

[Open Peer Review on Qeios](#)

Smart Motorcycle Helmet: A CNN-Based Multi-Task Learning Approach

Rodrigues Jesan¹, Kumar Pant²¹ Universidad Univer² Anna University

Funding: Anna University

Potential competing interests: No potential competing interests to declare.

Abstract

Automated motorcycle helmet detection through video surveillance is crucial for enhancing road safety through education and enforcement. Existing approaches face limitations, such as difficulty in tracking individual motorcycles and distinguishing drivers from passengers. To address these challenges, we propose a CNN-based multi-task learning (MTL) method for identifying and tracking individual motorcycles, with a focus on rider-specific helmet use. We introduce the HELMET dataset, comprising 91,000 annotated frames from 10,006 motorcycles across 12 observation sites in Myanmar, providing a benchmark for future detection approaches. Our MTL approach, leveraging concurrent visual similarity learning and helmet use classification, achieves improved efficiency and accuracy. The proposed method operates at over 8 FPS on consumer hardware, yielding a weighted average F-measure of 67.3% for detecting riders and helmet use. Our work showcases deep learning as an accurate and resource-efficient means of collecting critical road safety data. We present an intelligent motorcycle helmet featuring infrared transceivers, an image sensor, an embedded computation module, a charging module, a microphone, and earphones. Designed for large vehicle approach intimation, the helmet uses image recognition modes for day and night conditions. Experimental results demonstrate successful vehicle registration plate recognition for large trucks/buses, achieving up to 75% accuracy during the day and 70% at night. The proposed intelligent motorcycle helmet detects approaching large vehicles in real time within a 5-meter distance.

1. Introduction

In contemporary times, motorcycles have become widely embraced as a convenient and popular mode of transportation across various regions, including China, Indonesia, India, the Philippines, Malaysia, Taiwan, Thailand, and Vietnam. Offering distinct advantages such as affordability, superior fuel efficiency, and a compact size suitable for easy parking in congested areas, motorcycles dominate the transportation landscape in many Asia-Pacific regions poses significant risks and potential dangers. Globally, there are approximately 200 million motorcycles, with a density of around 33 motorcycles per 1,000 people ^[1]. The fatality rate per vehicle mile traveled for motorcycles in 2007 was 37 times higher than that of passenger cars, according to data from the National Highway Traffic Safety Administration (NHTSA) in the United States.

Disturbingly, the World Health Organization (WHO) reports an annual toll of 2.17 million motorcyclist deaths from traffic accidents, with an additional 20 to 50 million sustaining injuries.

In response to the critical issue of road safety for motorcyclists, numerous countries have implemented laws mandating the use of helmets. Research indicates that wearing a helmet can significantly reduce traffic accident fatalities by nearly one-third (approximately 29%). Taiwan, for instance, grapples with safety concerns given its substantial motorcycle population, constituting 63.83% of the total motor vehicles by the end of 2023 [2]. Alarming statistics reveal that over 3,500 people in Taiwan perish annually in road traffic accidents, with motorcyclists accounting for more than 70% of these incidents [3]. The risk of death per kilometer traveled on a motorcycle is 20 times higher than that for car drivers or passengers. Large vehicles, such as trucks and buses, are frequently implicated in motorcycle collisions, exacerbated by factors like blind spots and front-view dead angles [4].

To address the pressing issue of heavy vehicle rear collisions and safeguard motorcyclists, this paper proposes an intelligent motorcycle helmet. This innovative helmet integrates infrared (IR) sensors with an image sensor, utilizing real-time image processing and recognition methodologies to identify large trucks and buses [5]. The system aims to prevent collisions and enhance road safety for motorcyclists, particularly in congested areas where traditional alerts may lose effectiveness. Moving forward, the paper introduces a novel system for helmet use detection in Intelligent Transportation Systems (ITS) using a combination of Transformer models for vision and the Cascade RCNN framework [6]. Leveraging the Swin Transformer as a feature extraction backbone, the proposed method outperforms existing CNN-based approaches in detecting helmet usage with a mean average precision (mAP) of 30.4 [7].

The study emphasizes the potential of Transformer-based models in solving computer vision problems within the realm of road safety. By introducing cutting-edge technologies and methodologies, the paper aims to contribute to the ongoing efforts to mitigate the risks associated with motorcycle accidents and improve overall road safety. Traditional methods for detecting active motorcycles typically follow a standardized procedure. Initially, a background subtraction technique is employed to isolate moving objects or vehicles from video data [8]. Subsequently, a binary classifier, such as a support vector machine (SVM), is applied to identify motorcycles. The next step involves localizing the head region of motorcyclists, followed by the utilization of an additional classifier to differentiate between helmet use and non-helmet use. To enhance the performance of the binary classifier, handcrafted features, like extracting a histogram of oriented gradients (HOG) from detected head regions, are commonly employed [9].

However, these methods encounter limitations when dealing with scenarios involving numerous motorcycles or multiple riders on a single motorcycle. In contrast to handcrafted feature design, deep learning methods aim to automatically derive representations from raw image data that are optimal for helmet use detection. In [10], helmet use is classified within detected head regions using a convolutional neural network (CNN). In [11] and [12], two separate CNNs are trained—one to distinguish motorcycles from other vehicles, and the other to classify helmet and non-helmet use in the head region of riders. To address the time-consuming nature of employing two separate CNNs for motorcycle and helmet use detection, [13] and [14] advocate using a single CNN for simultaneous detection of motorcycles and helmet use. However, the tracking of individual motorcycles across single frames of recorded video is only incorporated in half of the existing

approaches outlined in Table 1. As video data from traffic surveillance infrastructure inherently follows a frame-based structure, helmet use data generated through automatic detection must be mapped onto individual motorcycles to enable an accurate evaluation of helmet use [15]. This necessitates the remapping of frame-based detection results for motorcycle and rider counts, as well as helmet use, onto individual motorcycles appearing in multiple frames. Unfortunately, some approaches lack this cross-frame tracking element.

To address the absence of tracking, alternatives include utilizing single-frame detection at a fixed point or line in the frame to prevent repeated detection of the same motorcycle or collecting helmet use data in every video frame without tracking, resulting in a loss of information regarding the number of motorcycles at an observation site [16]. Both shortcuts compromise the quality of helmet use data and hinder the utilization of multiple frames of an individual motorcycle for helmet use and rider detection. Concerning rider number and position detection, only one of the approaches listed in Table 1 provides detailed information on this aspect [17]. While alternative approaches use head counts on the motorcycle as a proxy for rider numbers, this method lacks the granularity needed for accurate detection of rider number and position [18].

2. Methodology

The identification of motorcycles within a single frame constitutes a traditional object detection task. For this purpose, we employed a cutting-edge object detection algorithm trained specifically for motorcycle detection in our dataset. Presently, object detection algorithms can be broadly categorized into two approaches: one-stage and two-stage. While two-stage algorithms generally exhibit superior accuracy in object detection, they tend to be comparatively slower due to the need to process frames twice—once to identify potential object locations and once more to detect the actual objects. In contrast, single-stage methods consolidate the steps of localizing potential objects and object detection into a single processing stage [19]. This consolidation results in a slight reduction in accuracy but a significant decrease in processing time.

A relatively recent single-stage method, [20] employs a multi-scale feature pyramid in conjunction with focal loss to effectively overcome limitations in detection accuracy. It achieves faster detection compared to two-stage methods while maintaining a higher detection accuracy than comparable single-stage methods like YOLO [21]. Consequently, we opted to utilize a RetinaNet model for motorcycle detection. Given the similarity between motorcycle detection and other object detection tasks, we chose not to train the model from scratch. Instead, we fine-tuned a RetinaNet model using pre-trained weights obtained from the COCO dataset [22]. This approach leverages existing knowledge to enhance the model's proficiency in accurately identifying motorcycles in our specific dataset.

Within the framework of edge computing, data processing occurs on servers positioned at the edge of the network. These servers establish direct connections with a myriad of sensors and controllers, enabling them to analyze information and execute machine learning algorithms for real-time decision-making [23]. This project employs an ESP32 module equipped with an integrated camera functioning as a Wi-Fi camera. The processed data stream is securely transmitted to the Google Cloud Platform via the Cloud IoT Core, facilitated by a Raspberry Pi board serving as a local server executing the

TensorFlow object detection model [24]. The data undergoes event-driven processing, triggering alerts as necessary. Additionally, a local server facilitates access to a web interface for offline monitoring of cameras, while Firebase cloud functions are responsible for archiving data on Firebase. This archival process facilitates the streaming of video to internet-connected users via the web interface.

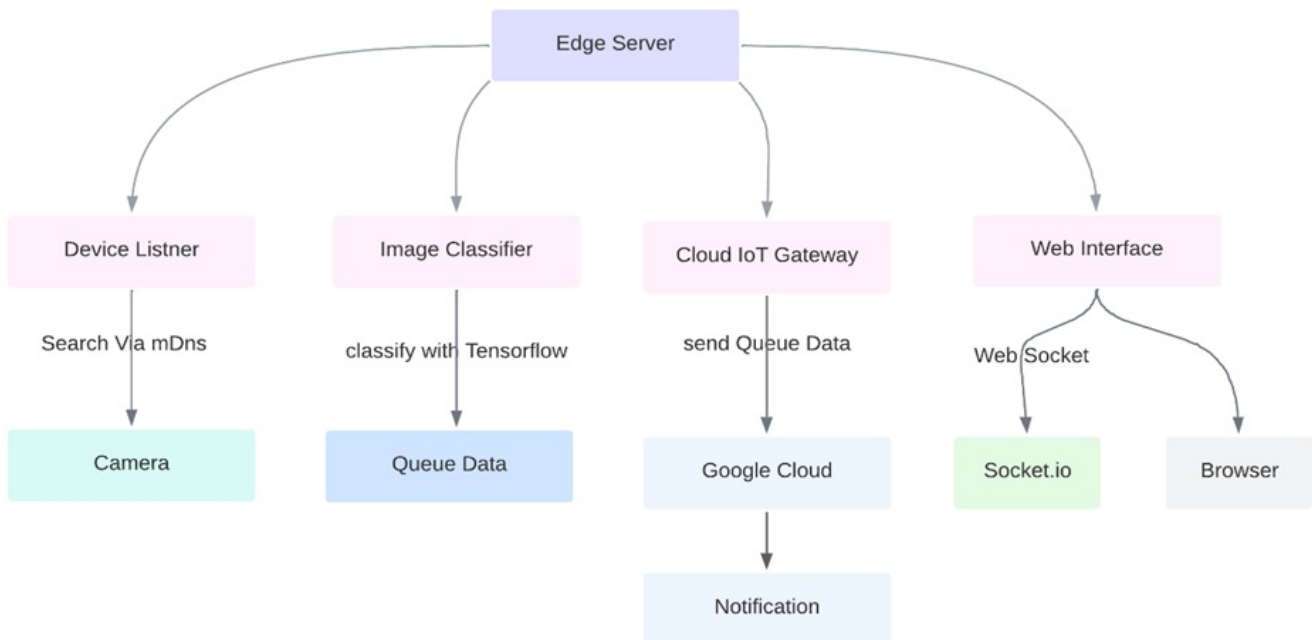


Figure 1. TensorFlow architecture overview

The Raspberry Pi Gateway scans the network via mDNS to locate local cameras, identifying objects and transmitting processed data to the cloud [25]. It also hosts a web interface for local data access. The @tensorflow/models package offers diverse pre-made machine learning models through NPM for various data types and purposes.

3. Model Training and classification

For a closed track of sufficient length, its helmet use is estimated by pooling the helmet use prediction of cropped image patches within the track. More specifically, let $(x^{(n)})_{n=1}^N$ be the cropped image patches that are assigned to a tracked motorcycle, then the track's helmet use class is estimated

The Swin Transformer architecture processes a sequence of tokens as its input, with these tokens being generated through the application of a patch partition layer on the input image, dividing it into N patches [26]. The architecture's hidden layers consist of multiple blocks, each comprising a multi-head self-attention module (MSA) as depicted in Figure 2. This MSA module employs an attention function on a set of query (Q), key (K), and value (V) vectors. It maps the query to a collection of key-value pairs, producing an output through the dot product of the query vector with all key vectors. A softmax function is then employed to scale the inner products and normalize them into weights (k), as defined in Eq.1:

$$\text{Attention}(Q, K, V) = \text{Softmax}(dkQKT) V$$

Here, dk represents the key dimension, and normalization is carried out by dividing by dk . The Swin Transformer introduces two variations of the multi-head self-attention module: the window-based Multi-head self-attention module (W-MSA) and the shifted window multi-head self-attention module (SW-MSA).

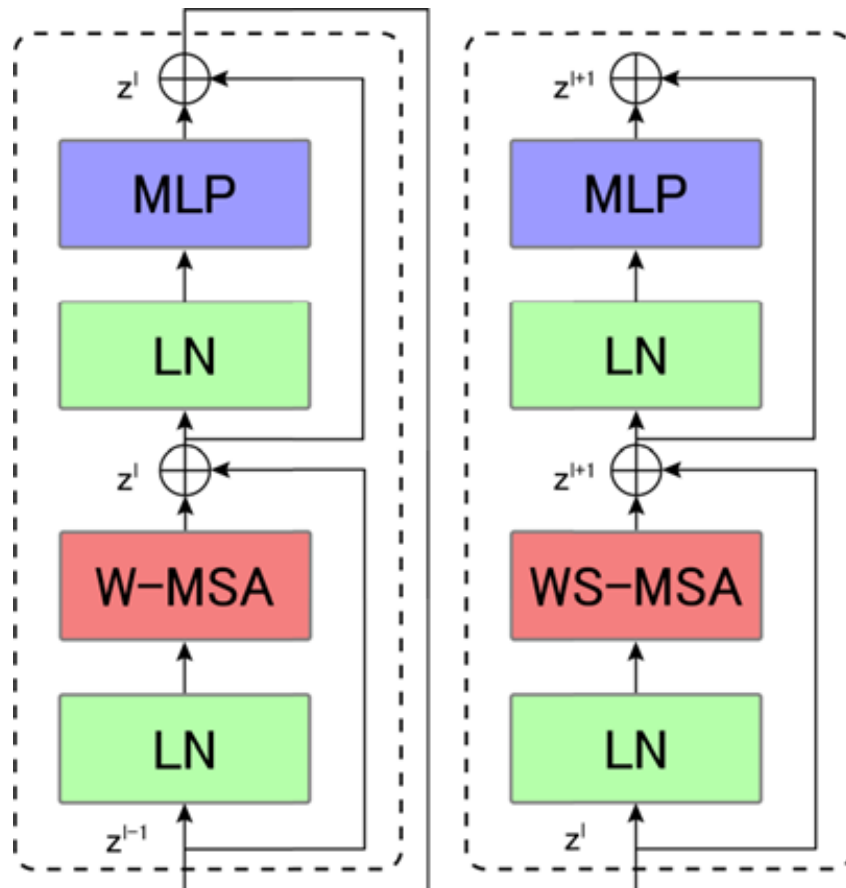


Figure 2. Two consecutive Swin Transformer blocks. The first consists of a regular window multi-head self-attention module (W-MSA), while the latter uses a shifted window configuration (WS-MSA).

The W-MSA module conducts attention calculations locally, applying self-attention on non-overlapping windows. To facilitate cross-window self-attention computations, the SW-MSA module performs identical calculations as the W-MSA module but after shifting the windows. This innovative approach enhances the Swin Transformer's ability to capture relationships within the input sequence using window-based attention mechanisms.

Training the model involves leveraging TensorFlow's Object Detection API, which simplifies the creation, training, and deployment of object detection models. Each model within TensorFlow's repertoire is characterized by its Speed, Mean Average Precision (mAP), and Output [27]. Typically, a higher mAP corresponds to a slower speed. In our case, we opted for the EfficientDet-Lite0 architecture for training. However, the choice of model architecture can vary depending on the priority between speed and accuracy. The EfficientDet-Lite [1][2][3][4] family comprises mobile/IoT-friendly object detection

models derived from the EfficientDet architecture, offering a range of options suitable for different trade-offs between speed and precision.

Table 1. Precision for the sample data

Model architecture	Size (MB)	Latency (ms)	Average Precision
EfficientDet-Lite0	4.5	186	27.69%
EfficientDet-Lite1	5.8	359	33.55%
EfficientDet-Lite2	7.2	456	36.97%
EfficientDet-Lite3	11.4	916	39.70%
EfficientDet-Lite4	19.9	1986	43.96%

We conducted a more in-depth analysis of the results, categorizing them by the number of occupants per motorcycle, and observed improved outcomes for classes with up to two riders, as detailed in Table 1. This improvement is likely due to the prevalence of single or dual-rider scenarios in real-world settings, resulting in a larger dataset for these particular classes. However, our scrutiny also brought to light some dataset limitations, notably the presence of imbalanced classes. Some classes lacked sufficient samples in both the training and test sets, with classes 21 to 32 having no examples, and classes 33 to 36 having fewer samples in the training set and none in the test set. This imbalance posed challenges for accurate detection of these categories by the models.

Our proposed model, incorporating the Swin Transformer as a backbone and integrating it with a Cascade RCNN framework for object detection, outperformed all other models in terms of mean average precision (mAP) across 36 classes. Notably, it surpassed the YOLOv7 ^[28] model, which achieved the second-highest mAP score. This superiority can be attributed to the Swin Transformer's more effective feature extraction capabilities, leveraging attention mechanisms to extract pertinent features from input images. In contrast, YOLOv7 relies on predefined anchor boxes and a CNN-based backbone for feature extraction, a method that may not be as proficient as the attention mechanism in identifying and extracting relevant features. While the YOLOv7 model shows a slight edge in weighted mean average precision, it falls short of outperforming the Swin Transformer and Cascade RCNN model in the overall mAP.

4. Intelligent Helmet with Advanced Vehicle Recognition and Alert System

While this intelligent motorcycle helmet excels in recognizing rear-approaching vehicles, it may encounter limitations in identifying larger vehicles. Consequently, its applicability becomes constrained, particularly in congested areas, as depicted in Fig. 3. Notably, existing studies have yet to address the need for an active intelligent helmet capable of detecting large truck/bus proximity, despite the severity of accidents involving such vehicles.

To address this gap, this paper introduces an intelligent helmet that integrates various electronic components, including miniature IR transceivers, a compact image sensor (camera), an embedded systems module, and a battery charging

module. The proposed system employs an image recognition method to ascertain the presence of an approaching heavy vehicle in the motorcyclist's vicinity. Furthermore, when such a vehicle is detected, earphones are utilized to promptly deliver a voice alert, prompting the motorcyclist to swiftly navigate away without the need for constant mirror checks. This design allows the motorcyclist to focus on the vehicles ahead without compromising safety.

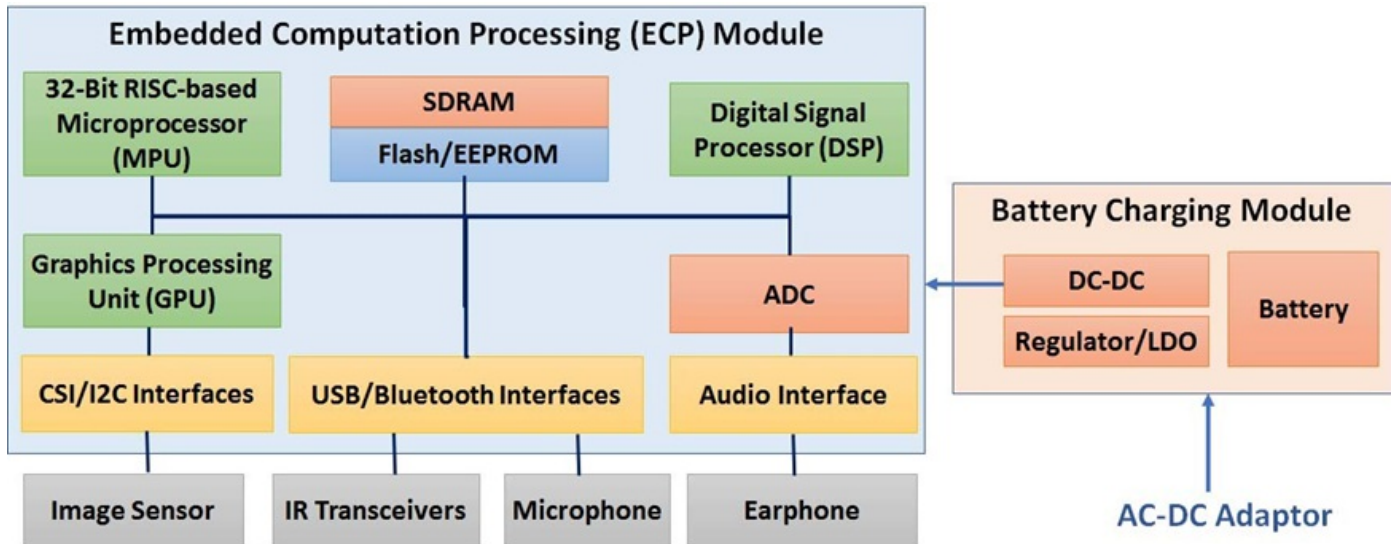


Figure 3. The system block diagram of the hardware implementation for the proposed intelligent helmet.

The integration of TensorFlow models in renewable energy into electric vehicles represents a groundbreaking convergence of two cutting-edge technologies with profound implications for sustainable transportation and energy efficiency. This integration harnesses the power of machine learning and renewable energy sources to optimize the performance, range, and environmental impact of electric vehicles (EVs) [29].

At its core, TensorFlow serves as a robust framework for developing and deploying machine learning models, including those tailored for renewable energy applications.

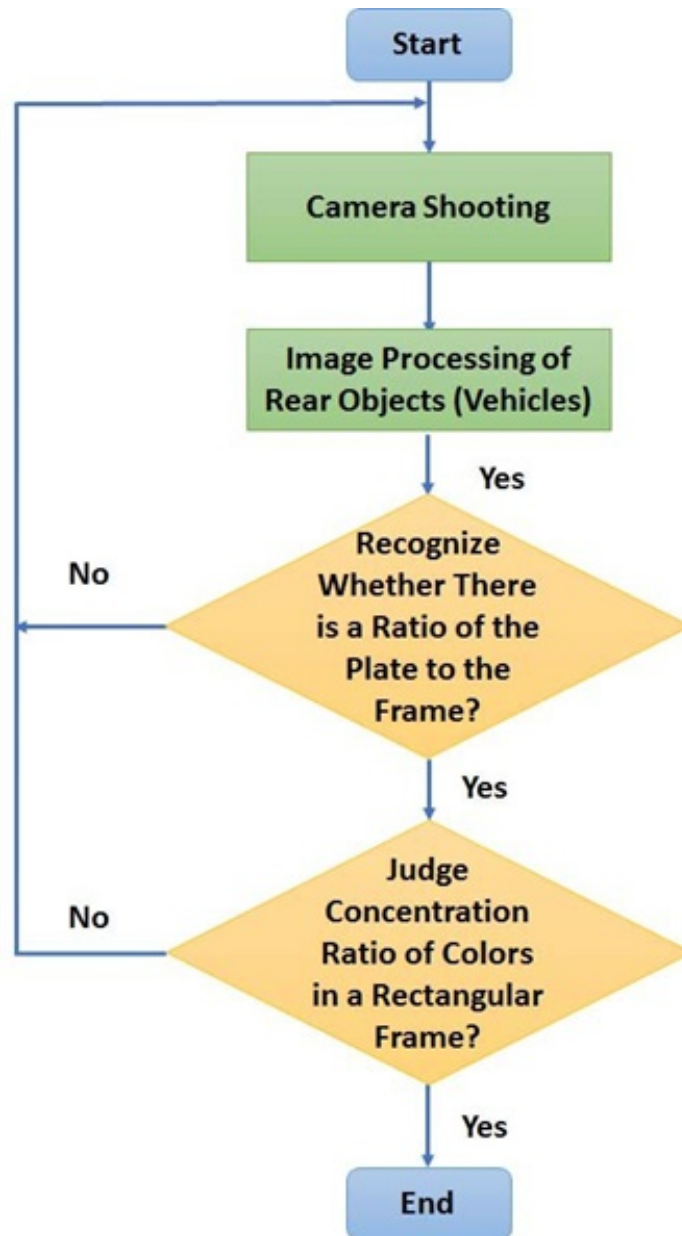


Figure 4. Flowchart of the detection mode

This paper employs the aforementioned six steps of the image processing method to analyze the original image captured, aiming to identify the black rectangle that may represent the license registration plate. Fig. 4 outlines two decision-making processes. The first process involves assessing whether there is a proportional relationship between the plate and the frame. The objective here is to discern the potential presence of a license registration plate in the image. The second process entails evaluating the color ratios within a rectangular frame to confirm the existence of license registration plates on larger vehicles in the image.

A standard ratio is applied to an object image with dimensions of "2.375" in length and "1" in width. This ratio serves as a criterion to determine the presence of the correct length and width for the license plate rectangle. If these criteria are not met, the system proceeds to capture a new image for re-recognition. Conversely, if the dimensions align, the next step involves assessing whether the rectangular frame meets the pre-established color density threshold. If it falls short, the

system prompts the capture of a new image for subsequent recognition.

5. Conclusion

This paper introduces an intelligent motorcycle helmet designed to enhance motorcyclist safety. The proposed helmet incorporates an image recognition approach focused on identifying large trucks and buses on the road. To achieve this, two recognition algorithms, one for daytime and another for nighttime, are developed specifically for detecting vehicle registration plates on large Taiwanese trucks and buses. The accuracy of the registration plate recognition is evaluated using a dataset of 600 images capturing rear-approaching trucks and buses on real roads, collected by 10 motorcyclists during both day and night conditions. The recognition accuracy rates are reported as approximately 78% at night and 85% during the day. Furthermore, the proposed intelligent helmet features integrated Bluetooth (BT) transmission, enabling the helmet to send notifications when a large truck or bus is approaching. Results from helmet use detection of tracked motorcycles reveal a weighted F-measure of 67.3%, demonstrating the capability of the approach to provide reliable estimates of motorcycle helmet usage, rider number, and position. The ablation study results indicate that the approach achieves a notably high accuracy against ablation experiments, albeit with some compromise on computational efficiency. Despite this, the approach processes at a speed of more than 8 frames per second (FPS) on consumer hardware, approaching real-time efficiency for 10 FPS video data. Overall, this work demonstrates the successful implementation of all four fundamental elements of helmet use registration through a CNN-based approach that is computationally efficient on consumer hardware.

References

- ^{a, b}W. -J. Chang and L. -B. Chen, "Design and Implementation of an Intelligent Motorcycle Helmet for Large Vehicle Approach Intimation," in *IEEE Sensors Journal*, vol. 19, no. 10, pp. 3882-3892, 15 May 15, 2019, doi: 10.1109/JSEN.2019.2895130.
- ^{a, b}M. L. N, A. E. Rao and M. P. Kalyan, "Real-Time Object Detection with Tensorflow Model Using Edge Computing Architecture," *2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, pp. 01-04, 2022.*
- ^{a, b}A. Bouhayane, Z. Charouh, M. Ghogho and Z. Guennoun, "A Swin Transformer-Based Approach for Motorcycle Helmet Detection," in *IEEE Access*, vol. 11, pp. 74410-74419, 2023, doi: 10.1109/ACCESS.2023.3296309.
- ^{a, b}Hemendra Kumar, Mohit Kumar, Mahiban Lindsay, "Smart Helmet for Two-Wheeler Drivers" *International Journal of Engineering Research And Advanced Technology, Volume 2, Issue 05, Pages 156-159, 2016.*
- [^]K T Maheswari et al, "Design and Implementation of Intelligent Head Protective Gear for Accident Detection and Notification," *IOP Conf. Ser.: Mater. Sci. Eng. 1084 012082, 2021.*
- [^]H. Abedi, S. Luo, V. Mazumdar, M. M. Y. R. Riad and G. Shaker, "AI-Powered In-Vehicle Passenger Monitoring Using Low-Cost mm-Wave Radar," in *IEEE Access*, vol. 10, pp. 18998-19012, 2022, doi: 10.1109/ACCESS.2021.3138051.
- [^]A. T. Jacob and N. Mahiban Lindsay, "Designing EV Harness Using Autocad Electrical," *2022 8th International*

- Conference on Smart Structures and Systems (ICSSS), Chennai, India, pp. 1-4, 2022.*
8. [^]H. Lin, J. D. Deng, D. Albers and F. W. Siebert, "Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning," in *IEEE Access*, vol. 8, pp. 162073-162084, 2020, doi: 10.1109/ACCESS.2020.3021357.
 9. [^]R. S. Charran and R. K. Dubey, "Two-Wheeler Vehicle Traffic Violations Detection and Automated Ticketing for Indian Road Scenario," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22002-22007, Nov. 2022, doi: 10.1109/TITS.2022.3186679.
 10. [^]D. V. K. Sarma and N. M. Lindsay, "Structural Design and Harnessing for Electric vehicle Review," 2023 9th International Conference on Electrical Energy Systems (ICEES), Chennai, India, pp. 107-111, 2023.
 11. [^]S. R. Rupanagudi et al., "A novel video processing based smart helmet for rear vehicle intimation & collision avoidance," in *Proc. Int. Conf. Comput. Netw. Commun. (CoCoNet)*, Dec. 2015, pp. 799–805.
 12. [^]S. Gupta and S. G. Mazumdar, "Sobel edge detection algorithm," *Int. J. Comput. Sci. Manage. Res.*, vol. 2, no. 2, pp. 1578–1583, Feb. 2013.
 13. [^]Lindsay N. Mahiban, Emimal M.. "Longevity of Electric Vehicle Operations". Qeios. doi:10.32388/ZAPC23.2., 2023
 14. [^]N. Garg, "Binarization Techniques used for grey scale images," *Int. J. Comput. Appl.*, vol. 71, no. 1, pp. 8–11, Jan. 2013
 15. [^]T. Huang, G. Yang, and G. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 1, pp. 13–18, Feb. 1979.
 16. [^]E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?" *Ann. Statist.*, vol. 37, no. 3, pp. 1172–1206, Jun. 2009.
 17. [^]N.K. Rayaguru, N. Mahiban Lindsay, Rubén González Crespo, S.P. Raja, "Hybrid bat–grasshopper and bat–modified multiverse optimization for solar photovoltaics maximum power generation", *Computers and Electrical Engineering*, Volume 106, 2023, 108596, ISSN 0045-7906, doi: 10.1016/j.compeleceng.2023.108596.
 18. [^]T. North, M. Wagner, S. Bourquin, and L. Kilcher, "Compact and high-brightness helmet-mounted head-up display system by retinal laser projection," *IEEE J. Display Technol.*, vol. 12, no. 9, pp. 982–985, Sep. 2016.
 19. [^]R. A. Alimbuyog, J. C. D. Cruz, and R. V. Sevilla, "Development of motorcycle data logging system with visual basic data simulation for accident analysis," in *Proc. IEEE 9th Int. Conf. Humanoid, Nanotech- nol., Inf. Technol., Commun. Control, Environ. Manage. (HNICEM)*, Dec. 2017, pp. 1–5.
 20. [^]C. Spelta, V. Manzoni, A. Corti, A. Goggi, and S. M. Savaresi, "Smartphone-based vehicle-to-driver/environment interaction system for motorcycles," *IEEE Embedded Syst. Lett.*, vol. 2, no. 2, pp. 39–42, Jun. 2010.
 21. [^]Mahiban Lindsay, M. Emimal, "Fuzzy logic-based approach for optimal allocation of distributed generation in a restructured power system", *International Journal of Applied Power Engineering (IJAPE)* Vol. 13, No. 1, March 2024, pp. 123~129, ISSN: 2252-8792, DOI: 10.11591/ijape.v13.i1.pp123-129
 22. [^]R. A. Alimbuyog, J. C. D. Cruz, and R. V. Sevilla, "Development of motorcycle data logging system with visual basic data simulation for accident analysis," in *Proc. IEEE 9th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ. Manage. (HNICEM)*, Dec. 2017, pp. 1–5.
 23. [^]J. Chiverton, "Helmet presence classification with motorcycle detection and tracking," *IET Intell. Transport Syst.*, vol. 6, no. 3, pp. 259–269, Sep. 2012.

24. [^]N. Mahiban Lindsay and A. K. Parvathy, "Power system reliability assessment in a complex restructured power system," *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 9, No. 4, pp. 2296~2302, August 2019.
25. [^]S. Issaoui, R. Ejbéli, T. Frikha, and M. Abid, "Embedded approach for edge recognition: Case study: Vehicle registration plate recognition," in *Proc. 13th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2016, pp. 336–341.
26. [^]B. Tejas, D. Omkar, D. Rutuja, K. Prajakta, and P. Bhakti, "Number plate recognition and document verification using feature extraction OCR algorithm," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2017, pp. 1317–1320.
27. [^]K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
28. [^]Lindsay, N.M., Parvathy, A.K.. "Simulation and Application on Power System Reliability for Bulk Electrical System", *Lecture Notes in Electrical Engineering*, vol 326. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2119-7_111
29. [^]R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587