COMPOSITE-SCALED EFFICIENTNET FOR ENHANCED MALAYSIAN TRAFFIC SIGN RECOGNITION

XIE HAN LIM[1], WENG KIN LAI[1,2] AND MEI KUAN LIM[3]

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering & Technology, Tunku Abdul Rahman University of Management & Technology, Kuala Lumpur, Malaysia.

[2]*Centre for Multimodal Signal Processing,* Faculty of Engineering & Technology, Tunku Abdul Rahman University of Management & Technology.

[3]School of Information Technology, Monash University Malaysia, Subang Jaya, Selangor.

E-MAIL: *laiwk@tarc.edu.my*

Abstract:

Accurate traffic sign recognition (TSR) is critical for autonomous vehicles, especially in diverse driving environments. While datasets such as the German Traffic Sign Recognition Benchmark (GTSRB) have been extensively studied, limited attention has been paid to specific regional challenges, especially in Malaysia. This study explores traffic sign recognition in Malaysia using a composite-scale convolutional neural network (EfficientNet). The Malaysian Traffic Sign (MTS) dataset contains signs with cultural uniqueness, language, and design differences, which are underexplored compared to global datasets. EfficientNet-B5 was selected for its balance between accuracy and efficiency. Enhancements including image resizing, architecture depth expansion, and data augmentation were applied. Results show that EfficientNet-B5 achieves significant improvements, especially on the MTS dataset, demonstrating the potential for scalable, real-time TSR for autonomous driving systems in Malaysia.

Keywords:

Traffic Sign Recognition, EfficientNet Models, Traditional Machine Learning, Deep Learning, Image Resizing, Adding Augmentation, Adding More Layers on CNN Models

1. Introduction

Advances in autonomous vehicle technology depend on the ability to accurately recognize and interpret traffic signs under a variety of real-world conditions [1]. While deep learning models have achieved remarkable success in traffic sign recognition (TSR), the impact of preprocessing techniques such as image scaling on model performance remains underexplored [2]. This study fills this critical gap by investigating the impact of image scaling algorithms on the performance of EfficientNet, a state-of-the-art convolutional neural network (CNN) architecture, in the TSR task.

The bulk of existing research—such as that by Y.L. Lee—focuses on the German Traffic Sign Recognition Benchmark (GTSRB), which provides a large, balanced, and highly structured dataset [3]. However, the Malaysian Traffic Sign (MTS) dataset presents unique challenges: smaller sample sizes, inconsistent resolutions, and cultural design variations. These characteristics necessitate customized approaches for training robust deep learning models.

This study addresses this research gap by evaluating the EfficientNet family of models, known for its compound scaling technique, across two datasets—MTS and GTSRB. The primary objective is to analyze the performance of EfficientNet-B5, supported by methodological improvements such as image resizing, deeper model architecture, and data augmentation. Emphasis is placed on Malaysian traffic signs to underscore the novelty and regional significance of the work. This study makes three key contributions:

1. We conduct a comprehensive empirical study of EfficientNet architectures (B0–B7) on the GTSRB dataset to identify the most optimal model for the MTS dataset
2. We demonstrate that larger input resolutions can enhance the accuracy by 3.62%
3. We propose methodological enhancements, including layer additions and data augmentation, which improves the performance of B5 on MTS dataset by 17.45%.

The paper is organized as follows. After this introduction, some prior work related to EfficientNet Models is discussed. The details of various methods to improve accuracy, the datasets and the details of the experimental setup are described in Section 3. Section 4 presents the experimental results obtained from the various methods to improve accuracy. Finally, conclusions and some areas for future work are presented in Section 5.

2. Related Work

Review of previous studies indicates that the pre-trained network methods may be classified into 3 main groups, which are *Traditional Machine Learning and Deep Learning, Convolution Neural Network (CNN), and EfficientNet Models.*

2.1. Traditional Machine Learning and Deep Learning

Traffic sign recognition (TSR) has evolved significantly with advancements in deep learning, particularly Convolutional Neural Networks (CNNs). Traditional machine learning methods, such as Support Vector Machines (SVMs), were initially employed for TSR but required manual feature extraction, limiting their adaptability and accuracy. *Lai* et al. [4] demonstrated that CNNs outperform SVMs in image recognition tasks, achieving 98.85% accuracy on the MNIST dataset compared to 93.92% for SVMs, highlighting the superiority of automated feature learning in deep learning.

2.2. Convolution Neural Network (CNN)

Recent studies have focused on optimizing CNN architectures for TSR. *Shustanov and Yakimov* [5] achieved 99.94% accuracy on the GTSRB using a modified Hough transform with CNN preprocessing, emphasizing the importance of real-time processing. Similarly, *Hechri and Mtibaa* [6] proposed a two-stage CNN with filtering techniques, attaining 99.37% accuracy on GTSRB, though their model struggled with triangular signs due to shape variability. These works underscore the need for robust preprocessing to handle noise and occlusion.

2.3. EfficientNet Models

EfficientNet has emerged as a scalable solution for TSR, balancing accuracy and computational efficiency. *Bouderbal* et al. [7] compared EfficientNet variants (B0–B7) with MobileNetV2 on GTSRB, showing that EfficientNet-B7 achieved 98.21% accuracy, outperforming MobileNetV2 (91.66%). This aligns with findings by *YL Lee* [3], where EfficientNet-B0 achieved 99.08% accuracy on GTSRB and 95.06% on the China Traffic Sign (CTS) dataset, though CTS's text-heavy signs posed challenges. The scalability of EfficientNet, via compound scaling of depth, width, and resolution, makes it suitable for resource-constrained applications like autonomous vehicles.

Beyond TSR, EfficientNet has proven effective in medical imaging. *Savas and Damar* [8] used transfer learning with EfficientNet-B5 to classify brain MRI images, achieving 98.39% accuracy after hyperparameter tuning. Similarly, *Naidji and Elberrichi* [9] employed EfficientNet-B7 with attention mechanisms for COVID-19 detection in chest X-rays, attaining 96.5% accuracy. These applications demonstrate EfficientNet's versatility and robustness across domains.

Despite these advances, gaps remain in applying EfficientNet to diverse traffic sign datasets like CTS and Malaysia Traffic Signs (MTS). Previous studies primarily focused on GTSRB, leaving room for optimization in multi-lingual or region-specific sign recognition. Our work addresses this by evaluating EfficientNet-B5's performance across GTSRB, CTS, and MTS, with enhancements such as image resizing and augmentation.

3. Methodology

In this study, special attention is given to the MTS dataset due to its regional importance and underrepresentation in TSR research. Unlike the well-established GTSRB dataset, the MTS dataset is smaller and contains culturally unique traffic signs, requiring targeted preprocessing and architectural strategies for effective learning.

3.1. Traffic Sign Images

In this section, we briefly outline the 2 main datasets used in this study: GTSRB [10], and MTS datasets.

The GTSRB is one of the most extensive benchmarks for evaluating traffic sign recognition (TSR) algorithms [11]. It contains 51,839 images covering 43 different types of traffic signs, fully representing the variations in real-world scenes with various lighting, weather conditions, and partial occlusions. Its extensive coverage and detailed annotations make it a valuable resource for benchmarking TSR techniques.

In contrast, the MTS dataset is tailored to the Malaysian road environment and presents a unique set of challenges. It contains only around 2,001 images across 50 categories, making it considerably smaller and less balanced than GTSRB. Malaysian traffic signs often include bilingual text in Malay and English, as well as region-specific warnings and instructions—such as the frequent use of "*BERHENTI*" (meaning "STOP"). These signs are more varied in appearance and include complex elements like symbols, shapes, colors, and language that require deeper contextual understanding. Furthermore, the MTS dataset contains images with significantly higher variability in lighting, extreme weather conditions (such as morning, night, heavy rain, fog, and glare), and resolution, making it a more realistic representation of real-world driving scenarios compared to the GTSRB. Figure 1 shows GTSRB and MTS's samples.

(a) GTSRB

(b) MTS

FIGURE 1. Sample Traffic Sign Images

Due to these differences, models trained on GTSRB typically perform well in structured and uniform environments but may struggle with the irregularities present in MTS. Thus, working with the MTS dataset requires more robust preprocessing techniques and model tuning to handle its complexities. Ultimately, while GTSRB is ideal for benchmarking and academic comparisons, the MTS dataset is more relevant for real-world applications in Malaysia and similar regions, highlighting the importance of region-specific datasets in autonomous vehicle development.

3.2. Image Resizing

Image resizing is a crucial preprocessing step in computer vision and deep learning that ensures input images have consistent dimensions, which is essential for neural network models. It involves scaling images using interpolation methods like bilinear or bicubic, and the choice of technique can impact model performance by either preserving or distorting image details. Resizing may include maintaining aspect ratio or adapting dimensions as needed, often paired with normalization to enhance efficiency. Tools like OpenCV and PIL simplify this process [12], [13] and advanced techniques such as adaptive resizing or padding are used in tasks like object detection to retain important features. Overall, resizing is a simple yet vital step to ensure compatibility between image data and model architecture.

3.3. Increasing number of layers on CNN Model

Adding more layers to a convolutional neural network can improve its ability to learn complex and hierarchical features; however, this enhancement comes with increased computational demands and a higher risk of overfitting, particularly when dealing with smaller or imbalanced datasets like the Malaysian Traffic Sign (MTS) dataset. EfficientNet addresses this challenge through compound scaling, which proportionally increases depth (number of layers), width (number of channels), and resolution (input image size), rather than focusing solely on depth. In this study, although deeper models like EfficientNet-B7 were evaluated, they did not outperform mid-sized models such as EfficientNet-B5, which demonstrated better accuracy and efficiency. This suggests that simply adding more layers does not always result in better performance and may even lead to diminishing returns. EfficientNet-B5 proved to be a more practical and effective choice, striking the right balance between model complexity and recognition accuracy, especially in the context of region-specific and complex datasets such as MTS.

3.4. Augmentation

Data augmentation is a technique that increases the size and diversity of a training dataset by applying random transformations such as rotation, flipping, scaling, and color adjustments to existing data. This helps improve model generalization and reduce overfitting, especially when the dataset is small or lacks diversity. The effectiveness of data augmentation depends on the relevance of the chosen transformation to the specific task. While simple techniques are commonly used, more advanced methods like GANs or style transfer can also generate diverse data, although their implementation is more complex. Libraries such as TensorFlow and PyTorch provide built-in support for data augmentation, making it easy to integrate into the training workflow. In summary, data augmentation is a cost-effective way to improve model robustness and performance, especially in fields where collecting labeled data is challenging, such as medical imaging and autonomous driving.

4. Results and Discussion

4.1. Investigating EfficientNET Model for optimal Traffic Sign Recognition

TABLE 1. Default Parameter Settings

Parameter	Details
Input Image Size	60 x 60 x 3
Bath Size	32
Maximum Training Epochs	100
Early Stopping Criteria	No improvement in Validation Loss for more than 10 epochs
Loss Function	Categorical Cross-Entropy
Optimization Algorithm	Adam with an initial learning rate 0.001

TABLE 2. Result of GTSRB using Default Parameter for various EfficientNET models

EfficientNet Model	Accuracy	Loss
B0	95.81	19.85
B1	96.23	19.03
B2	96.26	16.89
B3	95.78	19.13
B4	97.13	10.35
B5	97.26	13.74
B6	96.18	13.84
B7	95.52	16.66

This section evaluates the performance of EfficientNet models for traffic sign recognition, with a primary focus on the MTS dataset due to its practical relevance and unique challenges. The evaluation began by benchmarking eight EfficientNet variants (B0 through B7) on the GTSRB dataset. This dataset, being large, well-labeled, and relatively balanced, served as an ideal platform for establishing baseline performance. ***EfficientNet-B5*** achieved the highest classification accuracy of ***97.26***% among all variants, outperforming deeper models like B6 and B7, which suffered slightly from diminishing returns and increased computational cost. This makes B5 a practical choice for embedded systems used in real-time autonomous vehicle applications.

TABLE 3. Test Accuracy of MTS Dataset on various EfficientNET models

EfficientNet Model	Test Accuracy		Difference
	GTSRB	MTS	
B0	95.81	76.08	19.73
B1	96.23	77.46	18.77
B2	96.26	66.74	29.52
B3	95.78	83.26	12.52
B4	97.13	71.88	25.52
B5	97.26	86.78	10.48
B6	96.18	81.92	14.26
B7	95.52	77.23	18.29

The table compares EfficientNet models on the MTS dataset, with B5 achieving the highest accuracy (86.78%)—a 10.7% improvement over B0 (76.08%). MTS presents challenges like limited data, inconsistent images, and bilingual signs, making recognition difficult. B5's success stems from its compound scaling of depth, width, and resolution, enabling robust feature extraction. It also shows the smallest accuracy gap (10.48%) between GTSRB and MTS, proving its strong generalization. In contrast, models like B2 and B4 struggle with larger drops (29.52% and 25.52%), revealing their limitations. These results demonstrate that advanced architectures like B5 perform best in complex, real-world traffic scenarios.

4.2. Image Resize on EfficientNET-B5 Model

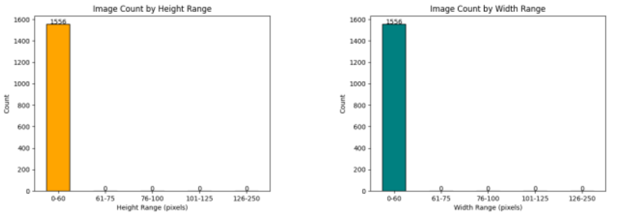

FIGURE 2. Image dimension visualization of the MTS ***Train*** Dataset

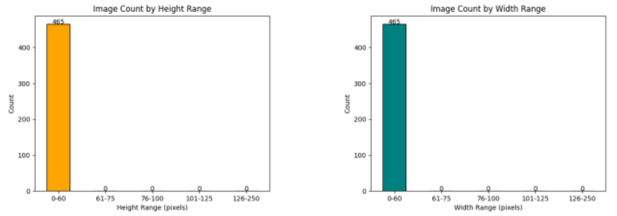

FIGURE 3. Image dimension visualization of the MTS ***Test*** Dataset

TABLE 4. Result of 60 x 60 (Original Image Size) VS. Other Sizing

Datasets	Image Size	Accuracy	Loss
MTS	60 x 60	86.78	70.73
	75 x 75	88.47	53.66
	100 x 100	90.40	45.60
	125 x 125	89.06	53.42

This part of the study evaluates the impact of image resizing on the EfficientNet-B5 model using the MTS dataset, with results summarized in Table 4 and visualized in Figures 2 and 3. Resizing images from 60×60 to 100×100 pixels improved accuracy from 86.78% to 90.40% and reduced loss from 70.73 to 45.60, indicating better feature recognition. However, larger input sizes (100×100) require about 2.78 times more computations. Despite this, the accuracy gain justifies the resizing for more critical applications, given the variability of the MTS dataset. Increasing the size to 125×125 caused a slight accuracy drop (89.06%) and higher loss (53.42%), suggesting diminishing returns. Figures 3 and 4 show that resizing to 100×100 aligns with the natural image size distribution in the MTS dataset, helping to standardize inputs and reduce variability.

4.3. Impact of Layers on EfficientNET-B5

TABLE 5. Hyperparameters for various layers

Hyperparameter	Values		
	1 Layer	2 Layers	3 Layers
Base Model		EfficientNet-B5	
Dense Layer 1 Units	256	512	512
Dense Layer 1 Activation		ReLU	
Dropout 1 Rate	0.5	0.5	0.5
Dense Layer 2 Units	-	256	256
Dense Layer 2 Activation		ReLU	

Dropout 2 Rate	-	0.5	0.5
Dense Layer 3 Units	-	-	128
Dense Layer 3 Activation		ReLU	
Dropout 3 Rate	-	-	0.5
Output Dense Layer		Softmax	

TABLE 6. Result of adding 1 Layer VS. More Layers

Datasets	Add Layers	Accuracy	Loss
MTS	0 Layer	89.73	141.05
	1 Layer	90.40	45.60
	2 Layers	92.50	39.45
	3 Layers	92.54	39.97

This section explores how modifying the EfficientNet-B5 architecture by adding more fully connected layers affects the model's learning ability and accuracy. The experiments involve three enhancement schemes shown in Table 5: adding one, two, and three fully connected layers to the original EfficientNet-B5 architecture. These enhancements aimed to improve the network's ability to capture fine-grained details, especially in Malaysian traffic signs that often include complex icons, mixed text, and color contrasts. On the MTS dataset, accuracy improved progressively with each added layer, peaking at 92.54% with three dense layers. This indicates that deeper models are better equipped to learn intricate patterns in challenging datasets. However, it's important to note that overly deep networks can introduce risks of overfitting, particularly when working with smaller datasets like MTS, and should therefore be implemented with proper regularization techniques such as dropout.

4.4. Impact of Data Augmentation on Model Performance

TABLE 7. Methodological Enhancements for Augmentation

Augmentation Methods	Tunning Value
Width Shift Range	0.2
Height Shift Range	0.2
Shear Range	0.15
Zoom Range	0.15
Fill Mode	Nearest

TABLE 8. Result of With Augmentation VS. Without Augmentation

Datasets	Add Augmentation	Total Train Images	Total Test Images	Accuracy	Loss
MTS	No Aug	1556	465	92.54	39.97
	Aug	9336	2790	93.53	31.37

This section analyzes the effect of applying data augmentation techniques on improving the generalization ability of the EfficientNet-B5 model. Table 7 shows the data augmentation strategies, including width and height translation, shear transformation, scaling, and various padding modes. These techniques are used to artificially expand the training dataset and simulate real-world situations that a vehicle vision system may encounter, such as rotation, distortion, or occlusion. On the MTS dataset, augmentation increased the number of training images significantly and led to an improvement in accuracy from 92.54% to 93.53%. Furthermore, the model's loss was reduced from 39.97 to 31.37, highlighting improved generalization. These results underscore the importance of augmentation for smaller, diverse datasets where overfitting is a common issue. By introducing artificial variability, the model becomes more resilient to changes in lighting, angle, and obstruction—conditions typical of real-world driving environments.

4.5. Performance Comparison

TABLE 9. Result of CNNs on all Datasets (Previous Work)

CNNs Model on MTS	Test Accuracy
Lenet-5	58.48
Lenet	58.78
AlexNet	71.88
MobileNet	74.40
ResNet-50	75.41
InceptionV3	75.73
VGG-16	75.97
EfficientNet-B0	76.08
EfficientNet-B5 + 3 FC Layers + Augmentation (EB5.3FC.Aug)	93.53

The last section provides a comprehensive comparison of the optimized our model (***EB5.3FC.Aug***) with earlier studies based on EfficientNet-B0 and other traditional CNN architectures. With the MTS dataset, ***EB5.3FC.Aug*** achieved the highest test accuracy of ***93.53%,*** a significant improvement from MobileNet's 74.40% and ResNet-50's 75.41%. Even EfficientNet-B0, a lighter version of the same family, only achieved 76.08%. The performance improvement is attributed to the compound scaling strategy of ***EB5.3FC.Aug***, which balances depth, width, and resolution. Although these improvements may seem insignificant in terms of numerical value, they are crucial in the field of autonomous driving, where accuracy and reliability are critical to safety. These results further solidify ***EB5.3FC.Aug's*** viability as a state-of-the-art solution for traffic sign recognition and set a new benchmark in this field.

TABLE 10. Comparative Test Accuracy of GTSRB and MTS Datasets

Datasets	CNNs Model	Test Accuracy
GTSRB	EfficientNet-B0	99.08
MTS	EB5.3FC.Aug	93.53

Table 10 presents a comparison of the test accuracy between two traffic sign datasets: the German Traffic Sign Recognition Benchmark (GTSRB) and the Malaysia Traffic Sign (MTS) dataset, using different EfficientNet models.

The GTSRB dataset achieved a test accuracy of 99.08% with EfficientNet-B0, while the MTS dataset attained 93.53% accuracy with ***EB5.3FC.Aug***. The higher accuracy of the GTSRB dataset may be attributed to factors such as a larger and more diverse collection of labeled images, standardized lighting and environmental conditions, or the maturity of the dataset, which has been widely used and refined in research. In contrast, the MTS dataset's slightly lower accuracy could reflect challenges like fewer training samples, greater variability in sign designs, or less optimal image quality. The choice of EfficientNet models—B0 for GTSRB and B5 for MTS—suggests a trade-off between computational efficiency and performance, with B5 likely selected for MTS to handle its potential complexities. Overall, the results highlight the influence of dataset characteristics and model selection on traffic sign recognition performance.

5. Conclusions

This study explored the application of EfficientNet models for traffic sign recognition, with a focus on the unique challenges posed by the Malaysian Traffic Sign (MTS) dataset. The results demonstrated that EfficientNet-B5, enhanced with image resizing, additional layers, and data augmentation, achieved a significant improvement in accuracy, reaching 93.53% on the MTS dataset. This performance surpassed earlier models like EfficientNet-B0 and traditional CNNs, highlighting the effectiveness of compound scaling and targeted preprocessing techniques. The findings underscore the importance of region-specific adaptations in traffic sign recognition, particularly for datasets with cultural and linguistic diversity. The success of EfficientNet-B5 in handling the complexities of the MTS dataset suggests its potential for real-world applications in autonomous driving systems, where accuracy and robustness are critical. Future work could explore further optimizations, such as advanced augmentation techniques or hybrid architectures, to address remaining challenges in diverse and dynamic environments. Overall, this study contributes to the growing body of research on scalable and efficient deep learning solutions for traffic sign recognition, with practical implications for autonomous vehicle technology in Malaysia and beyond.

Acknowledgements

XH Lim and WK Lai would like to express their gratitude to TAR UMT for the financial support in this work.

References

[1] L. Tan, 'Artificial Intelligence In Autonomous https://www.imaginationtech.com/future-of-automotive/how-do-autonomous-cars-work/artificial-intelligence-in-autonomous-vehicles/.

[2] Lee, 'Examining Autonomous Car Accidents and Vehicles', Statistics', https://www.lgrlawfirm.com/blog/examining-autonomous-car-accidents-and-statistics-2/

[3] Y. L. Lee, "Design and development of EfficientNet for traffic sign recognition," B.Eng. thesis, Tunku Abdul Rahman Univ., 2023, 39 pages.

[4] Y. Lai, "A comparison of traditional machine learning and deep learning in image recognition," *J. Phys.: Conf. Ser.*, vol. 1314, no. 1, p. 012148, 2019.

[5] A. Shustanov and P. Yakimov, "CNN design for real-time traffic sign recognition,"*Procedia Eng.*, vol. 201, pp. 718–725, 2017.

[6] A. Hechri and A. Mtibaa, "Two-stage traffic sign detection and recognition based on SVM and CNNs," *IET Image Process.*, vol. 14, no. 5, pp. 939–946, 2020.

[7] I. Bouderbal et al., "An analytical study of efficient CNNs tuning and scaling for traffic signs recognition," in *Proc. IEEE ICRAMI*, 2021, pp. 1–6.

[8] S. Savas and C. Damar, "Transfer-learning-based classification of pathological brain MRIs," *ETRI J.*, vol. 46, no. 2, pp. 263–276, 2024.

[9] M. R. Naidji and Z. Elberrichi, "Automatic detection of COVID-19 from chest X-rays using EfficientNet-B7," *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, pp. 1443–1456, 2024.

[10] Stallkamp, J. et al. (2012) 'Man vs. computer: Benchmarking machine learning algorithms for Traffic sign recognition', Neural Networks, 32, pp. 323–332. doi:10.1016/j.neunet.2012.02.016.

[11] Lim, X.R. et al. (2023) 'Recent advances in traffic sign recognition: Approaches and datasets', Sensors, 23(10), p. 4674. doi:10.3390/s23104674.

[12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[13] A. Clark, "Pillow (PIL Fork) Documentation," https://pillow.readthedocs.io/en/stable/, 2015. [Accessed: Apr. 26, 2025].