier with feature engineering to recognize Malaysia traffic signs with an accuracy of 99.9% on the Malaysia traffic sign database [8].

III. Deep Convolutional Neural Network

A. Network Architecture

CNN has become the dominant machine learning technique for visual object recognition [18]. More recently, CNN has seen many innovations from its initial implementation over 20 years ago. Starting from the 5 layers deep LeNet5 [19] in 1998, 8 layers deep AlexNet [20] in 2012, 26 layers deep VGG19 [21] in 2014 and more recently in 2015, ResNets [4] by Microsoft featured over 100 layers in network depth. At the point of this writing, there exist many variations in CNN used for object classification. Table I tabulates some of the CNN variations trained on various object classification dataset.

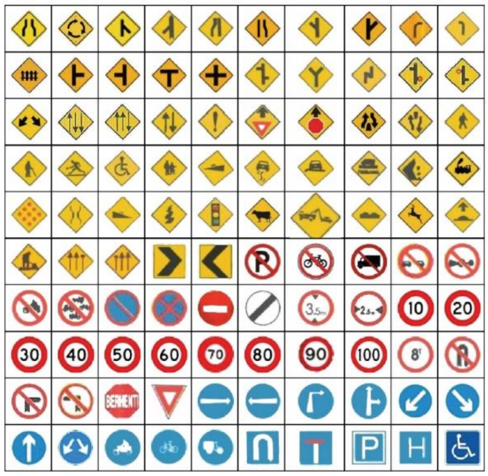


Figure 2. Malaysian traffic signs [8].

The most commonly used layer arrangements in CNN is to place pooling layers after each convolutional layers. These patterns can be observed in many variants such as the LeNet, AlexNet and the VGG19. Figure 2 illustrates the arrangements of the alternating layers. However, as discovered in [22], deeper networks suffer from a condition known as the vanishing gradient where the information passed from the output layers begin to diminish as it reaches the beginning layers of the network. In 2015, skip connections in ResNets were introduced to mitigate the vanishing gradient problem. These connection allow information to flow from the later layers to the earlier layers of the network which makes it possible to construct networks with more than 100 layers in depth. Applying the same idea of skip connections, DenseNet also utilized skip connections which bypasses certain layers to allow direct connection from one layer to every other layer preceding it. Figure 3 illustrates a DenseNet block. DenseNet has been proven to outperform ResNets on various datasets with a fraction of the number of parameters [18].

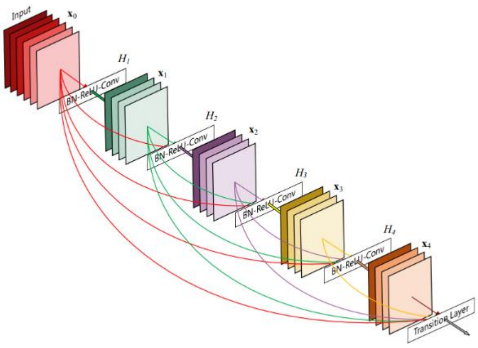


Figure 3. DenseNet architecture [18].

B. *Transfer Learning with pre-Trained Models*

Many deep learning models exhibit a common phenomenon: the features learned by the lower layers of the networks are somewhat similar even though they are trained on different datasets [24]. This opens up the possibility of reusing the features learned by the lower layers on one dataset and transferring the features to perform on other datasets. This method is known as transfer learning. Yosinski et al. demonstrated that transfer learning leads to improved generalization performance on datasets [24].

Given the limited traffic sign images in our dataset, we are not able to train very deep CNN models due to overfitting concerns [25]. Hence, we leverage on the availability of pre-trained models on other datasets and fine-tune by retraining the top densely connected layers and the output layer. All models are pre-trained on the Imagenet dataset with 1000 classes [26]. Table I tabulates the pre-trained models that are used in this study. Model refers to the name of the pre-trained model, Parameters refers to the number of trainable weights, Depth refers to the number of computational layers of the network.

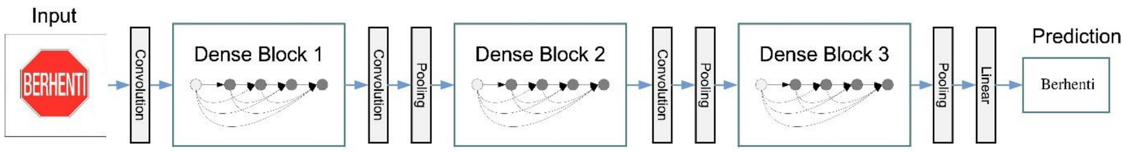


Figure 4. DenseNet model architecture for traffic sign classfication.

IV. Experimental Results

A. *Dataset*

Due to the constraint of having low-resolution images in [12] and [8], we opt to sample a higher resolution image dataset of the Malaysian traffic signs. The dataset consists of 5 different classes with a total of 100 images per class. The resolution of each image is at least 200×200 pixels. Listed below are the classes of traffic signs as part of the dataset:

- Berhenti
- Beri Laluan
- Jalan Sehala
- Lampu Isyarat
- Dilarang Masuk

Figure 5 illustrates the sample images for each class. Images were taken from various angles and lighting conditions.

B. *Training*

We utilized the transfer learning training scheme to train all of the listed pre-trained models in Table I by finetuning only the top layers for each model. Dataset is divided into 80% training, 10% validation and 10% test set. Training images were augmented by performing random rotations, translations, shearing, and zooming. All models were trained using the cross-entropy loss function (shown in Eq. 1) with Adadelta optimizer [27]. All dense layers in the models utilize the Rectified Linear Unit (ReLU) activation function (shown in Eq. 2). Each model is trained for a maximum of 100 epochs. In order to combat overfitting, the Early Stopping training technique is adopted [28]. Training is halted when the validation accuracy does not improve for 10 consecutive epochs.

$$L = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \quad (1)$$

where M is the number of classes, y is a binary indicator (0 or 1) if class label c is the correct classification for observation o, p is the predicted probability observation o is of class c.

$$R(z) = \max(0, z) \quad (2)$$

TABLE I. Different Variations of Deep Convolutional Neural Networks Used as pre-Trained Models

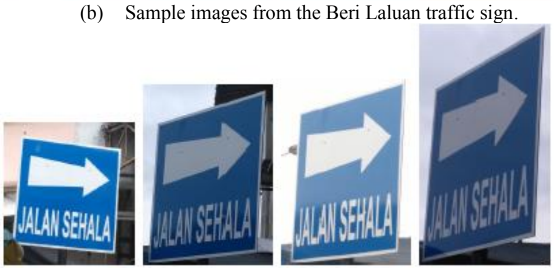
Model	*Total Parameters*	*Depth*
LeNet	60,000	5
AlexNet	62,378,344	8
VGG16	138,357,544	23
VGG19	143,667,240	26
ResNet50	25,636,712	168
NASNetMobile	5,326,716	*
NASNetLarge	88,949,818	*
MobileNet	4,253,864	88
InceptionV3	23,851,784	159
InceptionResNetV2	55,873,736	572
DenseNet121	8,062,504	121
DenseNet169	14,307,880	169
DenseNet201	20,242,984	201

V. RESULT AND DISCUSSION

We evaluated the accuracy of the models against the training, validation set and also the test set. The results are summarized in Table II. Included in the table is a 5-layer CNN trained from scratch (Vanilla CNN) as a benchmark. Since the dataset used is balanced, we only evaluate the performance of the model on the accuracy metric.

(a) Sample images from the Berhenti traffic sign.

(b) Sample images from the Beri Laluan traffic sign.



(c) Sample images from the Jalan Sehala traffic sign.

(d) Sample images from the Lampu Isyarat traffic sign.

(e) Sample images from the Dilarang Isyarat traffic sign.

Figure 5. Sample images from the collected dataset.

Observation of Table II indicates that all models score relatively well on the training dataset and the validation dataset. However, evaluation of the test dataset highlights the models that can generalize well to novel images. The best performer on the test dataset is the DenseNet169 pre-trained model with 98.33% accuracy score. The worst performer is the ResNet50 model with only 21.67% correctly classified images despite scoring extremely well on the training dataset. The poor performance on the test dataset can be attributed to overfitting of the model to the training dataset. Figure 6 shows the confusion matrix and Table III highlights other classification metrics.

TABLE II. PERCENTAGE OF CORRECTLY CLASSIFIED SAMPLES IN THE DATASET

Model	***Train Set (%)***	***Validation Set (%)***	***Test Set (%)***
Xception	0.9966	0.9761	0.9167
VGG19	0.9893	0.9878	0.9167
VGG16	0.9947	0.9961	0.9333
ResNet50	0.9970	0.2000	0.2167
NASNetMobile	0.9931	0.8978	0.8333
NASNetLarge	0.9970	0.9067	0.7667
MobileNet	0.9986	0.9961	0.9667
InceptionV3	0.9766	0.9228	0.7667
InceptionResNetV2	0.9928	0.9656	0.9000
DenseNet121	0.9954	0.9928	0.9167
DenseNet169	0.9972	0.9883	0.9833
DenseNet201	0.9970	0.9911	0.9500
Vanilla CNN	0.9972	1.0000	0.7833

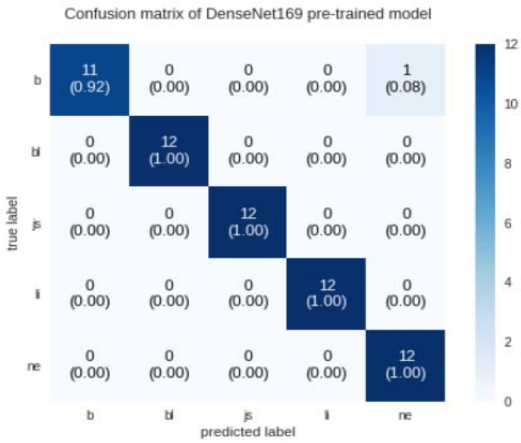


Figure 6. Confusion matrix on the test set.

TABLE III. TEST SET CLASSIFICATION RESULTS ON THE DENSENET169 PRE-TRAINED MODEL

Class	***Traffic***	***Samples***	***Correct***	***Incorrect***	***Precision***	***Truth***	***Accuracy***	***Recall***	***F1-Score***
0	Berhenti	12	11	1	92	11	98.33	1.0	0.96

Class	*Traffic*	*Samples*	*Correct*	*Incorrect*	*Precision*	*Truth*	*Accuracy*	*Recall*	*F1-Score*
1	Beri Laluan	12	12	0	100	12	100	1.0	1.0
2	Jalan Sehala	12	12	0	100	12	100	1.0	1.0
3	Lampu Isyarat	12	12	0	100	12	100	1.0	1.0
4	Dilarang Masuk	12	12	0	100	12	100	1.0	1.0

VI. CONCLUSION

In this study, we have presented preliminary results Malaysian traffic sign recognition using transfer learning from pre-trained deep convolutional neural networks. Transfer learning has enabled us to utilize many state-of-the-art deep CNN models that are otherwise unfeasible to train with the amount of data that we possess. Not only the models can be trained very quickly, the models also perform reasonably well on the proposed dataset with no hyperparameter optimization. Extension of this work includes expanding the proposed dataset to include more traffic signs and proposing an automated architectural search instead of utilizing the available CNN architectures.

ACKNOWLEDGMENT

We would like to thank Universiti Tenaga Nasional for funding this study under the UNIIG2018 internal research grant: J510050819.

REFERENCES

[1] Z. Huang, Y. Yu, J. Gu, and H. Liu, "An efficient method for traffic sign recognition based on extreme learning machine," IEEE transactions on cybernetics, vol. 47, no. 4, pp. 920–933, 2016.

[2] S.-C. Huang, H.-Y. Lin, and C.-C. Chang, "An in-car camera system for traffic sign detection and recognition," in 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSASCIS), pp. 1–6, IEEE, 2017.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[5] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, et al., "Deep speech 2: End-to-end speech recognition in english and mandarin," in International conference on machine learning, pp. 173–182, 2016.

[6] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, "Image captioning with semantic attention," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4651–4659, 2016.

[7] L. Deng and Y. Liu, Deep Learning in Natural Language Processing. Springer, 2018.

[8] 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, vol. 2012, p. 1, 2012.

[9] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, "Deep image: Scaling up image recognition," arXiv preprint arXiv:1501.02876, 2015.

[10] X. B. Quan and W. X. Xiong, "Real-time embedded traffic sign recognition using efficient convolutional neural network," IEEE Access, 2019.

[11] A. Jain, A. Mishra, A. Shukla, and R. Tiwari, "A novel genetically optimized convolutional neural network for traffic sign recognition: A new benchmark on belgium and chinese traffic sign datasets," Neural Processing Letters, pp. 1–25, 2019.

[12] A. Madani and R. Yusof, "Malaysian traffic sign dataset for traffic sign detection and recognition systems," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 8, no. 11, pp. 137–143, 2016.

[13] S. Reilly, I. Clancy, S. Foy, and M. Madden, "Extended malaysian traffic sign dataset (emtd)," Apr 2018.

[14] M. M. Lau, K. H. Lim, and A. A. Gopalai, "Malaysia traffic sign recognition with convolutional neural network," in 2015 IEEE International Conference on Digital Signal Processing (DSP), pp. 1006–1010, IEEE, 2015.

[15] 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," Mathematical Problems in Engineering, vol. 2015, 2015.

[16] K. T. Islam and R. G. Raj, "Real-time (vision-based) road sign recognition using an artificial neural network," Sensors, vol. 17, no. 4, p. 853, 2017.

[17] K. T. Islam, R. G. Raj, and G. Mujtaba, "Recognition of traffic sign based on bag-of-words and artificial neural network," Symmetry, vol. 9, no. 8, p. 138, 2017.

[18] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269, IEEE, 2017.

[19] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[22] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in International conference on machine learning, pp. 1310–1318, 2013.

[23] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep convolutional neural networks," Entropy, vol. 19, no. 6, p. 242, 2017.

[24] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in Advances in neural information processing systems, pp. 3320–3328, 2014.

[25] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, vol. 1. MIT press Cambridge, 2016.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in CVPR09, 2009.

[27] M. D. Zeiler, "Adadelta: an adaptive learning rate method," arXiv preprint arXiv:1212.5701, 2012.

[28] L. Prechelt, "Early stopping-but when?," in Neural Networks: Tricks of the trade, pp. 55–69, Springer, 1998.