2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 979-8-3503-8564-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICICV62344.2024.00042

Survey on Traffic Violation Prediction using Deep Learning based on Helmets with Number Plate Recognition

P.Pandiaraja
Department of Computer Science and Engineering
M.Kumarasamy College of Engineering,
Karur, Tamilnadu,India -639113
sppandiaraja@gmail.com

S.Abisheck
Department of Computer Science and Engineering
M.Kumarasamy College of Engineering,
Karur, Tamilnadu,India -639113
sekarabisheck@gmail.com

A.Mohan
Department of Computer Science and Engineering
M.Kumarasamy College of Engineering,
Karur, India Tamilnadu, -639113
amohan0055@gmail.com

M.Ramanikanth
Department of Computer Science and Engineering
M. Kumarasamy College of Engineering,
Karur, Tamilnadu,India-639113
ramanikanthmadheswaran@gmail.com

***Abstract*- Nowadays, the most widely used mode of transportation is two-wheelers. Both bike riders and pillions are strongly encouraged to wear helmet. A significant application and developing field of study in machine learning and image processing, object tracking in video surveillance is of considerable interest to many academics. Object detection is the process of identifying the existence of items in a scene using a bounding box and the types or categories of the objects that have been placed. This study reviews tracking methods, classifies them into several sorts, and emphasizes the most significant and practical methods. Finally, this study analyzes potential study directions after reviewing general tactics as a part of the literature review on various approaches. This study uses an image processing system to identify motorcycle riders without helmets. Additionally, utilize the OCR method to extract user information from images of license plates and classify them using OCR. then figure out the fine amount. Making SMS services inform customers in order to avoid motor-bike accidents is the final step. The performance of the proposed framework is assessed in terms of precision and speed.**

Keywords-Number Plate Recognition, Object detection, Object Recognition, Optical Character recognition, Traffic Violations.

I. INTRODUCTION

Traffic safety is extremely important since accidents and injuries are increasingly likely to happen as a result of the growing congestion on our roadways in the modern world. Helmets have long been linked to the security and defence of motorcyclists and cyclists. However, helmets have more than just symbolic value; they are crucial for reducing the severity of brain damage incurred in accidents. To increase road safety, governments and law enforcement agencies have passed strict helmet-wearing regulations [11]. In this discussion, will look at the applicability of helmet traffic offences, the grounds for these laws, and the potential consequences for both people and society when helmet rules are ignored. The primary goal of helmet laws is to safeguard the lives and general welfare of motorcyclists, bicyclists, and other users of two-wheeled vehicles [12]. These rules typically require riders to wear helmets that meet high safety requirements, ensuring that the head and brain are suitably protected in the event of an accident. Since they have been proven to significantly reduce the risk of fatalities and head injuries in motorbike accidents, helmets are a crucial component of road safety. A crucial application in many safety-critical areas, including as building sites, sporting venues, and transportation, is the detection of helmets in pictures and video feeds. This system uses machine learning and image processing to detect people wearing helmets, ensuring that safety rules are followed and improving overall safety.

II. RELATED WORKS

Vision-based solution as suggested by Susa, Julie Ann B et. al [1] could be used to find out whether some of the cyclists are using misfitted helmets. With a deep learning approach (YOLOv3 model), the system applied the helmet detection model for this study. Multiple check-ups on different public roads may end up in traffic congestion. Their safety and requirement for use of approved helmets can only make a case for the early development of a detection system. This study will employ the YOLOv3 methodology. During testing and deployment, the system achieved scores between 90 per cent and above on all testing attributes.

However, Waris, Tasbeeha, and Tasmiyah [2] devised a faster R-CNN (Region-based Convolutional Neural Network) approach for automatically detecting bikers with no helmet. The system converts the video input to frames for purposes of detecting violation instances of helmet use. A self- made film from various places in Lahore, Pakistan, and online repositories constituted the dataset. The proposed technique is found to be highly accurate with an experimental study showing a whooping 97.69% accuracy. It could help to take appropriate legal measures against the people who violate traffic regulations. This work can be expanded in future to include other capabilities including number plate detection and possibly other traffic offences.

Tran, Duong Nguyen Ngoc and others [3], describe an approach of tracking distinctive motorbikes together with individual rider's helmet usage. According to most investigations, this helmet- use categorization technique is more efficient as compared to earlier ones. This strategy employs new techniques and algorithms on that established foundation so as to achieve even greater levels of accuracy and preciseness. By employing cutting edge technology and good practice developed a very powerful tool for identifying helmet wearing in different contexts, places and situations. Indeed, this measure will prove invaluable towards boosting awareness of safe riding and cutting on cases of unavoidable crashes and deaths.

Lightweight object detection network based on ML-YOLOv3 by Deng, Lixia, et al [4] This proposes three network optimization schemes that could dramatically cut down on the model's computational cost without losing the punch of the detection impact. Our proposed CSP-Ghost-ResNet model with the helmet dataset reduces model complexity while producing almost the same improvement compared to YOLOv3. ML-Darknet reduces the detection capability but also greatly reduces the modelling cost. This article also redraws the PAN-CGR-Network. This also means that it reduces the cost of computation.

Recently, there has been innovation adaptation of YOLOv5, a prevalent object detection model for enhancing its capabilities under different situation. A suggestion of a An, Qing, et al [5] altered YOLOv5 network, which targets extracting characteristic of tiny images. This is facilitated by resizing and shifting, boosting correspondence between anchor and target boxes. Also, they coupled the CPS module of the CBAM attention unit with the GAM strategy to improve the interplay between pixels globally and prevent the degradation of features as they crossed through the backbone and neck nets of the prior YOLOv5s architecture. To this end, three-dimensionally arranged channel attention and convolutional spatial attention sub-modules, coupled with combination layers are used to fine tune every convolutional block in order to enhance the model's capability of handling multi-dimensional spatial features as well as essential target characteristics.

Additionally, Hayat, Ahatsham, et al. [6] evaluated the performance of YOLOv3, YOLOv4, and YOLOv5x for safe helmet identification. YOLOv5x had the highest mAP for smaller things at a little over 92%. Nevertheless, it encountered difficulties in cases of a safety helmet, and the view was hindered by objects. By incorporating more diversified sets of data that include such difficult situations, a model's efficiency could be improved. In addition, future improvements in the object detection may also cover the detection of other pieces of safety clothing such as vests, work gloves, and goggles, improving workers' safety.

Consider replacing the Res8 module in Darknet53 with a parallel RSSE network module as suggested by Song, Hongru, et al. [7]. The model can also be strengthened in terms of network speed as well as smaller target recognition precision. Finally, using residual module Res2 instead of CBL5 modules on YOLOv3 method.

The present work introduces an improved version of RSSE-YOLOv3 that is more suitable for detecting dense occlusions. This enhances detection performance as revealed by ablation tests conducted on the helmet dataset.

To achieve real time and effective safety helmet detection, Chen, Junhua, et al [8] present improvements in the YOLOV4 algorithm. This modified YOLOv4 algorithm incorporates deep-wise separable convoluting for reducing the number of model parameters and employs PP-LCNet as the main network component. In order to enhance the capability to grasp target information, a novel improved feature fusion framework is constructed, and the coordinate attentional mechanism is incorporated inside the three result feature planes of the backbone. This proposes a new Siou loss function aimed at increasing detection accuracy by taking into consideration the direction information. Workers continue failing to follow the safety helmet norms in different workplaces, yet safety hazard remains a top concern. The traditional methods of carrying out helmet inspections involved patrols within the field and watching of the surveillance rooms. These, though, involve hard work, consume a lot of time, and might result in exhaustion, mistakes, or missed inspections by inspectors. Consequently, novel procedures are coming into light using sensors and picture analysis techniques, which will automize examination of security helmets wearing by build-up site employees.

Tsai, Chun - Ming, et al [9] YOLOv7-CBAM and YOLOv7-SimAM, two novel deep learning models that combine YOLOv7- E6E, CBAM, and SimAM, were presented. While the YOLOv7- CBAM and YOLOv7-SimAM models were trained on pictures of size is 1280, the YOLOv7- E6E model was trained on images of size 1920. These models were used to identify the test pictures in Track-5, and the outcomes were sent to the Track 5 assessment system for the AI City CHALLENGE. The experimental findings on the 100 test movies for Track 5 of the 2023 AI City CHALLENGE show the success of our techniques, with mAP scores for YOLOv7- E6E, YOLOv7-CBAM, and YOLOv7- SimAM of 0.6112, 0.6389, and 0.6422, respectively. Out of more than 36 competing teams, our suggested techniques came in 6th, 5th, and 4th, respectively, on the public scoreboard. One false positive motorcycle and two false positive D-Helmets were nonetheless created using YOLOv7- CBAM. Overall, the suggested models show how deep learning techniques may be used to enhance traffic safety measures and emphasise the significance of further study in this field. Since the source code for these models is readily accessible, others may build on these developments and advance the State-Of-The-Art.

Aboah, Armstrong, et al [10] To build a robust helmet detecting model using a tiny amount of annotation, "few shot data sampling method" was invented and applied. These algorithms are used to select a few pictures out of the large collection while generating more enhanced training images through means of data augmentation. Therefore, developed a solid helmet detection model using few annotations, which is important since it significantly reduces the times and efforts for annotation. Researchers have developed a real-time helmet identification system which is robust against shifting weather and time of the day based on the state-of-art YOLOv8 object detection model and data augmentation techniques. YOLOv8 is supposed to be fast and accurate and that is why one can use it for the real time application. Furthermore, some data augmentation methods including occlusion and perspective issues were employed in this paper. To boost accuracy and overall certainty, TTA was implemented during the inferential stage of the study. Table. I illustrated the overall details about literature survey.

TABLE I OVERALL DETAILS ABOUT LITERATURE SURVEY

Refer No	Techniques	Merits	Demerits
[1]	YOLOv3 mode	Classified helmet or non-helmet	Not support real time data
[2]	R-CNN	Video based helmetdetection	Computational complexity
[3]	CNN model	Classify the helmet wearing and non-wearingpersons	Support only imagedatasets
[4]	Cross Stage networks	Lightweight objectdetection network	Need large datasets for training
[5]	GAM with neural network	Effectively improves the speed	Time complexity can behigh
[6]	Circle Hough Transform (CHT)	Automatic detection of safety helmets	Image based analysis
[7]	RepVGG framework	Improving the detection accuracy	Mis-classification errorrate can be occurred
[8]	PP-LCNet framework	Multiple features are extracted and fusion	Small images are used
[9]	CBAM (Convolutional Block Attention Module)	Improving traffic safety measures	Does not support publiclyavailable datasets
[10]	Frame Sampling Algorithm	Single-stage objectdetection model	Support only videodatasets

III. EXISTING METHODOLOGIES

The leading cause for human mortality today is traffic related injuries. Motorbike accidents could cause serious injuries. All motorcycle riders should wear helmets. Nevertheless, many others willingly ignore the helmet use requirement. Provided are a number of safety devices, including helmets which should be worn by workers when conducting operations and inspections [13]. To this end, employees may at times fail to follow the rule and hence there is need of CCTV cameras that capture everything as seen in the whole plant and supervisors who should ascertain if the workers worn their helmets. There can be hundreds of monitoring screens but it is not easy to observe at least a helmet violation instance that has led to devastating collisions at any on moment of time [14]. With the fast development of image recognition technology, computer vision-based inspections have emerged as arguably the most important industrial field of application. The present one involves the checking whether employees wear their helmets and the distinction of the colors they use for them [15]. A color based hybrid descriptor is presented to extract the features of colored helmets made up of LBP, HSI, and CH. Subsequently, the features are grouped into the red helmet, yellow helmet, blue helmet and non-helmet categories using the H-SVM.

A. Local Binary Pattern For Helmet Detection

Local Binary Patterns (LBP), a texture analysis technique, may be used to identify helmets in photographs. LBP is useful for determining the texture of helmets in a number of situations due to its robustness in creating local texture patterns in a picture. LBP operates by evaluating the intensity of a pixel in relation to its neighbours. Each pixel in the image is given a binary code based on whether the pixels around it have intensities higher or lower than the pixel in the middle. After that, the binary value is converted to decimal form. In order to identify helmets, LBP may be used to extract texture features from an area of interest (ROI) around the head that may also contain the helmet. This feature extraction process describes the regional texture patterns seen inside the ROI. Create a collection of labelled pictures that includes both pictures of individuals wearing helmets and pictures of people who aren't wearing them. These pictures make up the training set for your LBP-based helmet identification model.

B. Hu-Moment Invariations

A collection of mathematical descriptors known as Hu Moment Invariants (HMI) is used in image processing and computer vision to describe the form and spatial distribution of an item in a binary or greyscale picture. Because they are invariant to translation, scaling, and rotation, these moment invariants are useful for form analysis and object recognition.Hu Moment Invariants, also known as H1, H2, H3, H4, H5, H6, and H7, are derived from an image's core moments. The following is a succinct description of each Hu Moment Invariant:

- H1 (Invariant to Scale): Measures the overall intensity of the image. It is invariant to scale changes in the object.
- H2 (Invariant to Rotation): Describes the skewness and orientation of the object. It remains unchanged when the object is rotated.
- H3 (Invariant to Rotation): Represents the balance or elongation of the object. Like H2, it is invariant to object rotation.
- H4 (Invariant to Scale and Rotation): Measures the fourth moment of the distribution of the object's pixel intensities, making it invariant to both scale and rotation.
- H5 (Invariant to Scale): Measures the compactness of the object. It is invariant to scale transformations.
- H6 (Invariant to Scale and Rotation): Represents the object's skewness. It is invariant to scale changes and rotations.
- H7 (Invariant to Scale): Measures the elongation or thinness of the object. It remains unchanged under scale transformations.

For helmet detection, Hu Moment Invariants can be used to characterize the shape and orientation of helmets, making it possible to distinguish helmets from other objects or background in images. By comparing the computed HMI of a region of interest (ROI) with pre-determined values for helmets, it can assist in identifying individuals wearing helmets, ensuring compliance with safety regulations.

C. Colour Histograms Based Object Detection

The distribution of colours in an image is statistically represented by colour histograms. They offer important details regarding the presence and quantity of different colours in a picture. For applications like image retrieval, object identification, picture segmentation, and more, colour histograms are often employed in image processing and computer vision. The colour space is divided into distinct bins by a colour histogram. There can be 256 bins per channel for an 8-bit colour space (such as RGB with 256 intensity levels for each channel). The granularity of the information about the colour distribution is influenced by the number of bins, which might vary. The image is scanned, and the colour values for each pixel are then given to the appropriate bin for each channel (for example, Red, Green, and Blue). The histogram measures the number of pixels that fit into each bin. It is standard practise to normalise histograms in order to make them comparable across various photos. To create a probability distribution, divide the count in each bin by the overall number of pixels in the image.

D. Support Vector Machine For Helmet Detection

Helmet characteristics have been shown using colour histograms and as shape-like elements. The separation of the dataset into a training and testing set. Thus, it's possible for even more cases of with and without helmets on. Parameters like accuracy, precision, recall, and F1 scores are used to evaluate the model's performance. In the case of real-time applications, where images or video frames are assessed by a trained SVM classifier, it separates between "helmets users" and "non-helmets users".

IV. PROPOSED METHODOLOGIES

As the traffic police cannot monitor non-helmeted motorcyclists based on their license plates automatically, they have to record such violations manually using either the vehicle's number or photo. However, this manual administration could sometimes lead to errors. To overcome these restrictions, developed an automatic helmet and face mask locating device based only on taking down number plates. Considering this, proposed a more accurate automated way of solving the problems in the present system with minimal human effort. The main purpose of this system is to point out people riding on motorcycles without proper headgears. Use the techniques in this chapter to identify helmet-offenders whilst on the ground.

An object detection and identification system can use the YOLO technique to identify the helmet. In addition, locate the user's identification by the use of Convolutional Neural Network for the recognition of the car number plate. This proposed method may detect the target object with different illuminations and occlusions. The system defines an object's category based on a feature that has been extracted. YOLO-Darknet is a system that uses deep learning architecture for computer vision based on pre-trained common objects in context. Using CNN, locate the objects and spot the text object for printing purposes. Fines can be sent through SMS to any person not putting on a helmet. Also, ensure that the database has been updated and check if payment is made in full. The number will be automatically blocked if the user fails to clear the payment. A message shall also be sent. The admin can only renew the licence plate after he/she receives the payment. Fig. 1 illustrates the overall design of the proposed system.

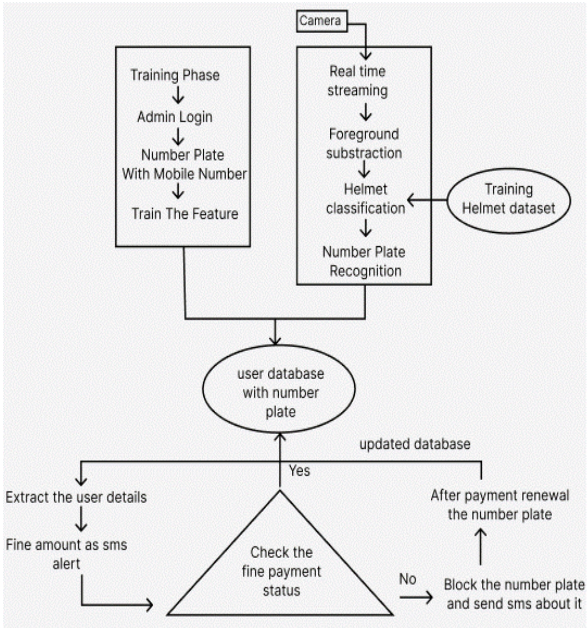

Fig. 1. Proposed architecture

A. YOLO Algorithm For Helmet Detection

YOLO is a technique that allows for instantaneous recognition of objects based on neural networks. This algorithm is so accurate and fast, this is why it enjoys popularity. It can be used on animals, humans, parking metres and traffic signals as an identification measure.

The YOLO algorithm employs the following three methods:

Remaining blocks:

- A bounding box is a shape that focuses on an object of an image.
- These are characteristic for each of the bounding boxes in this picture.
- Height (bw)
- Height (bh)

Regression using bounding boxes:

- Box regression bounding
- Bounding box refers to an outline that accentuates an object in a photograph.
- The following characteristics are present in each bounding box in the image:
- Height (bh)
- Width (bw)
- This is represented by "letter c" under class such as human, vehicle, traffic signal, etc.
- Centre bounding box (bx, by)

Intersection Over Union (IOU):

- Boxes are utilized in the object detection phenomenon called IOU to demonstrate overlaps. Therefore, YOLO uses IOU in order to produce the appropriate output box surrounding the objects.
- Each grid cell predicts the bounding boxes and their confidence ratings. The IOU equals one if the predicted bounding box and the actual box coincide. This approach eliminates non-identical bounding boxes with respect to the particular box Provide the name of the most accurate item according to the features.

B. Number Plate Recognition

Number plates are used to mark every car in the whole world. Number plate recognition uses a photo handling technology to ascertain cars using its number plate. For everyday traffic management and security functions like control of access to specific locations or tracking stolen vehicles, number plate recognition systems are applied.

The following actions occur once the number plate is detected:

1. To capture the photographic image of the license.
2. To distinguish between character groups.
3. A photo showing recognized license plate, displayed on a GUI, and stored together with the time and day in a database – for future references.
4. SMS the user's fine.

C. Optical Character Recognition

OCR has been the centre of attention since its purpose to decompose the visual document image into separate letters and numbers. For years, OCR's aim of decomposing image-based visual documents into separate letters and figures has received continuous attention. OCR has been the center of attention since its purpose to decompose the visual document image into separate letters and numbers.

However, in spite of years and years of development in the field, there still cannot be found a satisfactory solution as far as OCR systems are concerned—ones which may be capable of matching human capabilities. The study area is a very complex one attracting many interested scholars and engineers thus leading to the creation of different academic laboratories and research organizations specialized in character recognition systems. The purpose of this review is to give a broad summary of researches on OCR.

The function of OCR software is to transform scanned pictures and texts into computer readable electronic information. Machines find it hard to read and understand images or pictures, unlike people, hence the need for machine-readable documents projects. However, there are many complexities in OCR that include different languages, fonts, writing styles, and diverse linguistical laws, etc. Consequently, researches draw upon methods from different areas of computer science such as imaging processing, pattern recognition, and natural language processing which are multi-faceted challenges. It will discuss about various phases that involve OCR process which include the following: acquisition, pre-processing, segmentation, feature extraction, classification, and post processing. An efficient OCR system can then be put together by integrating these sub-processes. These systems are used in different fields e.g. smart library and real time number plate recognition proving their usefulness in reality.

V. EXPERIMENTAL RESULTS

Each phase is discussed extensively in this paper. The use of an effective OCR as a follow-up project.These techniques could be used together to develop a system. Moreover, it can be employed in many other.For example, smart libraries, number- plate recognition, and others real-time applications like practical uses.

• A perfect positive forecast is a TP.

• False positives (FP)-incorrect prediction of positivity.

• Number of true negatives (TN): absolute precision in

pre-empting adverse events

• False negative (FN): the number of true negatives arising due to false negatives Prediction.

Error rate

The error rate (ERR) is calculated as a percentage of all faulty forecasts to all test data. The absolute best error rate that may be achieved is 0.0, while the worst is 1.0.

Any classifier's main goal will be to reduce this error rate as much as possible. Table II. illustrates the Algorithm with error rate. Fig 2 illustrates the Algorithm with error rate .

$$ERR= (FP+FN)/(TP+TN+FN+FP)$$

TABLE II ERROR RATE

Algorithm	Error rate
RANDOM FOREST	0.75
SUPPORT VECTOR MACHINE	0.5
YOLO ALGORITHM	0.4

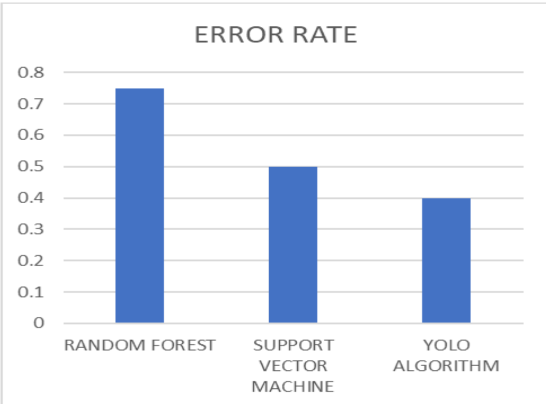

Fig. 2 Error rate

From above graph, the error rate was proposed for YOLO algorithm that is lesser than existing one.

Accuracy: This refers to the ratio between the total number of flawless forecasts and the entire test data.

(ACC). Moreover, some prefer it as 1-ERR. Maximum = 1.0 and Minimum Accuracy is 0.0. Table III illustrates the

Algorithm with Accuracy rate Fig. 3 illustrates the Algorithm with accuracy chart.

ACC= (TP+TN)/(TP+TN+FN+FP) x 100

TABLE III ACCURACY

Algorithm	**Accuracy rate**
RANDOM FOREST	50%
SUPPORT VECTOR MACHINE	65%
YOLO WITH OCR	80%

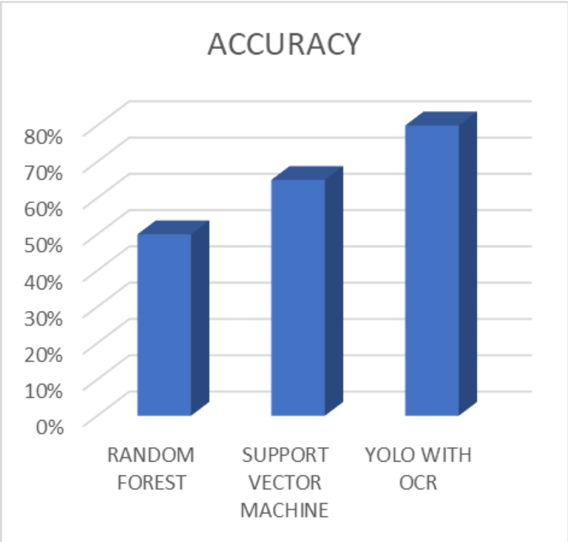

Fig 3: Accuracy chart

According to the graph above, the proposed YOLO with CNN algorithm offers a higher degree of accuracy than the current approach.

VI. CONCLUSION

The proposed system just deals with image datasets for helmet detection, and various methods for helmet detection have been provided in this research work. Thus, with real time video capturing may propose an automatic way of recognizing those bikers without helmets and getting back on their licence plates use YOLO and OCR in a bid to trace riders who do not wear helmet. However, it is not sufficient to point out or identify these motorists. The system will also record the motorcycle drivers' license plate numbers. It has employed YOLO in identifying real-time helmet usage from those without helmets. However, YOLO cannot be employed to determine all objects since a single cell might contain many. Furthermore, an automated license plate recognition system that utilizes advanced deep learning methods has been developed for use by Indian drivers. As compared to other techniques, the selected process outperforms at energizing large scale data of Indian character sets high variabilities consisting numbers plates generation and form a new dataset of Indian typographic variety. Successful training of the model was achieved using the Sequential CNN technique. Using this number, the transport office can access the rider's motorcycle registration number and therefore access the rider's detail from the licence vehicle database. Finally, it is bikers who appear to be concerned that they might suffer the consequences.

REFERENCES

[1] Susa, Julie Ann B., et al. "An efficient safety and authorized helmet detection using deep learning approach." 2022 International Conference on Smart Information Systems and Technologies (SIST). IEEE, 2022..

[2] Waris, Tasbeeha, et al. "CNN-Based Automatic Helmet Violation Detection of Motorcyclists for an Intelligent Transportation System." Mathematical Problems in Engineering 2022 (2022).

[3] Tran, Duong Nguyen-Ngoc, et al. "Robust automatic motorcycle helmet violation detection for an intelligent transportation system." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[4] Deng, Lixia, et al. "A lightweight YOLOv3 algorithm used for safety helmet detection." Scientific reports 12.1 (2022): 10981.

[5] An, Qing, et al. "Research on safety helmet detection algorithm based on improved YOLOv5s." Sensors 23.13 (2023): 5824.

[6] Hayat, Ahatsham, and Fernando Morgado-Dias. "Deep learning-based automatic safety helmet detection system for construction safety." Applied Sciences 12.16 (2022): 8268.

[7] Song, Hongru. "Multi-scale safety helmet detection based on RSSE-YOLOv3." Sensors 22.16 (2022): 6061.

[8] Chen, Junhua, et al. "Lightweight helmet detection algorithm using an improved YOLOv4." Sensors 23.3 (2023): 1256.

[9] Tsai, Chun-Ming, et al. "Video analytics for detecting motorcyclist helmet rule violations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[10] Aboah, Armstrong, et al. "Real-time multi-class helmet violation detection using few-shot data sampling technique and yolov8." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[11] J. Cao, Y. Pang, J. Xie, F. S. Khan, and L. Shao, ''From handcrafted to deep features for pedestrian detection: A survey,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 9, pp. 4913–4934, Sep. 2022.

[12] S. Srivastava, S. Narayan, and S. Mittal, ''A survey of deep learning techniques for vehicle detection from UAV images,'' J. Syst. Archit., vol. 117, Aug. 2021, Art. no. 102152

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ''An image is worth 16×16 words: Transformers for image recognition at scale,'' 2020, arXiv:2010.11929.

[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, ''Swin transformer: Hierarchical vision transformer using shifted windows,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 9992–10002.

[15] A.Chairat, M. N. Dailey, S. Limsoonthrakul, M. Ekpanyapong, and D. Raj K. C., ''Low cost, high performance automatic motorcycle helmet violation detection,'' in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2020, pp. 3549–3557.