Nature Environment and Pollution Technology
An International Quarterly Scientific Journal
p-ISSN: 0972-6268 (Print copies up to 2016)
e-ISSN: 2395-3454
Vol. 23 No. 4 pp. 2461-2468 2024
Original Research Paper
https://doi.org/10.46488/NEPT.2024.v23i04.052
Open Access Journal

Enhancing Driving Safety and Environmental Consciousness through Automated Road Sign Recognition Using Convolutional Neural Networks

M. H. F. Md Fauadi[1]†, M. F. H. Mohd Zan[2], M. A. M Ali[1], L. Abdullah[1], S. N. Yaakop[3] and A. Z. M. Noor[4]
[1]Fakulti Teknologi dan Kejuruteraan Industri dan Pembuatan, Universiti Teknikal Malaysia Melaka, 76100 Melaka, Malaysia
[2]HSK Consult, Bandar Tasik Selatan, 57000, Kuala Lumpur, Malaysia
[3]Bahagian Governan dan Kecemerlangan, Jabatan Pendidikan Politeknik dan Kolej Komuniti, 62100 Putrajaya, Malaysia
[4]Universiti Kuala Lumpur Malaysian Spanish Institute, Kulim Hi-Tech Park, 09000 Kulim, Kedah, Malaysia
†Corresponding author: M.H.F. Md Fauadi; hafidz@utem.edu.my

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 31-01-2024
Revised: 18-03-2024
Accepted: 05-04-2024

Key Words:
Traffic sign recognition
Convolutional Neural Network
YOLOv3 network
Environmental consciousness

ABSTRACT

Traffic accidents remain a pressing public safety concern, with a substantial number of incidents resulting from drivers' lack of attentiveness to road signs. Automated road sign recognition has emerged as a promising technology for enhancing driving assistance systems. This study explores the application of Convolutional Neural Networks (CNNs) in automatically recognizing road signs. CNNs, as deep learning algorithms, possess the ability to process and classify visual data, making them well-suited for image-based tasks such as road sign recognition. The research focuses on the data collection process for training the CNN, incorporating a diverse dataset of road sign images to improve recognition accuracy across various scenarios. A mobile application was developed as the user interface, with the output of the system displayed on the app. The results show that the system is capable of recognizing signs in real time, with average accuracy for sign recognition from a distance of 10 meters: i) daytime = 89.8%, ii) nighttime = 75.6%, and iii) rainy conditions = 76.4%. In conclusion, the integration of CNNs in automated road sign recognition, as demonstrated in this study, presents a promising avenue for enhancing driving safety by addressing drivers' attentiveness to road signs in real-time scenarios.

INTRODUCTION

With the growing demand for intelligent transport systems and the pursuit of safer road environments, automated road sign recognition has emerged as a promising technology. Convolutional Neural Networks (CNNs), a class of deep learning algorithms inspired by human visual processing, have revolutionized various fields, including image recognition tasks (Behera et al. 2022, Khan et al. 2023, Fredj et al. 2023, Kiliçarslan et al., 2023, Lee et al. 2021, Razi et al. 2023). Leveraging the power of CNNs, researchers, and engineers have made significant strides in developing automated road sign recognition systems that exhibit high accuracy and efficiency. This study aims to explore the application of CNNs in automatically recognizing road signs, and addressing challenges posed by diverse road sign designs, environmental conditions, and real-time processing requirements.

This research introduces a novel solution to tackle challenges in road sign recognition by presenting a robust and efficient framework that surpasses existing methodologies in UAV inspection scenarios. The successful integration of CNN and You Only Look Once version 3 (YOLOv3) methodologies highlights the advantages of employing deep learning techniques to enhance road sign recognition algorithms. The framework's development involved extensive investigations into dataset generation techniques, picture classification methods, and network parameter determination, all aimed at optimizing the algorithm's performance.

Recognizing traffic signs is crucial for automated driving and driver assistance systems. However, various factors such as partial occlusion, diverse views, varying illuminations, and weather conditions pose significant challenges for computers in visually identifying traffic sign images. Researchers are actively addressing this complex task using established or specifically developed computer vision algorithms. Before the release of standardized benchmarks like the German Traffic Sign Benchmark (GTSRB) and German Traffic Sign Database (GTSDB), researchers lacked publicly accessible datasets for comparison, hindering the evaluation and comparison of methods. Despite the availability of benchmarks, limitations persist, such as the focus on symbol-based traffic signs with regular shapes and colors, the reliance on static images, and the lack of comparability in identifying existing signs in a scene.

Researchers have made significant advancements in accurate road sign recognition by building a big dataset of road sign images and training CNNs on vast repositories of annotated data. Real-time implementation of CNN-based road sign recognition in cars is expected to enhance driving assistance systems by providing drivers with up-to-date information on speed limits, warnings, road signs, and directions (Luo et al. 2017, Chen et al. 2019, Dewi et al. 2022). The use of CNNs for road sign recognition has been shown to potentially reduce accidents caused by drivers misinterpreting or failing to observe road signs (Khan et al. 2023, Luo et al. 2017). This could help to improve the safety and dependability of driving. However, as stated there are still limitations to the technology which include processing of partially obstructed images or noisy images. This study aims to critically address the limitations and opportunities for further refinement to ensure the seamless integration of CNN-based road sign recognition in modern driving environments. Apart from CNN, other methods have also been deployed to enable autonomous vehicles (Sudhakar & Priya 2023, Ramlan et al. 2022, Rosli et al. 2018, Noor et al. 2017, Bin Md Fauadi & Murata 2010).

This study integrates the CNN and YOLOv3 approaches to create a unique framework for an algorithm that recognizes traffic signs. The suggested method improves the quality of acquired road sign pictures by high-resolution restoration of blurry images using CNN. The enhanced clear images are processed with more training images to increase the dataset size and enhance the network's recognition performance. YOLOv3 is then utilized for precise road sign recognition in real-time situations. Extensive investigations were carried out on various aspects of the proposed framework. Furthermore, the dataset collection approach, picture classification methods, and network parameter determination were critically analyzed to optimize the algorithm's performance.

Traffic sign recognition is a critical factor for driver assistance systems and automated driving. Additionally, as the system needs to deal with a realistic environment, partial occlusion, varied angles, different illuminations, weather, and other factors make it challenging for the system to visually identify photographs of traffic signs. Most methods for recognizing traffic signals in an image involve two primary steps: detection and classification. Many academics are using well-established or specially created computer vision algorithms to tackle this difficult issue (Huang et al. 2020).

There was no publicly available dataset for comparison before the publication of the German Traffic Sign Benchmark (GTSDB) and German Traffic Sign Benchmark (GTSRB) (Stallkamp et al. 2012). As a result, researchers have a uniform dataset to assess and contrast their approaches with. Nevertheless, there are still issues with GTSRB and GTSDB:

i. Both benchmarks encompass only three types of symbol-based traffic signs with regular shapes and colors, which are comparatively easier to detect and classify than text-based traffic signs.
ii. GTSDB solely comprises static images, yet in practical scenarios, continuous video footage captured by an in-vehicle camera proves more beneficial for detection and classification (Luo et al. 2017).
iii. The ultimate goal of traffic sign recognition is to identify existing signs in a scene, but the two benchmarks lack comparability in this aspect.

The unique contribution of this research lies in the innovative framework that integrated CNN and YOLOv3 methodologies, addressing challenges in road sign recognition and surpassing existing methodologies in UAV inspection scenarios. The innovative framework developed in this research offers a distinct advantage by compensating for the potential lack of state-of-the-art equipment in road sign recognition systems. While traditional approaches may rely heavily on advanced hardware or costly infrastructure to achieve accurate results, the integration of CNN and YOLOv3 methodologies provides an alternative solution. By leveraging deep learning techniques within the proposed framework, the system can achieve high levels of accuracy and efficiency without necessarily requiring the latest and most expensive hardware components. This means that the framework can be implemented in scenarios where access to state-of-the-art equipment may be limited or cost-prohibitive, making it a more accessible and practical solution for various applications.

The remainder of the paper is divided into the following sections: The method for recognizing the road signs is described in Section II, along with the function of each component of the framework. This section also provides a thorough introduction to the theoretical underpinnings of the proposed CNN algorithm and the proposed YOLOv3's detection principle. Section III discusses the dataset, experimental findings, assessment of our algorithms, and comparison to alternative approaches. The last section of this essay is Section IV.

MATERIALS AND METHODS

Proposed Model

Fig. 1 illustrates the main procedure of CNN for detecting

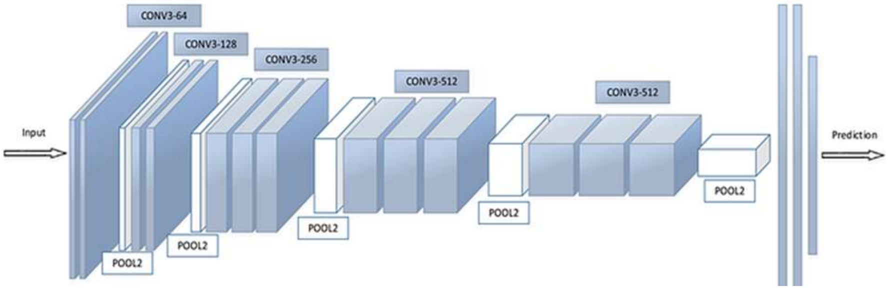


Fig. 1: CNN model.

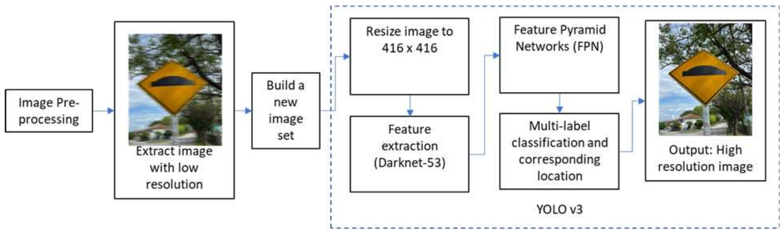


Fig. 2: Proposed flow.

the primary electrical components (Chen et al. 2019). In this research, preprocessing of the sign images involves the steps. Initially, the original image dataset is categorized into two groups. The first group comprises qualified images suitable for use as the training set, while the second group consists of blurred images with lower resolutions. Subsequently, the blurred image set undergoes super-resolution reconstruction using CNN, thereby enhancing its resolution to match that of the original images. The resulting processed images are then merged with the original images to form the appropriate inspection image sets.

Furthermore, based on Peng et al. (2021) and Qingyun et al. (2020), the original inspection image set is resized to a resolution of 416×416 within the YOLOv3 architecture. The resized images are subsequently input into Darknet53, enabling the extraction of relevant features associated with electrical components. The feature pyramid networks (FPN) are subsequently utilized to generate predictions across three distinct scales using the feature outputs from Darknet-53. The comprehensive predictions obtained from YOLOv3 encompass essential parameters, including bounding box information, objectness score, and class predictions. To refine the predictions, YOLOv3 employs a filtering process to remove anchors that exhibit substantial overlap with the ground truth object, subject to a selected threshold. Following this filtering step, the network proceeds to output the classification outcomes and corresponding positioning information for each bounding box. Ultimately, the YOLOv3 network yields comprehensive detection results. The proposed flow for this study is depicted in Fig. 2.

Experimental Setup

The dataset for this research project comprised a total of 450 images, consisting of 125 negative images and 325 positive images. The negative images encompassed road scenes containing objects like buildings, trees, cars, and roads devoid of road signs. These images were captured in Ayer Keroh, Melaka for several days. It is worth noting that all images used in the dataset were obtained directly from phone-captured photographs, and no images from external internet sources were utilized. The dataset exhibited diverse conditions, including variations in lighting and time of day, providing a comprehensive representation of real-world scenarios. Table 1 shows the summary of data collection.

The experiment was conducted on a personal computer equipped with an Intel(R) Core(TM) Pentium processor running at 3.7 GHz. It was configured with 8GB of RAM and operated on a 64-bit operating system. Data collection

Table 1: Data collection summary.

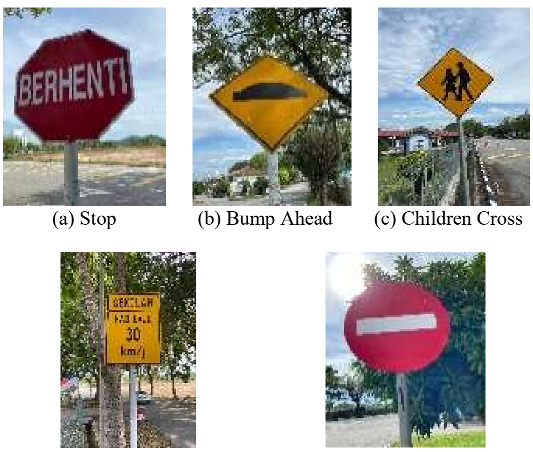
Setup for Data Collection	
Distance (Camera-to-sign)	3-7 meters
Time	9-11 am 9-11 pm
Weather Condition	Clear and rainy
Camera	12-megapixel iPhone 11 rear camera

employed an iPhone 11 camera featuring a high-definition (HD) 12-megapixel (MP) rear camera.

Moreover, the personal computer is equipped with Python 3.7.2 (64-bit), enabling clear and coherent programming with the added benefit of automatic memory management. The Python libraries used in this project include Opencv and Numpy, essential for image processing and numerical computations, respectively. Additionally, Google Colab serves as the software platform for model training, eliminating the need to install or configure additional software on the computer. This streamlined approach ensures a straightforward and efficient training process for road sign image recognition.

Dataset Preparation

To construct the targeted dataset, five types of road signs as illustrated in Fig. 3 (a) Stop, (b) Bump Ahead, (c) Children Crossing, (d) Speed Limit and (e) No Entry - were meticulously collected as the primary data. The labels for these road signs are listed in Table 2. These five road sign types are commonly encountered on Malaysian roads. A total of 750 images were gathered, comprising 150 images for each road sign category. Subsequently, the dataset was automatically constructed using Google Colab, streamlining the data preparation process.



(a) Stop (b) Bump Ahead (c) Children Cross

(d) Speed Limit (e) No Entry

Fig. 3: Road signs.

Table 2: Labels used for the road sign.

Road Sign	Label (in Malay language)
Stop	Berhenti
Bump Ahead	Bonggol
Children Cross	Kanak-Kanak Melintas
Speed limit	Had laju 30km/j
Children Cross	Kanak-Kanak Melintas

Network Structure of the Proposed CNN

The CNN network initially resizes the extracted blurred inspection image to the desired target size using the Bicubic interpolation algorithm, denoted as Y. The primary objective of super-resolution reconstruction is to recover Y to the high-resolution image H, resembling the original resolution image X. This is achieved through training to obtain the corresponding "end-to-end" mapping function F(Y). The architectural representation of the CNN network is illustrated in Fig. 1, comprising a multi-layer CNN. The network is divided into three levels, corresponding to the three successive steps involved in image super-resolution reconstruction:

- The initial convolutional layer extracts image blocks from Y and represents these features at a lower resolution level.
- The subsequent convolutional layer performs non-linear mapping to generate high-resolution features.
- The final convolutional layer reconstructs high-resolution images, effectively producing images closely resembling the original high-resolution images.

Training and Classification Using YOLOv3

YOLOv3, a real-time object identification technique utilizing neural networks, was employed for training and classification. Due to its remarkable combination of speed and accuracy, this algorithm has gained widespread adoption among users. Notably, it has successfully detected a diverse range of objects, including traffic lights, pedestrians, parking meters, and animals. Key advantages of YOLOv3 include its speed, high accuracy, and learning capabilities in object representation and detection.

The feature detector Darknet-53 in YOLOv3's architectural design draws inspiration from established models such as ResNet and Feature Pyramid Network (FPN). With 52 convolutions featuring skip connections akin to ResNet and three integrated prediction heads similar to FPN, Darknet-53 exhibits the capacity to process images at various spatial compressions. This amalgamation of influential designs empowers Darknet-53 with the ability to effectively detect features in images across different spatial r esolutions.

RESULTS AND DISCUSSION

The assessment of the overall system performance primarily focuses on evaluating detection and classification accuracy across diverse conditions encompassing varying views, angles, distances, light intensities, and driving environments. Results are systematically organized into a tabular format, incorporating values for True Positive, True Negative, False Positive, False Negative, error rates, and accuracy metrics. The comprehensive evaluation of the system involves employing computational theories through Google Colab, generating quantitative metrics crucial for assessing the system's efficacy and robustness. The evaluation metrics utilized in this study are defined as follows:

- True Positive: Video frames with road signs precisely detected and recognized as such by the experiment.
- True Negative: Video frames without road signs precisely detected and recognized as lacking road signs by the experiment.
- False Positive: Video frames without road signs incorrectly detected and recognized as having road signs by the experiment.
- False Negative: Video frames with road signs incorrectly detected and recognized as lacking road signs by the experiment.

Detection and Recognition based at Varying Distances

The experiment involved detection and recognition procedures conducted at varied distances from the road sign, encompassing short and long ranges, captured during both day and night. Short-distance images were taken within 3 meters, while long-distance images were captured from 7 meters. Three distinct tests were performed, each involving fixed distances and road settings, accounting for daytime and nighttime conditions. Images, taken using an iPhone 11 with enhanced resolution, depict the detection and recognition outcomes of the five road sign types at a 3-meter distance in daylight. To ensure accuracy, each image was captured thrice, and average accuracy was computed.

Similarly, experiments were conducted at night between 8:00 pm to 10:00 pm, with illumination lower than 20 lux, aiming to evaluate system performance under varied lighting conditions. Fig. 4 illustrates the detection and recognition outcomes of the five road sign types at distances of 3 to 7 meters during nighttime conditions. Similar to daytime evaluation, images were captured thrice at the same distance, and average accuracies were determined. These comprehensive evaluations under different conditions aim to ensure robust and reliable system performance.

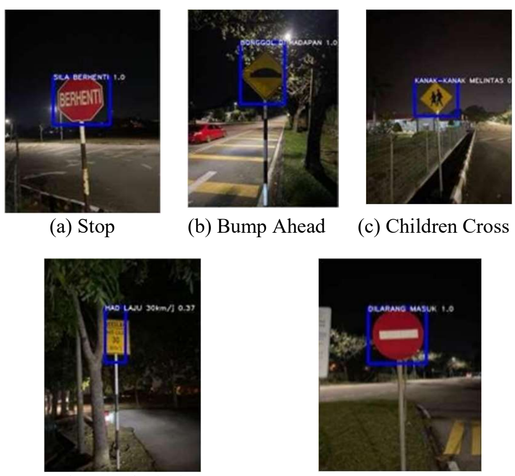
Accuracy in Relation to Distance - Day and Night

The results demonstrate data collected from three repetitions for each road sign at distances of 3 to 10 meters. The experiment's highest accuracy of 95.00% was achieved for the Speed Limit label, while the lowest accuracy of 75.00% was observed for the Stop label.

Additionally, during nighttime recognition at a 3-meter distance with an illumination of 10 lux, the Speed Bump demonstrated the highest accuracy of 88%, whereas the Speed Limit label recorded the lowest accuracy of 62.00%. This revised version aims to improve the academic tone, clarity, and coherence of the content while ensuring the accurate presentation of experimental procedures and results. The results are shown in Figs. 5 and 6 respectively.

Image Retraining

Fig. 7 illustrates the results, indicating that the accuracy of image detection for Stop and Children crossing signs obtained the lowest scores compared to image detection at a 3-meter distance. It was observed that both road signs did not undergo complete training in Google Colab due to



(a) Stop (b) Bump Ahead (c) Children Cross

(d) Speed Limit (e) No Entry

Fig. 4: Sample of images captured during night time.

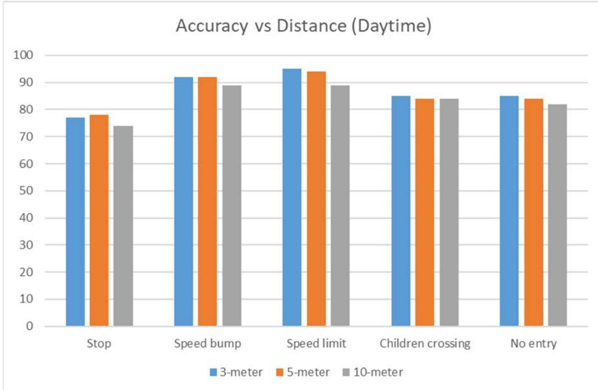


Fig. 5: Results accuracies (daytime).

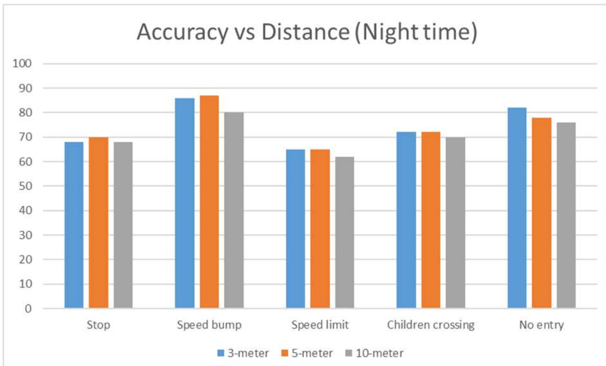


Fig. 6: Results accuracies (night time).

the platform's limited training time, capped at 60 hours. Consequently, retraining becomes imperative to enhance the accuracy score, allowing these road signs to achieve higher accuracy levels. However, achieving optimal accuracy for "Berhenti" and "Bonggol" signs might necessitate a lifetime purchase of training resources, given the insufficient training time available in Google Colab for comprehensive learning.

The study presents a comparative analysis of detection distances at 3-, 5- and 10-meters during daytime, both before and after retraining. Post retraining for image detection at a 10-meter distance, the accuracy for "Stop" and "Speed bump" signs exhibited significant improvement, increasing from 74% to 84% for "Stop"; and from 88% to 92% for "Speed bump." The enhancement is depicted in Figs. 7 and 8, illustrating notable accuracy rates for all signs during daytime and night time testing, respectively.

Detection and Recognition Under Rainy Conditions

This experiment specifically aimed to evaluate the detection and recognition capabilities of the system in rainy weather conditions, assessing its ability to identify road signs under adverse circumstances. The image captures were conducted around 9:00 a.m. during daylight hours. To ensure a comprehensive assessment, three separate tests were conducted, each utilizing distinct images to showcase the system's accuracy in detecting images under rainy conditions. All images were captured using an iPhone 11 with enhanced resolution. The results also shed light on the system's accuracy in detecting road signs consistently from a particular angle during rainy weather.

The data presented in Fig. 9 was obtained from three repetitions for each road sign under rainy conditions. The experiment's highest accuracy in detection and recognition remains at 0.96, equivalent to a percentage score of 96.00%, corresponding to the "No Entry" label. Conversely, the lowest accuracy was observed for the "Children Crossing" label, scoring 0.90. These findings highlight the system's varying performance in recognizing different road signs under challenging rainy conditions.

CONCLUSIONS

In conclusion, the study successfully developed an automated

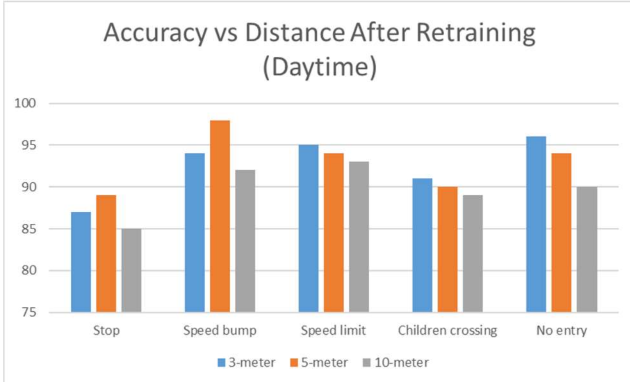


Fig. 7: Results accuracies after retraining (daytime).

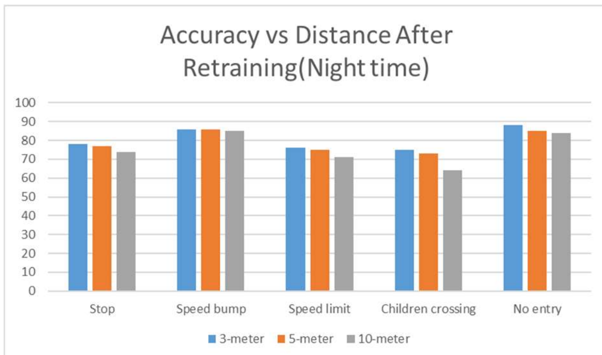


Fig. 8: Results accuracies after retraining (night time).

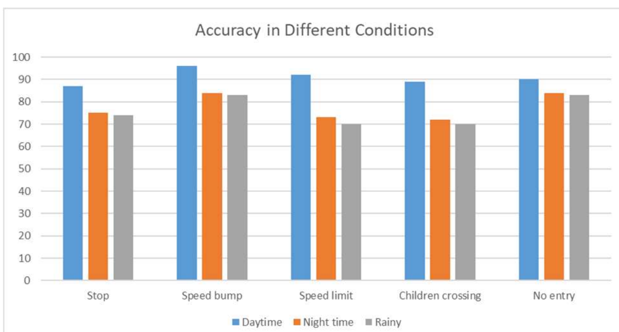


Fig. 9: Results accuracy during rainy conditions.

CNN-based system for traffic sign recognition. Following a thorough evaluation in a range of conditions, including varying illumination, distances, and weather, the system demonstrated respectable average accuracy levels: 89.8% in the daytime, 75.6% at night, and 76.4% in the rain. While these results are encouraging, there are significant drawbacks to using a phone camera as the input device. Most importantly, there was inadequate training time, which resulted in reduced accuracy for some road sign labels (such as “Stop” and “Children Crossing”). Future research should focus on reducing these limitations by using high-speed cameras that can capture moving images at exposure lengths of less than 1/1000 seconds, potentially improving image quality and precision. Additionally, further exploration into optimizing training procedures and dataset augmentation techniques could contribute to improving overall system performance. By addressing these challenges and exploring new avenues for refinement, future iterations of automated road sign recognition systems can aspire to achieve even higher levels of accuracy and reliability, ultimately enhancing driving safety and efficiency.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknikal Malaysia Melaka for enabling the research to be carried out.

REFERENCES

Behera, S.A.S.M.I.T.A., Bhoi, S.S., Mishra, A.S.U.T.O.S.H., Nayak, S.S., Panda, S.K. and Patnaik, S.S., 2022. Comparative study of convolutional neural network and long short-term memory network for solar irradiance forecasting. *Journal of Engineering Science and Technology*, 17(3), pp.1845-1856.

bin Md Fauadi, M.H.F. and Murata, T., 2010. Makespan minimization of machines and automated guided vehicles schedule using binary particle swarm optimization. *Lecture Notes in Engineering and Computer Science*, 2182, pp.1897–1902.

Chen, H., He, Z., Shi, B. and Zhong, T., 2019. Research on recognition method of electrical components based on YOLO V3. *IEEE Access*, 7, pp.157818-157829.

Dewi, C., Chen, R.C., Jiang, X. and Yu, H., 2022. Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4. *Multimed Tools Appl*, 81(26), pp.37821-37845.

Fredj, H.B., Chabbah, A., Baili, J., Faiedh, H. and Souani, C., 2023. An efficient implementation of traffic signs recognition system using CNN. *Microprocess Microsyst*, 98, p.104791.

Huang, Y.Q., Zheng, J.C., Sun, S.D., Yang, C.F. and Liu, J., 2020. Optimized YOLOv3 algorithm and its application in traffic flow detections. *Applied Sciences*, 10(9), p.3079.

Khan, M.A., Park, H. and Chae, J., 2023. A lightweight convolutional neural network (CNN) architecture for traffic sign recognition in urban road networks. *Electronics*, 12(8), p.1802.

Kiliçarslan, S., Közkurt, C., Ba□, S. and Elen, A., 2023. Detection and classification of pneumonia using novel Superior Exponential (SupEx) activation function in convolutional neural networks. *Expert Syst Appl*, 217, p.119503.

Lee, W.K., Abdullah, M.D., Ong, P., Abdullah, H. and Teo, W.K., 2021. Prediction of flank wear and surface roughness by recurrent neural network in turning process. *Journal of Advanced Manufacturing Technology (JAMT)*, 15(1), pp.55-67.

Li, H., Dong, Y., Liu, Y. and Ai, J., 2022. Design and implementation of uavs for bird's nest inspection on transmission lines based on Deep Learning. *Drones*, 6(9), p.252.

Luo, H., Yang, Y., Tong, B., Wu, F. and Fan, B., 2017. Traffic sign recognition using a multi-task convolutional neural network. *IEEE Trans Intell Transp Syst*, 19(4), pp.1100-1111.

Mittal, P., Singh, R. and Sharma, A., 2020. Deep learning-based object detection in low-altitude UAV datasets: A survey. *Image and Vision Computing*, 104, p.104046.

Noor, A.M., Fauadi, M.M., Jafar, F.A., Nordin, M.H., Yahaya, S.H., Ramlan, S. and Aziz, M.S.A., 2017. Fuzzy Analytic Hierarchy Process (FAHP) Integrations for Decision Making Purposes: A Review. *Journal of Advanced Manufacturing Technology (JAMT)*, 11(2), pp.139-154.

Peng, F., Miao, Z., Li, F. and Li, Z., 2021. S-FPN: A shortcut feature pyramid network for sea cucumber detection in underwater images. *Expert Syst Appl*, 182, p.115306.

Qingyun, F., Lin, Z. and Zhaokui, W., 2020. An efficient feature pyramid network for object detection in remote sensing imagery. *IEEE Access*, 8, pp.93058-93068.

Ramlan, S., Fauadi, M.H.F.M., Razali, N.H. and Hao, X., 2021. Agent-Based Chemical Mechanical Planarization Qualification for Semiconductor Wafer Fabrication. *Journal of Advanced Manufacturing Technology*, 15(3), pp.41-53.

Razi, A., Chen, X., Li, H., Wang, H., Russo, B., Chen, Y. and Yu, H., 2023. Deep learning serves traffic safety analysis: A forward-looking review. *IET Intell Transp Syst*, 17(1), pp.22-71.

Rosli, N.S., Fauadi, M.H.F.M., Awang, N.F. and Noor, A.Z.M., 2018. Vision-based defects detection for glass production based on improved image processing method. *Journal of Advanced Manufacturing Technology (JAMT)*, 12(1), pp.203-212.

Stallkamp, J., Schlipsing, M., Salmen, J. and Igel, C., 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*, 32, pp.323-332.

Sudhakar, M. and Priya, R.M., 2023. Computer Vision Based Machine Learning and Deep Learning Approaches for Identification of Nutrient Deficiency in Crops: A Survey. *Nat. Environ. Pollut. Technol.*, 22(3).